Faculty of Engineering and Information Technology
University of Technology Sydney

Multivariate Sequential Contrast Pattern Mining and Prediction Models for Critical Care Clinical Informatics

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Shameek Ghosh

December 2017

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not been previously submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used have been reported in the thesis.

Signature of Candidate

_____

Acknowledgments

Foremost, I would like to express my deepest gratitude to my supervisor Prof. Jinyan Li for his continuous support to my doctoral study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance has helped me in learning to carry out strong and effective research and in the preparation of this thesis. I could not have imagined having a better mentor and advisor for my doctoral research.

Also, I would like to thank my co-supervisor Prof. Hung Nyugen, Dr. Mengling Feng, Prof. Ramamohanarao Kotagiri, and Prof. Longbing Cao for their continuous support and scientific advice during my research. Without their professional guidance, this thesis would not have been possible.

Additionally, I would like to thank my colleagues, Dr. Qian Liu, Jing Ren, Renghua Song, Yi Zheng, Chaowang Lan, Hui Peng, and Yuangsheng Liu for their strong support and numerous stimulating discussions.

Besides, I offer my regards to all of my co-workers at the Advanced Analytics Institute and Faculty of Engineering and IT, and thank them for their support in the completion of this dissertation.

Shameek Ghosh
December 2017, UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Published

Peer-reviewed Journals

- **Shameek Ghosh**, Jinyan Li, Longbing Cao, Kotagiri Ramamohanarao (2017). Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. **Journal of Biomedical Informatics**, Elsevier, Volume 66, February 2017, Pages 19-31.

- **Shameek Ghosh**, Mengling Feng, Hung Nguyen, Jinyan Li (2016). Hypotension Risk Prediction via Sequential Contrast Patterns of ICU Blood Pressure. **IEEE Journal of Biomedical and Health Informatics**, Volume 20, Issue 5, 2016, pp. 1416-1426.

Peer-reviewed Conferences

- **Shameek Ghosh**, Hung Nguyen, Jinyan Li (2016). Predicting short-term ICU outcomes using a sequential contrast motif based classification framework. *in* Proceedings of IEEE Annual International Conference of the Engineering in Medicine and Biology Society (**EMBC-2016**), pp. 5612-5615. IEEE.

- **Shameek Ghosh**, Jinyan Li (2015), Using sequential patterns as features for classification models to make accurate predictions on ICU events. *in* Proceedings of IEEE Annual International Conference of the Engineering in Medicine and Biology Society (**EMBC-2015**), IEEE.

- **Shameek Ghosh**, Mengling Feng, Hung Nguyen, Jinyan Li (2014). Risk Prediction for Acute Hypotensive Patients by Using Gap Constrained Sequential Contrast Patterns. *in* Proceedings of the American Medical Informatics Association Annual Symposium (**AMIA-2014**), pp. 1748-1757.

- **Shameek Ghosh**, Mengling Feng, Hung Nguyen, Jinyan Li (2014), Predicting heart beats using co-occurring constrained sequential patterns. *in* Proceedings of Computing in Cardiology Conference (**CinC-2014**) pp. 265-268. IEEE.

Abstract

Data mining and knowledge discovery involves efficient search and discovery of patterns in data that are able to describe the underlying complex structure and properties of the corresponding system. To be of practical use, the discovered patterns need to be novel, informative and interpretable. Large-scale unstructured biomedical databases such as electronic health records (EHRs) tend to exacerbate the problem of discovering interesting and useful patterns. Typically, patients in intensive care units (ICUs) require constant monitoring of vital signs. To this purpose, signicant quantities of patient data, coupled with waveform signals are gathered from biosensors and clinical information systems. Subsequently, clinicians face an enormous challenge in the assimilation and interpretation of large volumes of unstructured, multidimensional, noisy and dynamically fluctuating patient data.

The availability of de-identified ICU datasets like the MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care) databases provide an opportunity to advance medical care, by benchmarking algorithms that capture subtle patterns associated with specific medical conditions. Such patterns are able to provide fresh insights into disease dynamics over long time scales.

In this research, we focus on the extraction of computational physiological markers, in the form of relevant medical episodes, event sequences and distinguishing sequential patterns. These interesting patterns known as sequential contrast patterns are combined with patient clinical features to develop powerful clinical prediction models. Later, the clinical models are

used to predict critical ICU events, pertaining to numerous forms of hemodynamic instabilities causing acute hypotension, multiple organ failures, and septic shock events. In the process, we employ novel sequential pattern mining methodologies for the structured analysis of large-scale ICU datasets. The reported algorithms use a discretised representation such as symbolic aggregate approximation for the analysis of physiological time series data. Thus, symbolic sequences are used to abstract physiological signals, facilitating the development of efficient sequential contrast mining algorithms to extract high risk patterns and then risk stratify patient populations, based on specic clinical inclusion criteria.

Chapter 2 thoroughly reviews the pattern mining research literature relating to frequent sequential patterns, emerging and contrast patterns, and temporal patterns along with their applications in clinical informatics.

In Chapter 3, we incorporate a contrast pattern mining algorithm to extract informative sequential contrast patterns from hemodynamic data, for the prediction of critical care events like Acute Hypotension Episodes (AHEs). The proposed technique extracts a set of distinguishing sequential patterns to predict the occurrence of an AHE in a future time window, following the passage of a user-defined gap interval. The method demonstrates that sequential contrast patterns are useful as potential physiological biomarkers for building optimal patient risk stratification systems and for further clinical investigation of interesting patterns in critical care patients.

Chapter 4 reports a generic two stage sequential patterns based classification framework, which is used to classify critical patient events including hypotension and patient mortality, using contrast patterns. Here, extracted sequential patterns undergo transformation to construct binary valued and frequency based feature vectors for developing critical care classification models.

Chapter 5 proposes a novel machine learning approach using sequential contrast patterns for the early prediction of septic shock. The approach combines highly informative sequential patterns extracted from multiple phys-

iological variables and captures the interactions among these patterns via Coupled Hidden Markov Models (CHMM). Our results demonstrate a strong competitive accuracy in the predictions, especially when the interactions between the multiple physiological variables are accounted for using multivariate coupled sequential models. The novelty of the approach stems from the integration of sequence-based physiological pattern markers with the sequential CHMM to learn dynamic physiological behavior as well as from the coupling of such patterns to build powerful risk stratification models for septic shock patients.

All of the described methods have been tested and bench-marked using numerous real world critical care datasets from the MIMIC-II database. The results from these experiments show that multivariate sequential contrast patterns based coupled models are highly effective and are able to improve the state-of-the-art in the design of patient risk prediction systems in critical care settings.

Chapter 1

Introduction

1.1 Background

Sequences exist everywhere in our daily life. In their simplest logical structure, a sequence can be described as an enumerated collection of objects, where repetitions are allowed. Similar to a set, it consists of members (also called elements, or terms). The cardinality of ordered elements in the corresponding set is called the length of the sequence. However, unlike a normal set, an ordered collection of objects consists of a set of objects where the sequential order of the objects or members hold importance. This means that the same elements can also appear multiple times at different positions in the sequence.

There exist numerous real world applications that use sequential data. Typical examples include clickstream logs, consumer shopping sequences, DNA sequences, share price sequences of a company, sequence of medications taken by a patient and so on. An essential aspect of discovering interesting sequences is related to determining domain specific events that occur in an order, which may correspond to uncovering interesting behaviour of the underlying system or agents in concern. We illustrate the importance of sequences using two specific examples in detail, as given below.

The first case is the customer shopping sequence, as shown in Table 1.1.

Table 1.1: Transaction Data Table

Tid	Transaction Time	Customer ID	Items	Quantities	Profit
T1	11-11-2014 10:00:00	C1	45	1	$10.50
T2	11-11-2014 10:01:05	C2	30,31,32	2,3,1	$5.20, $2.00, $3.00
T3	11-11-2014 10:02:12	C3	29,16	1,2	$7.00, $5.00
T4	11-11-2014 10:03:16	C1	28	6	$2.80
T5	12-11-2014 10:04:35	C5	45	2	$10.50
..	..	..	..	..	..
T3465	11-11-2014 18:00:00	C3	22,32	2	$1.00,$3.00

As a toy example, the table is from a retail stores database which contains customers transactions records. The first column contains IDs that are assigned to the corresponding transactions. The second column contains the time stamps for transactions. Users who purchased by store membership card or credit card are recorded in the third column. The last three columns record the items which were purchased, the quantity of items and their respective unit profits.

Thus, each row in the table can be viewed as a customer-purchased basket of objects. Moreover, a customer will not just shop only once (one transaction is one row in Table 1.1 in the retail store. Rather, he or she may shop multiple times a day. For example, the transactions of customer C1 and C3 can be viewed as two sequences, i.e. $< T1, T4 >$ and $< T3, T3465 >$ respectively. It is also understandable that when transactions are analyzed over a longer period of time such as "all transactions in the month of January", the sequence of transactions for each customer (for example, C1 and C2) in "the month of January" would be longer versus "all transactions in a day".

To improve profits and productivity, the job of a manager in a retail store is to improve the turnover and revenue of the retail business. In this context, a use case may involve mining of customer buying behaviour. For example, users would generally buy CD-ROMS, digital cameras following the purchase of a computer. However, a retail store would require knowledge of all such frequently occurring sequence of transactions to advertise a specific prod-

uct to a given customer for maximizing the probability of purchase. Thus, the retail store management is required to discover customers frequent shopping habit sequences, and activate the most appropriate sales and promotion strategies at the right time. Accordingly, the retail manager will probably look into the shopping histories of customers, and be presented with specific sequential patterns in a customer's buying habits that influences their shopping behaviors on a regular basis. Such sequential patterns occurring among a population of customers help design marketing strategies to match the customers needs, seasonal sales planning, and thus improve productivity as well as company profits. Consequently, revenue is improved.

The second case is that of an online shopping website. Nowadays, e-commence websites such as Amazon.com and Groupon.com are increasing becoming very popular. People tend to buy things online rather than go to a physical store because of the convenience, variety, low price and many other advantages. These websites, however, have to deal with a great number of accesses every day.

Table 1.2: Web Access logs

user_id	session_id	timestamp	referring_url	page_url	action
100	1	23-10-2014 12:05:00	www.twitter.com?user_id=ABC	www.groupon.com/view skydiving	View
100	1	23-10-2014 12:05:15	.....	www.groupon.com/purchase_skydiving	Checkout
100	1	23-10-2014 12:06:45	.....	www.groupon.com/purchase_complete	Purchase
200	1	23-10-2014 11:35:00	www.facebook.com?user_id=XYZ	www.groupon.com/view_skydiving	View
200	1	23-10-2014 11:35:30	. . .	www.groupon.com/purchase_skydiving	View
200	2	23-10-2014 12:10:05	www.facebook.com?user_id=XYZ	www.groupon.com/view_yoga	View

One of the backend jobs is to record the customer behaviors such as clicks and scrolls to a backend web log database, as shown in Table 1.2. Each row in Table 1.2 represents an action of a user: when, where, what and how. Thus, a single users behaviors are elements of a sequence. For example, user id = 100 probably noticed the skydiving promotion advertisements on Twitter and wanted to use the opportunity to experience skydiving. The user directly clicked the link and purchased this promotional offer. All these actions are captured by Groupons servers behind the web pages, and then stored in their web log databases. There are millions of such users online

every day, which means the same number of sequences in the databases are recorded. As time passes, not only do the sequences get longer, but new sequences are also added.

Website data analysts are keen to know which items are most related to others. With this knowledge, they can accurately recommend items to online users. As an example, "people who buy this item also buy A, B and C" is often seen in Amazon, and many users eventually purchase those recommended things which they did not originally plan to buy. It is definitely important for analysts to review and discover patterns in user behaviors to ensure the precision of their recommendations.

1.2 Mining of Sequential Patterns

In the 1990s, mathematicians, statisticians, and computer scientists proposed Knowledge Discovery and Data mining (KDD), which involves using a range of models, algorithms and tools to analyze various types of data. In the academia, groups of researchers are interested in finding patterns in the transactions, sequences and graphs, etc.

The specific areas of frequent patterns and sequential pattern mining are highly relevant to the topic in this thesis. In frequent pattern mining, the frequently repeated sub-itemsets in a transaction database are discovered as patterns. It was first proposed in the work by Rakesh Agrawal et al (1993), in which the renowned downward closure property (also named the Apriori Property) was introduced. With the foundation of the frequency based mining algorithms (namely, downward closure property), many followup papers were subsequently published. For example, Park et al. propose an effective hash-based algorithm for the candidate set generation (Park, Chen & Yu 1995) . Savasere et al. presented an algorithm reducing both CPU and I/O overheads by applying partition techniques (Savasere, Omiecinski & Navathe 1995). Several works (Agrawal & Shafer 1996, Cheung, Han, Ng, Fu & Fu 1996) use parallel and distributed techniques in the area of association

rule mining. An incremental approach is discussed in (Cheung et al. 1996), and sampling methods are proposed in (Toivonen et al. 1996).

Later, sequential pattern mining has been popular since its introduction by Agrawal and Srikant (1995). In this work, the sequential pattern mining was defined as follows:

"Given a database of sequences, where each sequence consists of a list of transactions ordered by transaction time and each transaction is a set of items, sequential pattern mining is to discover all sequential patterns with a user-specified minimum support, where the support of a pattern is the number of data sequences that contain the pattern."

For simplicity, it can be said that sequential pattern mining seeks to discover frequent subsequences as patterns in a sequence database (Han, Pei, Mortazavi-Asl, Pinto, Chen, Dayal & Hsu 2001).

In the first case in Section 1.1, item 45 and item 32 both appear twice in different customers transactions (C1 and C5 have 45, C2 and C3 have 32), which makes support for these items higher than for any other items. If the minimum support (a threshold to filter infrequent sequential patterns, and retain frequent ones) is set to 2, then $< 45 >$ and $< 32 >$ are two frequent sequential patterns. Sequential pattern mining has proven to be essential for handling order based critical business problems. For retail data, sequential patterns are useful for shelf placement and promotions, as the first case in 1.1. In the industry, sequential patterns are used for targeted marketing, customer retention, and many other tasks. Other areas in which sequential patterns can be applied include web access pattern analysis, weather prediction, production processes, and network intrusion detection. Note that most studies of sequential pattern mining concentrate on categorical (or symbolic) patterns,whereas studies on numerical curve analysis usually belong to the scope of trend analysis and forecasting in statistical time-series analysis.

In the last two decades, data mining researchers have proposed many techniques and algorithms for mining sequential patterns. For instance, GSP (Srikant & Agrawal 1996) uses a Generating-Pruning method and makes mul-

tiple passes over the data to target the patterns; SPADE (Zaki 2001) builds an ID-list for each candidate, and joins two k-candidates to generate a new (k + 1)-candidate; PrefixSpan (Han et al. 2001) extends the pattern growth approach in the FP-Growth algorithm (Han, Pei & Yin 2000) for frequent sequential pattern mining; CloSpan (Yan, Han & Afshar 2003) proposes an efficient algorithm for mining closed sequential patterns; SPAM (Ayres, Flannick, Gehrke & Yiu 2002) presents a bitmap representation of the original sequence database, and proposes pruning methods for the I-Step/S-Step extensions; PAID (Yang, Kitsuregawa & Wang 2006)and LAPIN (Yang, Wang & Kitsuregawa 2007) use an item-last-position list and prefix border position set instead of the tree projection or candidate generate-and-test techniques introduced so far; DISC-all (Chiu, Wu & Chen 2004) prunes infrequent sequences according to other sequences of the same length, and employs lexicographical ordering and temporal ordering. FreeSpan (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000) starts by creating a list of frequent 1-sequences from the sequence database called the frequent item list (f-list), and then constructs a lower triangular matrix of the items in this list. Moreover, there have been two thorough surveys of the sequential pattern mining algorithms (Mabroukeh & Ezeife 2010, **?**).

1.3 Mining Useful Patterns for Critical Care Decision-support

As described in Section 1.2, abundant literature has been dedicated to research in frequent sequential patterns and tremendous progress has been made, which include efficient and scalable algorithms in various domains. Yet the mining of interesting patterns of various underlying complex structures demanded by domains in medicine, open up multiple challenges for medical data mining that still remain unsolved. In particular, with the advent of large-scale biomedical databases, exciting opportunities have opened up in the areas of sequential pattern mining.

Biomedical databases can be categorized into multiple types, which may store microarray gene expression data, protein sequences or electronic health records. Even though a huge amount of pattern mining research has gained ground in bioinformatics, the area of healthcare analytics has comparatively been slow in the adoption of pattern mining techniques. The slow pace of healthcare analytics also suggests the availability of problems that have been difficult to solve traditionally and hence not been tackled much.

1.3.1 Problem Statement

Today, most clinicians across the world, continue to practice the traditional process of trial-and-error medicine. Accordingly, when a patient presents with symptoms, the doctor makes a most likely diagnosis, then prescribes a drug and, then a treatment recommendation. To help a clinician in these activities, the most popular diagnostic tool-kits frequently make use of population based scoring techniques. However, such scoring systems seldom take into account dynamically changing symptoms or events in a patient's medical history. As a result, a significant percentage of diagnoses carried out across the world, lead to slower discovery of a patient's true ailment leading to delayed treatment, which has consistently been adding to health-care costs across numerous countries. Given the premise that a patient's ailment can be caused by numerous static and dynamic clinical factors, current scoring tool-kits used by clinicians require sophisticated improvements for consuming large-scale patient data and make dynamic predictions about the patient state.

The current thesis is motivated by a need to develop predictive systems that account for fast changing fluctuations in a patient's physiological condition for personalized medicine. Hence, the long term aim of such methods are to assist doctors who use the personalized medicine approach to take into account the patients unique physiology.

1.3.2 Research Motivations

Raw EHR (electronic health records) data, in particular, when critically analysed, can help extract important patient information and help develop a map of the patients history, which can aid the diagnostic process used by the hospital and the physician for improving patient care. Core analysis of medical data is essential in multiple departments of a hospital or health care centres.

With the advent of complex healthcare systems generating massive data, there is a specific dearth of tools and techniques that can quantitatively support the fast analysis of complex, high-frequency data streams emanating out of such environments. Yet, accessibility of such medical databases for widespread research has been comparatively restricted owing to multiple procedural reasons. In recent years, there have been many efforts worldwide to provide access to such databases as part of collaborative research. These include:

- the Stanford Translational Research Integrated Database Environment - STRIDE (Lowe, Ferris, Hernandez, Weber et al. 2009)
- the Australian and New Zealand Intensive Care Society Adult Patient Database (Stow, Hart, Higlett, George, Herkes, McWilliam, Bellomo, Committee et al. 2006)
- the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database (Saeed, Villarroel, Reisner, Clifford, Lehman, Moody, Heldt, Kyaw, Moody & Mark 2011)

Among these the MIMIC-II is available free of charge for public use, on completion of an online training course and signing of a data use agreement. MIMIC-II is available via PhysioNet, which is an online resource for the study of physiological data and shares numerous problems for computational physiology.

Critical care databases like the MIMIC-II provide an excellent avenue for carrying out data mining research. It can thus be pointed out that the

big data generated in intensive care units (ICUs) have a massive potential to usher in novel clinical discoveries, leading to the development of early warning, detection and prevention systems in a wide range of serious patient conditions.

Patients who are critically ill require intensive care for their survival. Typically, an intensive care unit (ICU) is a department where a patient's vital functions are regularly monitored, with mechanical support or medication, until the patient regains his or her normal functional abilities again. The majority of critical care patients require such ICU care for only a few days, which may consequently result in a high chance of survival. Some patients may continue in the ICU for longer periods, with their likelihood of mortality increasing as the stay period increases. Present risk prediction models in critical care (Rosenberg 2002) may be used to compare the risk of mortality or severity of disease in patient populations, but are considered sub-optimal in predicting the probability of survival of individual patients. As such, there exist no tools which can reliably predict an individual patient's chance of developing a complication in the future, such as organ failure.

Thus, it is crucial to be able to detect clinical problems early enough, so that preventive or curative treatments can be applied on time. In practice, an intensivist analyses all the patient related data in order to foresee a change in the patient's condition and administer the appropriate treatment. Since humans are not able to simultaneously deal with more than 5 to 7 different parameters and an average ICU patient is estimated to be described by more than 250 different parameters, it is likely that there is more information in the data than what is currently being extracted from it by humans. Accordingly, data mining could assist clinicians by analysing the ICU data and detecting problems earlier than an experienced intensivist would, and could also be used to generate models that would assist the intensivist in deciding for the best treatment for a specific clinical problem.

Hence, with an abundance of rich ICU data, clinical data mining systems can add significant benefits by extracting useful information from these rich

databases, in comparison and by complementing traditionally used scoring systems. Based on outcomes of such knowledge discoveries, the benefits could be wide and large-scale impacting both hospital administrative and patient financial decisions. Hence, the proposed research is motivated by a need for discovering dynamic interaction of physiological events and the role these events play in informing patient morbidity and mortality.

1.4 Limitations and Challenges

Although sequential pattern mining algorithms successfully extract patterns from the sequence databases, their primary interestingness measurement is the frequency of a pattern. This means, any frequent sequential pattern is treated as a significant one. However, in clinical practice, most frequent sequential patterns are not useful and informative for clinical decision-making, since they do not have clinical value, and can be spurious in nature. Particularly, clinicians are interested in patterns that are prominent in an intervention population given a specific treatment and are strong indicators of a disease risk. In most clinical scenarios, truly interesting sequences may be filtered out because of their low frequency.

Methods in emerging pattern mining tend to address this problem by considering the growth rate of a pattern as an interestingness measure (Dong & Li 1999). In this framework, emerging patterns were defined as itemsets whose support increases significantly from one data set to another. Emerging patterns are said to capture emerging trends in time-stamped databases, or to capture differentiating characteristics between classes of data. When first defined by Dong and Li (1999), the purpose of emerging patterns was to capture emerging trends in time-stamped data, or useful contrasts between data classes. Subsequent emerging pattern research has largely focused on the use of the discovered patterns for classification purposes, for example, classification by emerging patterns (Dong & Li 1999, Li, Ramamohanarao & Dong 2000) and classification by jumping emerging patterns. An advanced

Bayesian approach (Fan & Ramamohanarao 2003) and bagging (Fan, Fan, Ramamohanarao & Liu 2006) were also proposed.

The quality measure of emerging patterns is the growth rate (the ratio of the two supports). It determines, for example, that a pattern with a 10% support in one data set and 1% in the other is better than a pattern with support 70% in one data set and 10% in the other (as $10/1 > 70/10$). Further, Fan and Ramamohanarao (2003) had worked on selecting the interesting emerging patterns, while Soulet et al. (2004) had proposed condensed representations of emerging patterns. Later a CART-based approach was used to discover emerging patterns in microarray data (Boulesteix, Tutz & Strimmer 2003). The method is based on growing decision trees from which the emerging patterns are extracted.

Specially in the field of healthcare, there have been studies in the last few years, demonstrating the usefulness of sequence based predictive models. Previously, Toma et al. (2007) showed a data driven technique to discover temporal episodes of organ failure scores to predict patient mortality. Noren et al. (2010) proposed a statistical model which summarized the temporal associations, medical events and drug prescriptions. Temporal abstractions for time interval mining were employed as a set of features inspired by the Bag-of-Words approach (Moskovitch & Shahar 2009). More recently, frequent sequential patterns were identified using a patient populations among using interactive and visual discovery (Gotz, Wang & Perer 2014). EEG and EMG time series recordings were employed by pattern mining algorithms using a binning step for facilitating the discovery of high quality patterns (Skapura & Dong 2015). Patient clinical factors like fluid balance evolution during the first days was used by temporal data mining for a patient's survivability (Casanova, Campos, Juarez, Fernandez-Fernandez-Arroyo & Lorente 2015). Further, electronic healthcare reimbursement claims have been used to extract frequent sequential patterns to analyze healthcare delivery and practice patterns across the United States (US) (Malhotra, Hobson, Valkova, Pullum & Ramanathan 2015). As evident, a significant proportion of these recent

studies focus on extracting frequent sequential patterns from patient populations for building prediction models.

However, in real-time clinical data mining problems, simplified applications of frequent patterns do not tend to be discriminative and predictive all the time. Instead, advanced methodologies are required which are able to integrate highly discriminative patterns, using smart pattern transformation techniques to build models that are both interpretive and highly predictive.

In addition, an interesting aspect of medical patterns is their tendency to co-occur while in a progressive state. This means meta-information about patterns that can integrate them using sequence ordering are required to improve the state-of-the-art in clinical informatics.

The applicability of contrast pattern mining is intuitive from a medical data mining perspective due to its quality measure being relatively synonymous with the concept of odds ratio (a popular measure used by clinicians for carrying out clinical trials). By their virtue of using a constrained pattern mining approach while comparing an intervention and control population, contrast pattern mining tends to be more suitable for finding patterns that are able to distinguish between two patient populations. However, there have been comparatively limited applications and extensions of methods in contrast mining to develop novel systems for critical care.

1.5 Research Issues

The objectives of this research are therefore related to mining interesting sequences of events (episodes) that are predictive of future critical conditions in patients admitted to intensive care units (ICUs). A sequential pattern mining approach is important and useful for various reasons such as :

- to identify dynamic variations in patient physiological patterns
- in clustering of treatment plans based on similar sequential patterns in ICU patients

- to help in mining of abnormal events or specific episodes for various diseases towards personalized patient care.

- Informative and discriminative features from sequential patterns are useful for classification and forecasting of important clinical outcomes in ICUs

Thus, an evolving set of interesting events would have the ability to provide an excellent form of interpretive and descriptive knowledge to physicians.

1.6 Research Contributions

The main contributions of this thesis are related to it's advancement of the state-of-the-art in critical care data mining models as described below. In this context, our works provide evidence that short term critical care event prediction systems can greatly be improved by the use of sequential contrast patterns that are able to capture dynamic fluctuations within a patient's physiological variables.

For prediction of acute hypotension in the ICU, we incorporate a sequential contrast mining methodology on discretised representations of patient mean arterial pressure to extract distinguishing sequential patterns. These discriminative hemodynamic sequential patterns are used to classify and risk stratify ICU patients. In the process, we demonstrate that using sequential patterns that are able to contrast between two population sub-groups to predict patient risk of critical events provide high performances.

This thesis also builds on a generalized pattern based classification framework, which automatically generates sequential contrast patterns as features from multivariate physiological time series to predict short term ICU events. To this purpose, informative contrast sequential patterns from clinical variables like mean arterial pressure and respiratory rate time series are transformed to a feature space using two mapping methods. Each pattern mapping method creates a new feature space involving binary valued attributes

and frequency based information, to predict ICU events like the onset of a future acute hypotensive episode (AHE) and patient mortality. Thus, the integration of these mapping methods involve employing sequential contrast patterns to define features for building generalized clinical classification models.

Finally, we propose the integration of multivariate sequential contrast patterns (SCPs) using Coupled Hidden Markov Models for septic shock event predictions in an ICU. To integrate SCPs with CHMMs, a novel transformation approach is proposed where the original patient sequence is transformed to a sequence of SCPs for each patient. Our experiments show that our framework is able to learn accurate event detection classifiers for real-world clinical tasks, which is a key step for developing intelligent clinical monitoring systems. In the process, we extend the idea of contrast patterns based classification methods to important problems in the clinical temporal domains.

1.7 Thesis Structure

The thesis is structured as follow:

Chapter 2 outlines the related research in the fields of frequent and sequential pattern mining. This includes providing a broad coverage of frequent sequential and emerging pattern mining, which is related to the topic of this thesis. Relevant works in temporal pattern mining are also described. Following this, we report about specific applications of pattern mining in critical care that aligns with the current research.

Chapter 3 presents our sequential contrast pattern mining methodology for mining discriminative sequential predictive patterns in the field of critical care informatics. A controlled methodology is described to extract patient populations which are treated as intervention and control sets. Discriminative patterns are then used by a majority voting technique to generate predictive alerts for acute hypotensive episodes in an ICU. It also presents our experimental evaluations on real-world EHR datasets while comparing

against other popular machine learning approaches.

Chapter 4 describes generalised pattern based classification approaches by transforming sequential contrast patterns into a feature space and using standard machine learning algorithms like SVM and Naive Bayes to build patient risk prediction models for critical events involving patient mortality. Our results on multiple critical care datasets demonstrate competitive performances.

Finally, Chapter 5 proposes a novel machine learning framework where multivariate sequential contrast patterns are combined using coupled hidden markov models (CHMM) to predict septic shock events. The approach transforms the original patient sequences to form a time series of sequential contrast patterns. These multivariate pattern sequences are then used for learning univariate and multivariate hidden markov models. Our experimental results indicate that ordering patterns to form a sequence can have strong predictive value while learning multivariate and coupled machine learning models.

Chapter 6 concludes the thesis and outlines the scope for future work.

Chapter 2

Literature Review

In this chapter, we first introduce the traditional frequent pattern mining framework, which contains sequential pattern mining. Later, we introduce the emerging and contrast pattern mining framework, which contains an overview of the research so far. Following this, we introduce prior research and the relevant frameworks used for mining temporal patterns. Finally, we discuss previous implementations and applications of pattern mining algorithms for problems in critical care clinical informatics.

2.1 Frequent Pattern Mining Framework

Frequent patterns are patterns that appear a considerable number of times in a dataset. These patterns can exist in a variety of formats such as:

- Itemset patterns: Representing a set of items (Agrawal & Swami 1993, Cheng, Yan, Han & Hsu 2007).
- Sequential patterns: Representing sequence based and temporal order among items (Srikant & Agrawal 1996, Zaki 2001, Han et al. 2001, Wang & Han 2004).
- Time interval patterns: Represent temporal relations among states with time durations (Höppner 2003, Papapetrou, Kollios, Sclaroff &

Gunopulos 2005, Winarko & Roddick 2007, Moerchen 2006, Batal, Sacchi, Bellazzi & Hauskrecht 2009, Mörchen & Fradkin 2010, Batal, Valizadegan, Cooper & Hauskrecht 2011).

Frequent pattern mining plays an essential role in the discovery and extraction of interesting regularities that appear in data. It was first proposed by (Agrawal, Imieliński & Swami 1993) to mine association rules for market basket datasets. Since then, abundant literature has been dedicated to this field and tremendous progress has been made. To this purpose, the objective was to analyze customer buying habits by discovering associations between items that customers frequently buy together. As an example, if a customer buys cereal, he is also likely to buy milk on the same trip to the supermarket. Here, cereal and milk are called items and the customers trip to the supermarket can be termed as a transaction.

Formally, let $\Sigma = I_1, I_2, ..., I_n$ denotes the set of all items, also known as the alphabet. An itemset pattern is a conjunction of items: $P = I_{q1} \wedge ... \wedge I_{qk}$ where $I_{qj} \epsilon \Sigma$. If a pattern contains k items, we call it a k-pattern (an item is a 1-pattern). We say that pattern P is a sub-pattern of pattern P0 (P0 is a super-pattern of P), denoted as $P \subset P_0$, if every item in P is contained in P0. The support of pattern P in database D, denoted as sup(P,D), is the number of instances in D that contain P. Accordingly, given a user specified minimum support threshold σ, we say that P is frequent pattern if $sup(P, D) \geq \sigma$.

Example 1. Given below is the transaction data in Table 2.1. Here, the alphabet of items is $\Sigma = \{A, B, C, D, E\}$ and there exist 5 transactions $T_1 to T_5$ (each representing a customer visit). Note that pattern $P = A \wedge C$ appears in transactions T_1, T_2 *and* T_4, hence the support of P is 3. If we set the minimum support $\sigma = 3$, then the frequent patterns for this example are: $\{A, C, E, A \wedge C\}$.

The original pattern mining framework was used to mine transaction data. However, similar concepts can be applied to relational attribute-value data, such that each instance is described by a fixed number of attributes such as the data in Table 2.2.

Table 2.1: Transaction Data Example

Transaction	List of items
T_1	A,C,D, E
T_2	A,B,C
T_3	A,D,E
T_4	A,C,E
T_5	E

Table 2.2: Relational Attribute Data Example

Age	Education	Marital status	Income
Young (≤ 30)	Bachelor	Single	Low ($\leq 50k$)
Middle Age ($30-60$)	Masters	Married	Low ($\leq 50k$)
Middle Age ($30-60$)	Bachelor	Married	Medium ($50k-100k$)
Senior (≥ 60)	PhD	Married	High ($\geq 100k$)

Attribute-value data is converted to an equivalent transaction data so that the data is discrete. This means the data should contain only categorical attributes. Here, each attribute-value pair is mapped to a distinct item. When the data contain numerical (continuous) attributes, these attributes should be discretized (Yang et al. 2006). For example, the age attribute in Table 2.2 has been converted into three discrete values: Young, Middle age and Senior.

Table 2.3 shows the data in Table 2.2 in transaction format. Here, converting an attribute-value data to a transaction data format ensures that transactions are having the same number of objects (unless the original data contained missing values). Following this transformation, pattern mining algorithms are applied on the equivalent transaction data.

Typically, pattern mining is challenging since the search space of patterns tends to be very large. For instance, the search space of all possible itemset patterns for transaction data is exponential in the number of items. So if Σ

Table 2.3: Table 2.2 as a set of Transactions

Transaction	List of items
T_1	Age=Young, Education=Bachelor, Marital Status=Single, Income=Low
T_2	Age=Middle age, Education=Masters, Marital Status=Married, Income=Low
T_3	Age=Middle age, Education=Bachelor, Marital Status=Married, Income=Medium
T_4	Age=Senior, Education=PhD, Marital Status=Married, Income=High

is the alphabet of items, there are $2^{|\Sigma|}$ possible itemsets (all possible subsets of items). The search space of itemset patterns for attribute-value data is exponential in the number of attributes. So if there are d attributes and each attribute takes V possible values, there are $(V + 1)^d$ valid itemsets. Note that the search space for more complex patterns, such as sequential patterns, graph patterns, or time interval patterns , is even larger than the search space for itemsets. Thus, the naive method to generate and count all possible patterns is not feasible. Frequent pattern mining algorithms make use of the minimum support threshold to restrict the search space to a reasonable subspace that can be explored more efficiently.

2.1.1 Sequential pattern mining

Frequent sequential pattern mining refers to the discovery of frequent subsequences as patterns in a sequence database. A sequence database consists of sequences which are ordered list of elements, and each element can be either an itemset or a single item. Such databases are quite common and widely used; for example, in customer shopping sequences, web clickstreams and bio-logical sequences. The formal definition of frequent sequential pattern mining is defined below.

Let $I = \{i_1, i_2, ..., i_n\}$ be a set of items. A sequence is defined as $s =< e_1, e_2, ..., e_m >$ where $e_k \subset I$, $ek = \phi$, $1 \leq k \leq m$. Without loss of generality, we assume that the items in each itemset are sorted in a certain order (such as alphabetical order). A sequence database is defined as $D = [sid_1, s_1], [sid_2, s_2], ..., [sid_l, s_l]$. The sid is the unique identification of

the corresponding sequence. A sequence $\alpha =< a_1, a_2, ..., a_p >$ is called a subsequence of another sequence $\beta =< b_1, b_2, ..., b_q >$, denoted by $\alpha \subset \beta$, if and only if $\ni j_1, j_2, ..., j_p$, such that $1 < j_1 < j_2 < ... < j_p \leq n$ and $a_1 \subset b_{j1}, a_2 \subset b_{j2}, ..., a_p \subset b_{jp}$. We also call β the supersequence of α, or β contains α. Given a sequence database D, the support of α is the number of sequences in D which contain α. If the support α satisfies a minimum support threshold, α is a frequent sequential pattern.

For example, we assume the itemset I sold in some retail stores is as follows.

$I = \{bread, milk, cheese, butter, cereal, oatmeal\}$

Table 2.4: Sequence Database

sid	tid	transactions
1	1	bread, buter, cereal
1	2	milk, cheese, oatmeal
1	3	bread, butter
2	1	cheese, butter
2	2	bread, milk, cheese, oatmeal
2	3	milk
3	1	bread, cheese, butter
3	1	bread, milk, oatmeal

A toy sequence database D with I would be as shown in Table 2.4. The database consists of three sequences, which represent the shopping histories of three customers. Both sequence $sid = 1$ and $sid = 2$ contain 3 itemsets (transactions), and $sid = 3$ contains 3 itemsets. Equally, D in Table 2.4 can be written as:

$s1 =< (bread, butter, cereal)(milk, cheese, oatmeal)(bread, butter) >$
$s2 =< (cheese, butter)(bread, milk, cheese, oatmeal)milk >$
$s3 =< (bread, cheese, butter)(bread, milk, oatmeal)$

Speaking of the containment relationship, $< butter(bread, milk) >$ can be a subsequence of $s2$ and $s3$ but not $s1$. Similarly, $< buttercheese >$ can be a subsequence of $s1$ and $s2$ but not $s3$. Containment relationships for subsequences turn out to be extremely important in fields like bioinformatics, where sequences may contain a lot of garbage characters, and the informative part may be hidden within them. Being able to discover frequent subsequences would allow to get rid of the uninformative symbols in DNA sequences for example.

Quite a few algorithms have been proposed since it was first introduced in (Agrawal & Srikant 1995). For instance, GSP (Srikant & Agrawal 1996) uses a "Generating-Pruning" method and makes multiple passes over the data to extract the patterns. SPADE (Zaki 2001) builds an ID-list for each candidate, and joins two k-candidates to generate a new (k + 1)-candidate. PrefixSpan (Han et al. 2001) extends the pattern-growth approach in FPGrowth algorithm (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000) for frequent sequential pattern mining. CloSpan (Yan et al. 2003) proposes an efficient algorithm for mining closed sequential patterns. SPAM (Ayres et al. 2002) presents a bitmap representation of the original sequence database, and proposes pruning methods for the I-StepS-Step extensions. PAID (Yang et al. 2006) and LAPIN (Yang et al. 2007) use an item-last-position list and prefix border position set instead of the tree projection or candidate generate-and-test techniques introduced so far. DISC-all (Chiu et al. 2004) prunes infrequent sequences according to other sequences of the same length, and employs lexicographical ordering and temporal ordering. FreeSpan (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000) starts by creating a list of frequent 1-sequences from the sequence database called the frequent item list (f-list), and then constructs a lower triangular matrix of the items in this list. All of the above algorithms rely on the downward closure property. Next, we briefly introduce the most popularly used algorithms as reported above.

AprioriAll

AprioriAll (Agrawal & Srikant 1995) is believed to be the first algorithm solve sequential pattern mining. First, it finds all frequent 1-patterns whose support values satisfy a user-defined minimum support. Then, it initializes and maintains two types of list containers, namely the candidate lists and the frequent pattern lists. For every (k + 1)-candidate constructed by joining two frequent k-patterns (the patterns with k items in the frequent pattern list), the support needs to be scanned from the original database. The process repeats until no further patterns can be found.

GSP

GSP (Generalized Sequential Patterns) (Srikant & Agrawal 1996) is a sequential pattern mining method that was developed by Srikant and Agrawal in 1996 and has been very popular since then. It is an extension of the Apriori algorithm (Agrawal & Srikant 1995) for sequence mining. The main structure is similar to AprioriAll (Agrawal & Srikant 1995) , and the details are as follows. First, it scans the database to obtain the frequent 1-sequences. Then it generates the next level candidates by joining the previous level frequent sequences, the same as AprioriAll. The differences are in the candidate generation and candidate support counting. In the candidate generation stage, they use a mechanism to prune the unpromising candidates. Thus in the same level (candidates of the same length), the number of candidates is no more than that of AprioriAll. In the support counting stage, a hash-tree data structure is used to reduce the number of candidates to be checked. The representation of the database is transformed to efficiently determine whether a specific candidate is contained in the database.

SPADE

SPADE (Sequential PAttern Discovery using Equivalent classes) (Zaki 2001) is also a level-wise sequential pattern mining algorithm that uses a vertical data format. The key difference between SPADE and AprioriAll (Agrawal & Srikant 1995) builds an ID-list(a list of the IDs of sequences and

elements) for each candidate. The support count of the candidate can be easily calculated from its ID-list, which greatly reduces the cost of scanning. Because of this, SPADE outperforms GSP to a large extent according to authors experimental results.

FreeSpan

FreeSpan (Frequent pattern-projected Sequential pattern mining) (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000) is the first projection-based depth-first algorithm proposed by Han et al. in 2000. Similar to the previous algorithms, FreeSpan scans the database once to obtain the frequent 1-sequences and put them in the f-list(frequent item list). Then it constructs a matrix called S-Matrix which contains the 2-sequences and their supports generated from the f-list, and the infrequent ones are filtered. Each sequential pattern in the S-Matrix corresponds to a projected database that all the sequences contain the sequential pattern itself. The next step is to construct level-2-sequences from the S-Matrix and find annotations for repeating items and projected databases in order to discard the matrix and generate level-3 projected databases. The process repeats until no candidates can be generated.

SPAM

SPAM (Sequential PAttern Mining) (Ayres et al. 2002) is a depth-first algorithm that integrates the ideas of GSP (Srikant & Agrawal 1996), SPADE (Zaki 2001) and FreeSpan (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000). A group of novel concepts such as the sequence-extension step (S-Step), itemset-extension step (I-Step) and the lexicographical tree are firstly introduced. Similar to FreeSpan, SPAM uses a depth-first strategy to traverse the lexicographical tree to extract the complete set of frequent sequential patterns. More importantly, SPAM encodes the ID-list from SPADE to a vertical bitmap data structure and puts them in the memory so that the 'joining' operation between two ID-lists is extremely fast. That is the key

reason why SPAM outperforms any of the previous algorithms.

PrefixSpan

PrefixSpan (Prefix-projected Sequential pattern mining) (Han et al. 2001) is an algorithm that extends the pattern-growth approach for frequent pattern mining and the first algorithm that does not generate a candidate. As an enhanced algorithm of FreeSpan (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000), PrefixSpan uses the "prefix" of the sequence to project the database. Then it scans the projected database for the items to be concatenated to the prefix, and counts the support for each item. The infrequent concatenation items will be discarded, and frequent items will be retained. Lastly, for each frequent concatenation item, a new prefix and its corresponding smaller projected database can be constructed. The process continues until no more frequent concatenation items can be scanned. In experimental results, PrefixSpan performs much better than both GSP and FreeSpan. The major cost of PrefixSpan is the construction of projected databases.

PAID and LAPIN

PAID (PAssed Item Deduced sequential pattern mining) (Yang et al. 2006) and LAPIN (LAst Position INduction sequential pattern mining) (Yang et al. 2007) essentially follow pattern-growth algorithms such as FreeSpan (Han, Pei, Mortazavi-Asl, Chen, Dayal & Hsu 2000) and PrefixSpan (Han et al. 2001). The main contribution of PAID is that it adopts a novel strategy to reduce the scanning cost. The technical detail is as follows. In a prefix-sequence projection, the last position (the itemset number) of an item can be used to judge whether or not the item can be extended to the current prefix. For instance, $s0 =< (ab) >$ is contained in $s1 =< (ab)a(cd)ea >$, $s2 =< (ab)(ae) >$ and $s3 =< (abc)aea >$. Since the last position of a in $s1$ is 5 (the fifth itemset contains a, similarly 2 in $s2$ and $s3$), there is no need to scan the sequences to obtain a. Instead, PAID only needs to compare the projection positions with the last positions of a in the three sequences. That

is a simple example to explain the basic idea of PAID, and more complex designs in the implementation of algorithm.

DISC-all

DISC-all (DIrect Sequence Comparison) algorithm (Chiu et al. 2004) was proposed by Chiu et al. in 2004. The key element of DISC-all algorithm is the DISC strategy. It discovers the frequent k-sequences without having to compute the support counts of the non-frequent sequences. In detail, the authors define the order of two sequences having the same length. Given two sequences, they examine the items of both from left to right and compare the leftmost distinct items by alphabetical order. For example, $< abh >$ is smaller than $< acf >$ because b, in the second place, is smaller than c. The DISC strategy then finds the minimum subsequences of each sequence, and sorts the sequences according to the ascending order of these subsequences with the same length. Therefore, the DISC-all algorithm can skip many non-frequent candidate subsequences and save costs. The updating process in the DISCall algorithm involves searching the (k-1)-prefix projected database, which is similar to the mining process of PrefixSpan (Han et al. 2001).

Generally, it can be said that specific events in time, such as website traversals, nucleotides in an amino acid, computer networks and characters in a text string are examples of where the existence of sequences may be significant and where the detection of sequential patterns might be useful. Typically, sequential pattern mining algorithms are categorized into one of three broad classes that perform the task Apriori-based, either horizontal or vertical database format, and projection-based pattern growth algorithms. Improvements in algorithms and algorithmic development in general, are motivated by the need to process more data at an increased speed with lower overheads. Additionally, much research in sequential pattern mining has been focused on the development of algorithms for specific domains such as biotechnology, telecommunications, spatial/geographic domains, retailing/market-basket, and event failure detections. This has led

to algorithmic developments that directly target real problems and explains, in part, the diversity of approaches, particularly in constraint development, taken in algorithmic development.

Some common challenges that remain in the field of sequential pattern mining can described as folows:

- Pattern expressiveness: While the patterns and rules produced from the majority of approaches are simple and by growing candidate patterns, in the sense they do not take into account the use of temporal logic algebras and their derivatives. Hence, this is an area that will likely need to develop further in the future for strong usage of sequential patterns in analytics or industrial practice.

- Annotating sequential patterns with time: Some sequential patterns are examples of relative ordering in time. However, with few exceptions, pinning some of the events to absolute time points and the implication this has for pattern mining has not been investigated. For example, there are few algorithms that could state that a given sequence occurs on Mondays but not on other days.More generally, accommodating interval semantics as well as point based tokens in the event stream would provide richer rulesets. Using interval semantics as tokens in the sequence, can allow the development of powerful and efficient algorithms.

- Not many solutions in sequential pattern mining consider the confluence of ontologies/taxonomies and has to date only received minor attention. Fields like healthcare can greatly benefit when pattern mining approaches are combined with ontology and semantic information about entities occurring inside patterns.

2.2 Emerging Patterns

In numerous fields, the lack of comprehensibility can be an important drawback causing reluctance to use certain data mining algorithms. For example,

when credit has been denied to a customer, in some countries a financial institution is required to provide the reasons behind the rejection of the application. Thus, vague and indefinite reasons for denial are considered illegal. And then there are fields, such as medical sciences where clarity and interpretability are key user requirements.

An important family of interpretable classifiers are based on the use of emerging patterns (Ramamohanarao & Fan 2007). Simply put, an emerging pattern tends to appear frequently on the items or objects in one class, but it is harder to find in items belonging to other classes. As a result, emerging patterns are used to predict the class of unknown items or instances, and they report the frequency support of the discovered patterns that allows the result to be interpreted in the correct context.

Emerging pattern classifiers have been used for essential knowledge discoveries to solve real world problems in fields such as streaming data analysis (Alhammady 2007), bioinformatics (Pasquier, Pasquier, Brisson & Collard 2008), human activity recognition (Gu, Wu, Tao, Pung & Lu 2009), intruder detection(Chen 2007), anomaly detection in network connection data (Ceci, Appice, Caruso & Malerba 2008), forecasting of rare events (Gavrishchaka & Bykov 2007), and privacy preserving data mining (Andruszkiewicz 2011).

Extracting emerging patterns from a training sample is challenging, due to the following reasons:

1. The downward closure property (Zaki & Hsiao 2005), used for frequent itemset discovery, does not hold for Emerging Patterns (EPs).

2. For high-dimensional datasets, there are many potential emerging pattern candidates. Mining for all emerging patterns turns out be an NP-hard problem (Wang, Fan & Ramamohanarao 2004).

3. Continuous features having marginal values such as 3, 2.999 and 3.001 are not equal, but they can probably be found in the same pattern. In contrast, global discretization of numerical features may lead to serious

degradation of the classification accuracy, since similar values could be assigned to different discrete values.

4. Emerging pattern mining algorithms are sensitive to minimal support threshold values, and so it could be very hard for the user to define a good threshold value. The minimum support threshold is the minimal amount of instances that should support a pattern to be considered as a potential candidate.

5. Handling missing data: Missing data raises many difficulties in scientific research since data analysis procedures were not always designed for to handle them (Schafer & Graham 2002). Common approaches like data editing provides an appearance of completeness. However, sometimes estimating a missing value can lead to producing answers that are inefficient, biased, and un-reliable (Schafer & Graham 2002).

6. Emerging pattern classifiers may suffer the risk of high levels of abstention in those instances when an unknown instance cannot be assigned a class. In most classifiers abstention can be due to tie of evidences, whereas pattern based classifiers may find abstention due to lack of evidences for classifying the instances. The lack of evidence appears when no pattern matches the query object.

For a supervised classification problem, a pattern is considered discriminative if it involves properties which seek to differentiate between classes. There are numerous ways of representing discriminative patterns within classifiers, though these are implicit in some classifiers. For example, with a decision tree (Quinlan 1986) or forest (Ho 1998), the decision paths from the root to the leaves can be implicit discriminative patterns expressed in a conjunctive form. For a rule-based system (Hämäläinen 2010), the rule antecedents imply turn out to be discriminative patterns for a given class.

In this context, the emerging pattern is an important type of discriminative pattern. The ability to discriminate between two classes for an emerging

pattern is due to its support being significantly larger in one class than in the opposite class (Dong & Li 1999). Moreover, the support of a discriminative pattern in the positive class needs to be higher than a certain minimal support threshold μ. The intuition behind using a minimal support is that an emerging pattern with low support can be spurious or uninteresting, which could lead to incorrect and unreliable classifications (Fan & Ramamohanarao 2006).

There are many algorithms to search for patterns within a database. Most of them are based on restrictions that reduce the computational complexity of using exhaustive techniques. The most important and prominent restriction is the downward closure property (Zaki & Hsiao 2005). Thus, a property X satisfies the downward closure if $\ni$ pattern P, if P satisfies X, then any pattern $\bar{P}$ more specific than P also satifies X.

Searching for discriminative patterns within a training sample is the key procedure in many comprehensible and interpretable classifiers, even though these are implicit patterns. A discriminative pattern for a single class covers at most a limited amount of objects in other classes. A more specific pattern covers more objects in the positive class, but it also tends to cover objects in other classes and this limit could increase. Hence, discriminative patterns do not satisfy the downward closure and one cannot mine them using algorithms such as Apriori (Hämäläinen 2010). Additionally, there are too many candidate patterns in high dimensional datasets and exhaustive algorithms are too costly due of the size of the search space (Dong & Li 1999). Most papers about emerging patterns use a transactional representation of the objects and patterns. Thus, an instance is represented as a collection of items, or an itemset. Here, an item is an ordered pair (Feature, value), such that the "value" belongs to the domain of the "Feature". If the original database contains numeric features, they can be discretized using methods like the Entropy algorithm (Fayyad & Irani 1993). There have a number of papers extending EPs to propose extended representations such as the disjunctive emerging patterns (Loekito & Bailey 2009), the extended crisp emerging

patterns (García-Borroto, Martínez-Trinidad & Carrasco-Ochoa 2010), and the fuzzy emerging patterns (García-Borroto, Martínez-Trinidad & Carrasco-Ochoa 2011).

Dong and Li (1999) introduced the ρ-emerging pattern for two class problems, which is an emerging pattern having $GrowthRate \geq \rho$. The $GrowthRate$ 2.1 measures how frequent a pattern is in its own class C_P with respect to its frequency in the opposite class C.

$$\begin{aligned} GrowthRate(P) &= 0 \; if \; support(P,C) = 0 \; \wedge \; support(P,C_P) = 0 \\ &= \infty \; if \; support(P,C) = 0 \; \wedge \; support(P,C_P) > 0 \\ &= \frac{support(P,C_p)}{support(P,C)} \;\; otherwise \end{aligned} \tag{2.1}$$

Among these, an important category of emerging patterns (EPs) are those which cover objects in the positive class, and are named as Jumping Emerging Patterns (Definition 1). The jumping emerging patterns are widely used in emerging pattern classifiers, since they have a strong predictive capability. Thus, jumping emerging patterns describe properties that are strongly reflective in a single class, so they should be distinctive.

Definition 1. A Jumping Emerging Pattern (JEP) is an emerging pattern with infinite growth rate (Li et al. 2000).

Later, Fan and Ramamohanarao (2006) proposed the Strong Jumping Emerging Pattern (Definition 2). These patterns have an infinite growth rate, but they are also minimal with respect to the subset inclusion.

Definition 2. P is a Strong Jumping Emerging Pattern (SJEP) if it fulfills the following conditions:

1. P has infinite growth rate.

2. No proper subset of P satisfies condition 1.

Thus for a pattern not having a proper subset that satisfies condition 1, means that the corresponding pattern is a strong jumping emerging pattern

and is the minimal JEP which satisfies the frequency support constraint. Fan et al. (2006) reports that non-minimal JEPs are not useful for classification, and can be unreliable from an accuracy viewpoint, specially while aggregating many of them to make decisions. SJEPs are also known as essential JEPs (eJEPs) (Fan & Kotagiri 2002).

Wang et al (2004) suggested that aggregating many minimal EPs can cause duplicate counting of the EPs contribution, leading to lower accuracy. For example, if the properties are denoted by A, B, C, D, E, and F and the patterns ABCD, ABCE, ABCF are all minimal, then counting their contribution as an individual pattern can make the pattern ABC to be counted three times.

To solve the duplicate counting the authors proposed the Maximal Emerging Pattern (Definition 3).

Definition 3. A Maximal Emerging Pattern (MaxEPs) is an emerging pattern whose supersets are not emerging patterns.

Hence, the merits and demerits of using minimal and maximal patterns for classification, are as follows:

Using only minimal, more general patterns:

- If a smaller set of features can distinguish between two classes, then using more features do not help and can add noise (Fan & Kotagiri 2002).

- They speed up the search process, saving computing costs (Fan & Kotagiri 2002).

- Large growth rate ensures EPs strong discriminative power; large supports, i.e enough coverage on the training dataset ensures that EPs are more resistant to noise (Fan & Ramamohanarao 2003).

- Minimal EPs have higher support, thus unknown instances are easier to match (Wang et al. 2004).

- Aggregation of many minimal EPs cause duplicate counting of individual EP's contribution leading to lower accuracies (Bailey, Manoukian & Ramamohanarao 2002).

Using only maximal, more specific patterns:

- These patterns expose more information about the higher order interactions between features and are comprehensive (Zhang, Dong et al. 2000).
- They reduce the duplicated EP contribution problem (Wang et al. 2004).
- Maximal patterns are harder to find in the query object, so the classifier may have fewer patterns to decide about the classification (Wang et al. 2004).

No matter what the advantages of using jumping emerging patterns are, it can be said that they cannot capture useful properties if the dataset is noisy. Real world datasets have significant noise due to machine or user errors. To make emerging patterns tolerant to noise, a small but not strictly zero support in other classes need to be allowed (Definition 4).

Definition 4. A Noise-tolerant emerging pattern (NEP) is a minimal pattern P that satisfies:

1. support(P,C_P) $\geq \delta_1$
2. support(P,C) $\leq \delta_2$

where C_P is the class of the pattern, $C \leq C_P$ is any other problem class, and $\delta_2 < \delta_1$ are two positive integer thresholds (Fan & Ramamohanarao 2006).

Other types of emerging patterns have been defined which incorporate appropriate constraints, such as chi emerging patterns (Ramamohanarao & Fan 2007), constrained emerging patterns (Bailey, Manoukian & Ramamohanarao 2003), emerging patterns having counts of occurrences (Kobyliński & Walczak 2008), and high-level emerging patterns (Muyeba, Khan, Warnars & Keane

2011). Nevertheless, they are very specific to certain applications, and their usage differs across multiple domains.

2.2.1 Estimating the quality of emerging patterns

After training, the quality of a pattern based classifier tends to be directly proportional to the quality of the internal structure it involves to represent the relationships between patterns found in the training sample. to this purpose, we can say that the classifier quality depends tightly on the mined emerging pattern quality.

Generally there does not exist a standard methodology to measure the quality of an emerging pattern set. Hence, the quality of an emerging pattern subset is frequently inferred based on the accuracy of a classifier constructed using this subset. Nevertheless, it is to be noted that the accuracy of the classifier can be affected by many other parameters, such as the support aggregation mechanism and the pattern organization.

Typically, there are some desired properties a pattern collection should satisfy -

- Discriminative power: Every pattern should cover a significant amount of instances in a positive class, and fewer instances in the opposite classes.
- Simplicity: There should be a limited number of patterns. Violating this property could seriously lead to degrading the classifier's comprehensibility.
- Non-redundancy: Each pattern should consist of some new knowledge, with respect to the other patterns. Redundant patterns could present redundant evidence on a query instance, thus biasing the classification towards a single class.
- Generality: Patterns covering a large amount of instances tend to be less noisy. On the other hand, very specific patterns could exist due to

chance.

Algorithms for mining emerging patterns follow the following strategies for obtaining a high quality pattern collection:

- Extract patterns belonging to a particular family, like the following examples:
  - Patterns that cover instances in a single class, such as jumping emerging patterns, used in the DEEPs classifier (Li, Dong, Ramamohanarao & Wong 2004).
  - Minimal patterns with respect to the subset inclusion of their respective properties, used in the SJEPC classifier (Fan & Ramamohanarao 2006).
- Filtering of a large set of patterns, and obtaining a subset with the desired properties, used in BCEP (Fan & Ramamohanarao 2003) and LCMineC (García-Borroto, Martínez-Trinidad, Carrasco-Ochoa, Medina-Pérez & Ruiz-Shulcloper 2010) classifiers.

2.2.2 Mining paradigms

There exist three major paradigms for mining emerging patterns. These employ particular data structures and algorithms. For each of these, we present the general algorithm, the main papers, and their strengths and weaknesses.

Border-based

Dong and Li (1999) introduced the concept of emerging patterns (EPs). They found that the number of EPs in a problem could be extremely huge, and proposed a simplified representation, using the subset-closedness: they considered all the EPs as a collection of minimal and maximal patterns over the

subset inclusion relation, called borders. In their work, they reported that borders can be efficiently extracted for numerous commonly used repository databases.

ConstEPMiner (Zhang et al. 2000) introduced a set of constraints to prune the search space of EPs and reduce computations. The authors proposed an algorithm to apply these constraints to extract a subset having strong predictive power and no redundancies. Only patterns that are more general (those with top growth rate) remain and the algorithm filters patterns with the same support, considering them as redundant. Although these constraints are the basis of many post-processing and filtering methods, the algorithm also removed some important patterns impacting the classifier performance.

Borders are used in the border-based approach to represent candidates and subsets of patterns. Border differential operations are used to discover patterns, using the following general algorithm (Ramamohanarao & Fan 2007):

1. Select the minimal support threshold for each class.
2. Find the borders for each class, using an algorithm like Max-Miner (Bayardo Jr 1998).
3. Compute the emerging patterns within the border using border difference operators

Representation tree-based

Bailey et al (2002) proposed the first tree based approach to a fast JEPs mining method. The authors adapted the frequent pattern tree FP-tree (Han, Pei, Yin & Mao 2004) algorithm to deal with datasets structured as classes. Moreover, the authors reported a study on the influence of the selection of the minimal support threshold towards the classification accuracy. They discovered that, in numerous real world databases, it can be worthwhile to use higher threshold values because of the substantial decrement in computational time, at the expense of little accuracy degradations. The authors

also demonstrated the impact of mining only patterns with length below a given threshold, arguing that smaller (more general) patterns tend to be most suitable for classification. Although they found a significant increase in extraction efficiency, they also found a reduction in classification accuracies in some databases.

Later, Li et al (2007) introduced the following important modifications to the adapted FP-tree, in order to speed up the process:

- Grouping mined patterns in equivalence classes, according to the described instances. This allowed to reduce redundant patterns, and to simplify the process of computing sophisticated statistics, which are used to select the most useful patterns.
- Suppressing highly frequent and rare items, given their limited tendencies to appear in emerging patterns.
- In multi-class problems, the algorithm simultaneously mines patterns from all classes. Prior methods handled multiple classes one by one, using a single class and the complement on each iteration.

A different tree, namely contrast pattern tree (CP-tree), was proposed by Fan and Ramamohanarao (2002). A CP-tree is an ordered multiway tree structure, wherein all the instances in the training sample are covered. The mining algorithm searches depth-first the CP-tree to discover the patterns. For computational efficiency, only the strong JEPs were considered.

An adaptive version of CP-tree based mining (Terlecki & Walczak 2008*b*, Terlecki & Walczak 2008*a*) reportedly raises the minimum support threshold during the mining process. The algorithm tries to extract the top-k patterns, so the threshold is increased based on the number of patterns mined so far with the current threshold value. This optimization boosts the mining speed, since more tree branches are pruned earlier.

Bailey et al. (2003) proposed a fast algorithm for computing hypergraph transversals and applied it to mining emerging patterns. This algorithm is

based on a guided partitioning heuristic, which seems to work fine in some databases with thousands of instances.

Fan et al. (2003) created the first post-processing emerging pattern filtering process. They extracted SJEPs from the training sample, ranked them, and iteratively selected those that covered at least a new instance. The ranking considered the pattern support and the length of the pattern, discarding the growth rate information. As per the authors, EPs have implicitly large growth rate, and it does not make sense to compare between their values.

Loekito and Bailey (2006) employed Zero-Suppressed Binary Decision Diagrams (ZBDDs) (Minato 1993) as the core data structure for mining emerging patterns. Itemsets were represented as a n-bit binary vector, where each Boolean value represents the presence/absence of the particular item. Then, binary operators such as set-union, set-difference, and set intersection are performed for mining the emerging patterns. ZBDDs work like CP-trees and FP-trees, while drastically improving performances.

In cases where data are scattered in multiple tables of a relational database, it is not necessary to do costly joins to mine the emerging patterns. Appice et al (2007) proposed Mr-EP, a method to capture the differences between the instances of two classes. Mr-EP can extract emerging patterns whose properties are spanned in separated data tables. A recent technique for this purpose also uses local projections of the databases (Terlecki & Walczak 2008*b*).

Algorithms for mining emerging patterns in the representation tree-based approach employ the following steps:

1. Selection of the minimal support threshold μ.
2. Global discretization of numeric features.
3. Representation of the transformed instances using a particular data structure.
4. Traversing the structure to find efficiently mine emerging patterns.
5. Post-processing and filtering of patterns.

Decision tree-based

Garca-Borroto et a. (2010) introduced LCMine. This method extracts a representative collection of emerging patterns from a family of decision trees, induced from data. The tree induction procedure is similar to traditional methods for building decision trees, but they explore more candidate splits in order to look for properties that better describe the training sample in terms of accuracy and simplicity.

Crisp Emerging Pattern Miner (CEPM) (García-Borroto, Martínez-Trinidad & Carrasco-Ochoa 2010) is an enhanced version of LCMine. CEPM is faster and more accurate than LCMine, because it includes the following improvements:

- CEPM uses a novel weighting scheme for mining diverse patterns and it uses a stop criterion based on pattern coverage. This way, it does not have to generate a fixed amount of trees like LCMine does.
- CEPM does not need a pattern filtering post-processing. Nevertheless, it obtains fewer and more accurate patterns than LCMine does.
- CEPM assigns weights to the objects according to the support they have with the current mined patterns. This information is used in the generation of the subsequent decision trees. This way, CEPM prioritizes new patterns covering unsupported objects or objects supported in a wrong class.
- CEPM uses a novel algorithm for estimating the minimal support threshold.

For mining fuzzy emerging patterns, Garca-Borroto et al. introduced the Fuzzy Emerging Pattern Miner (FEPM) (García-Borroto et al. 2011). The mining algorithm is similar to LCMine, but it uses a fuzzy decision tree to allow extracting fuzzy patterns.

The algorithms in the decision tree-based have the following general steps:

1. Induce a diverse decision tree.
2. Extract patterns from the induced decision tree. Each pattern corresponds to the conjunction of the properties from the root node to a leaf node.
3. If stop condition is not met, return to Step 1.
4. Merge the patterns extracted from all induced decision trees.
5. Filter patterns.

It is important to note that the algorithms in this paradigm do not include a global discretization step, because they discretize only feature values appearing in the objects that belong to each tree node. The mining method has the following aspects:

- Type of decision tree to be built: fuzzy or crisp.
- Induction algorithm to build the decision trees.
- Method to obtain diverse decision trees. Classical methods to induce decision trees obtain a single tree, which is not enough to find a representative collection of patterns.
- Stop condition. This condition evaluates if the patterns mined so far are representative enough for the database.

Mining methods belonging to other paradigms are able to find all the emerging patterns in a database. Nevertheless, decision tree-based miners do not usually find all the emerging patterns, but commonly obtain a good collection of high-quality patterns. This is supported by the following reasons:

- In databases containing numerical features, there is a finite number of traditional emerging patterns, but an infinite number of extended emerging patterns. Then, it is impossible to mine all the patterns.

- Decision trees split the database using first the most discriminative properties. If the method for obtaining diversity follows this rule, the patterns mined are the most discriminant among all the patterns. So, they are the best patterns for classification.
- The experimental results presented in their respective papers show that decision tree-based miners are more accurate than traditional miners over significant database collections.

A useful characteristic of a supervised classifier is that the user can comprehend the classification results in terms of knowledge domain, particularly in those cases where the classification is contradictory with the user expectations. Unfortunately, top accurate classifiers are usually non comprehensible, while most comprehensible classifiers attain lower accuracy in most databases. On the contrary, emerging patterns classifiers build accurate and easy to understand models. Further, the commonalities (and differences) between the above described approaches in addition to algorithms in subgroup discovery (Gamberger & Lavrac 2002, Klösgen & May 2002) have been highlighted extensively Nada and Lavrac (2009).

2.3 Temporal Patterns

In previous sections, we mainly described the related research on pattern mining for attribute-value data (i.e atemporal data). Now, we focus our attention to using temporal datasets, which require various tools and techniques than those used for atemporal data.

Temporal data generally refers to any type of data which explicitly or implicitly captures the notion of time and defines a specific order. As an example, even if time is not provided explicitly and only a sequential ordering is given, we may still consider the data to be temporal (e.g., DNA sequences).

Temporal data is univariate, when the data instances consist of measurements of a single variable over time. Temporal data is multivariate when the data instances consist of measurements of multiple variables over time.

If time between consecutive events is uniform, we can say that the temporal data is regularly sampled in time (the same for all pairs of consecutive events). Else, the data is considered to be irregularly sampled in time. The latter is often the case for electronic health records, which is the focus of this thesis.

Temporal data may also be classified based on values of its observations. If the values are numeric, we have a numeric time series. If the values are discrete (belonging to a finite alphabet Σ), we have symbolic sequences. For example, a DNA sequence is a symbolic sequence, where the alphabet represents the 4 possible nucleotides $\Sigma = \{A, G, C, T\}$. A real world example of multivariate symbolic sequences involves log messages which are emitted from multiple machines or alarms that are emitted in a telecommunication network (Mannila, Toivonen & Verkamo 1997). Note that symbolic sequences can also be obtained from numeric time series using discretization (Lin, Keogh, Lonardi & Chiu 2003).

In many cases, the data do not consist of time points, but of time intervals. Time intervals have durations and are associated with specific start and end times. As an example, the data may express temporal concepts like "the patient underwent cancer chemotherapy from day 11 until day 15 of his hospitalization". Here, we consider some state sequences, where each state holds during a specific time interval.

Finally, for a temporal data model, the database may consist of a single long sequence or multiple (short) sequences. Examples of the former can be weather data (Höppner 2003) or stock market data (may be collected over many years). Examples of the latter may be web-click data, customer shopping profiles (Agrawal & Swami 1993), telephone calls, electronic health records (Hauskrecht, Valko, Batal, Clermont, Visweswaran & Cooper 2010), and so on. Long sequences are usually mined using a sliding window approach, where a window of a specific width is slid along the sequence and only patterns that are observed within this window may be considered valid (Mannila et al. 1997, Höppner 2003, Moerchen 2006).

Classifying temporal data

Next, we review some commonly used techniques for classification of temporal data. It should be noted that temporal classification and time series forecasting have differences in their methods. The task of temporal classification can be defined as follows: "Given an unlabeled sequence or time series T, assign it to one of predefined classes". In contrast, the task of time series forecasting can be defined as follows: "Given a time series T that contains n data points, predict its future values at future time points - n+1,n+2,...". Here, we discuss temporal classification methods, which are more related to the topic of the thesis.

In temporal classification problems, each sequence (time series) belongs to one of finitely many predefined classes and the objective is to be able to learn a model which can classify future sequences. There exist many practical applications of temporal classifications, involving classifying Electroencephalography signals (Xu, Guan, Siong, Ranganatha, Thulasidas & Wu 2004), speech recognition (Rabiner 1989), gesture recognition (Li, McCann, Pollard & Faloutsos 2009), and more.

A number of methods (Tseng & Lee 2005, Exarchos, Tsipouras, Papaloukas & Fotiadis 2009) classify symbolic sequences by employing a two-staged approach, which mines all frequent sequences (i.e sequential patterns) in the first stage and selects the classification sequences in the following stage. As opposed to the two-stage approach, Ifrim et al. (2011) employed interleaving techniques for pattern selection and frequent pattern mining. The author employs gradient-bounded coordinate descent to efficiently select discriminative sequences without having to explore the whole space of subsequences. Their evaluations demonstrated that this method could achieve comparable performance to the state of the art kernel-based support vector machine methods for classification of symbolic sequences.

Next we discuss the algorithms for mining time point data and time interval data.

2.3.1 Substring patterns

The simplest type of temporal patterns that can be extracted from time point symbolic sequences are sub-string patterns (Fischer, Mäkinen & Välimäki 2008). These are subsequences of symbols that appear consecutively in a sequence (without gaps). Discovering such patterns is mostly used in bioinformatics and computational biology for matching sequences of amino acids and nucleotides.

2.3.2 Sequential patterns

Sequential patterns tend to be more general than substring patterns since they do not need to be consecutive in the sequence (allowing gaps). The standard sequential pattern mining framework only cares about the order of events rather than their exact timestamps. Thus, sequential pattern mining need not require the original sequences to be regularly sampled in time. Note that the application of sequential pattern mining extends to univariate or multivariate symbolic sequences.

In the space of temporal pattern mining, the number of sequential patterns can be reduced using temporal constraints.

Temporal constraints

Mining the complete set or even the closed set of frequent sequential patterns usually leads to results that are extremely large for analysis by humans. One way to limit the number of sequential patterns can be to impose temporal constraints on the patterns. A temporal constraint is to restrict the total duration of the pattern. For example, one may specify that the total pattern duration must not exceed a given time period (e.g., 3 months). This constraint translates to defining a sliding window of width w and mining only sequential patterns that can be observed within this window. Another common temporal constraint is to define the maximum gap that is allowed between consecutive events in a pattern. Thus, we may specify that the

difference between consecutive events should not be more than g time units (e.g., 2 weeks).

Incorporating temporal constraints in the Apriori approach is described in (Srikant & Agrawal 1996) and the pattern growth approach is described in (Pei, Han & Wang 2007).

2.3.3 Time-interval patterns

Villafane et al (2000) is the earliest work in the area of mining time interval patterns. Their temporal patterns are restricted to having only containment relations, which corresponds to Allens contains relation. An example of such patterns is "during a FLU infection, a certain strain of bacteria is often found on the patient".

Kam and Fu (2000) were the first to propose using Allenś relations to define temporal patterns. Their temporal patterns, called the A1 patterns, were based on a nested representation which only allowed the concatenation of temporal relations on the right hand side of the pattern. For example, P1 ϵ ((A1 before D2) overlaps B3) is interpreted as: "state A1 is before state D2 and the interval that contains both A1 and D2 overlaps with state B3".

Hoppner (2003) proposed the first non-ambiguous representation for defining time interval patterns. The idea is to first define the normalized form of temporal patterns, where the states of a pattern are always sorted in increasing index according to their start times, end times and value. Now in order to define a temporal pattern with k states (a k-pattern), we should specify the relations for all pairs of states. For mining these types of temporal patterns, (Höppner 2003) used a sliding window approach to extract the local temporal patterns (i.e., patterns with limited total durations). He defined the support of a pattern to be the total time in which the pattern can be observed within the sliding window. Note that this definition is different from the popularly employed frequency support definition, which is the number of times a pattern appears in the data. His algorithm extends Apriori for sequential patterns to handle the more complex case of time interval

patterns.

2.4 Pattern mining in Critical Care Applications

Historically, data that is generated as a process of medical care is not just underused but is rather wasted. Traditionally, the reason behind this was related to difficulty in access, organization and usage of data stored in paper charts. Moreover, there was major variability in clinical documentation procedures which added on to the problem. In this context, medicine has remained a highly empirical process without the existence of smart ways to systematically tackle, capture, analyse and integrate information contained in the massive data generated during patient care. As a result, existing systems are typically disconnected from individual experiences and preferences, thus completely missing out on opportunities for effective personalized health and critical care delivery services.

As previously noted by Fialho et al (2013), the ICU has risen to be a compelling case for clinical data analysis. Typically, the value and impact of multiple interventions and treatments for a specific patient is just unproven, without the existence of any high quality data and well supported theories of hospital protocols and treatments. Specifically for the ICU, discovered knowledge of best practices is extremely thin in comparison to the data generated from such a complex environment. In medical circles, it is also widely believed that in a complex environment like the ICU, there is a need for variations in timely responses for patient subsets and contexts. It is thus pertinent that modern predictive methodologies can be used to take advantage of critical care databases and can thus create knowledge-bases that can be used for efficient delivery of patient care. In this context, several commercial and non-commercial critical care databases have been developed that capture patient illnesses, demographic information and physiological signals.

Sophisticated pattern mining methodologies can be used to analyse patient-specific physiological data and provide fast real-time alerts regarding severity outcomes of patient conditions. It is thus necessary to develop closed-loop predictive technologies that can act on real-time continuous physiological data and incorporate fast feedback towards an extremely efficient patient care system.

2.4.1 Short-term predictive modelling

Such modelling techniques are concerned with predicting the evolution of the individual patient and are associated with early identification of changes in the health state of the patient at the level of minutes, hours or days. Predictions can be obtained from the analysis of raw signal data generated from the different ICU information sources, whether numeric or textual. The vast majority of short term predictive modelling activities are based on data mining techniques that were developed for classification and regression tasks in general, where the inputs are assumed to be independent of time, and have then been applied to be used as features in the time-series domain. Some examples of interesting short term predictive studies involving a wide array of machine learning algorithms are reported as below.

- Bayesian networks (BN) have been employed for the prediction of fluid requirement on day two of ICU stay, as a study of inflammatory response in 3,000 patients, which resulted in a predictive accuracy of 78 % (Celi, Christian, Alterovitz & Szolovits 2008).

- A rule-learning algorithm was employed to predict impending physiologic instability across 12,000 ICU patients, resulting in 90 % sensitivity and 60 % specificity (Eshelman, Lee, Frassica, Zong, Nielsen & Saeed 2008).

- Prediction based on historic data collected, 15 and 30 minutes in advance of events of hypotension in a multi-centre database of over 260

traumatic brain injured patients via a Bayesian artificial neural network (ANN) resulted in a 41 % sensitivity and a 86 % specificity (Van Looy, Verplancke, Benoit, Hoste, Van Maele, De Turck & Decruyenaere 2007).

- Prediction of hypotension episodes, 1 or 2 hours in advance via ANN resulting in an area under the receiver operating characteristic curve (AUC) of 0.92, a 83 % sensitivity and 86 % specificity (Donald, Howells, Piper, Chambers, Citerio, Enblad, Gregson, Kiening, Mattern, Nilsson et al. 2012).

- Prediction of second day ICU discharges after non-emergency cardiac surgery via Gaussian processes (GP) conducted on a cohort of 500 patients. This resulted in an AUC of 0.76, and demonstrated a significantly better discriminative power than the EuroSCORE and the ICU nurses, and equal performance compared to ICU physicians (Meyfroidt, Güiza, Cottem, De Becker, Van Loon, Aerts, Berckmans, Ramon, Bruynooghe & Van den Berghe 2011).

Although, these studies are not meant to be an exhaustive exploration of all the available techniques in the ICU literature, they rather serve to briefly inform of the diverse spectrum and maturity of the field of predictive modelling in ICU.

2.4.2 Long-term predictive modelling

Historically, models based on demographic and administrative static data have been considered to be golden standards for long term or outcome prediction. The main reason being that they have been developed and validated in very large databases that can go back several decades. Likewise, these types of variables predate the electronic era, which eased the collection costs and feasibility when compared to monitored clinical data. Examples of long term survival prediction studies are described next.

- In a large study of over 47,000 patients, prediction of survival at 180 days after hospital discharge of patient resulted in an AUC of 0.73 for an administrative only model in comparison to the use of clinical variables which improved performance to 0.83 (Bohensky, Jolley, Pilcher, Sundararajan, Evans & Brand 2012).

- In a very large mortality prediction study on 55 Dutch ICUs and across 66,000 patients , improved performance and robustness were demonstrated by a model based on clinical data , against a model based on administrative data, having AUCs of 0.85 and 0.77 respectively (Brinkman, Abu-Hanna, van der Veen, de Jonge & de Keizer 2012).

- In a study of over 38,000 patients data from several information sources during the first 24 hours of ICU stay were used to develop ANN, SVM, decision trees (DT) and conventionally used logistic regression (LR) models, all of which resulted in similar discriminatory performance with AUCs above 0.87. Additonally, these models had similar performance as the routinely used scoring system APACHE III, albeit requiring less predictive variables (Kim, Kim & Park 2011).

The majority of long term ICU prognosis deals with mortality prediction for different risk sub-populations. In such scenarios, as discussed, there exist well established golden-standards with which to compare model performance. Long term prediction outcomes, such as mortality are commonly used for benchmarking purposes, for evaluating the financial and patient care performance of an ICU or hospital as a whole. However, unless models are sufficiently well-calibrated and discriminative to provide accurate predictions for the individual patient they are of little use to daily clinical practice. For highly performing models, a difference in prognosis between the predictions of models and the clinician's opinion could lead to more in-depth tests that can evaluate the health-state of a patient. Such models can also be employed to provide value for counselling of relatives and patients, and can be deployed in hospitals where there is a general lack of expert clinicians.

2.4.3 State-of-the art in ICU informatics

There have been certain organized research groups internationally, which have been focusing on the area of predictive analytics in critical and health care systems. A major amount of thrust on this area has been initiated after the development of the t he MIIMICII research database (Saeed et al. 2011). Fialho et al (2013) reported using a disease based modeling strategy in comparison to a general method, towards performance improvements, to predict the progress of fluid resuscitation to vasopresuure use in ICUs thus treating fluid response as an outcome variable. Mandelbaum et al (2013) employed multivariate logistic regression models for in-hospital mortality and RRT (renal replacement therapy) predictions, based on serum creatinine and urine output measurements. Customized mortality prediction models, using bayesian and neural networks, have reported better accuracy in comparison to traditional methods like SAPs (Simplified Physiology Score) for ICU patients (Celi, Galvin, Davidzon, Lee, Scott & Mark 2012, Celi, Tang, Villarroel, Davidzon, Lester & Chueh 2011). Automated intelligent methods have been reported to record more reliable blood pressure measurements associated with hypotension (Hug, Clifford & Reisner 2011). Sayadi et al (2010) developed dynamic bayesian framework models to classify ventricular complexes from ECG signnals. Clifford et al (2009) described several key problems and methods related to data collection and storage errors, noise reduction, addressing missing data, quality analysis of acquired signals, robust data fusion, false alarms in ICU etc. Typically, certain efforts have been also directed to the development and improvement of real time alarm algorithms in the ICU and relevant comparisons were reported against previous and present generation bedside monitor alarm algorithms (Zhang, Silvers & Randolph 2007, Zhang & Szolovits 2008, Wong, Clifton & Tarassenko 2012). A review of patient monitoring systems, methods and their requirements in the ICU has also been reported by Schmid et al (2013). From the machine learning viewpoint, semi-supervised learning algorithms have be reportedly been used to improve detection of intracranial pressure

alarm systems in ICU, as compared to supervised learning methods that require extensive training phases (Scalzo & Hu 2013). Additionally, static rule based induction methods have also been employed for predicting hemodynamic instability in ICU patients (Eshelman et al. 2008). Evolutionary optimization algorithms have also been used to select dynamic physiological features which were used to build prediction models among patients with sepsis and hypotension (Mayaud, Lai, Clifford, Tarassenko, Celi & Annane 2013). Lehmann et al (2012) recently employed Bayesian non-parametric methods to determine clusters of patients having similar physiological signal dynamics of blood pressure and examined it's utility in predicting mortality. Their study stressed the importance of analyzing the dynamics of physiological time series and emphasized the importance of methods that could effectively analyze such complex time series data (Li-wei, Nemati, Adams, Moody, Malhotra & Mark 2013, Nemati, Li-wei & Adams 2013). Moreover, the Physionet platform described by the authors has typically harnessed on the crowdsourcing policy of allowing participants to evaluate their learning algorithms for several challenges that they host, every year. The belief that temporal patterns in physiological time series data could be of immense use has also been covered by a recent review by Stacey and McGregor (2007). An important application area where analytical efforts have been emphasized upon, is also the neonatal ICU (NICU). Thommandram et al (2013) used static rules encoded in to the real time Artemis framework (McGregor, Catley, Padbury & James 2013) to classify neonatal spells from physiological data streams.

2.4.4 Research issues

In general, from the sequential pattern mining point of view, important case studies in critical care informatics may need to be taken up for ICU datasets for further investigations. There are numerous areas where retrospective electronic health records driven data investigations using sequential pattern mining can generate hidden clinical patterns and rules that are useful to understand progression of symptoms while leading to a specific critical event.

Some important examples of such investigations can involve:

- Mining of interesting blood pressure (BP) patterns causing hemodynamic instabilities like hypotension
- Mining of significant clinical patterns in relation to Septic Shock Prediction
- Investigating causal patterns in relation to Acute Renal Failures
- Mining of interesting sequential patterns in Cardiac Outputs (CO)

Our research is thus motivated by a need to explore and demonstrate the importance of mining interesting sequential patterns like contrastive sequences for early prediction of critical patient outcomes that can facilitate timely medical interventions.

Chapter 3

Hypotension Risk Prediction via Sequential Contrast Patterns of ICU Blood Pressure

Acute hypotension is a significant risk factor for in-hospital mortality at intensive care units. Prompt medical interventions are thus extremely important for dealing with acute hypotensive episodes (AHE). In this chapter, we describe the design of an efficient risk prediction system that can significantly help in the identification of critical care patients, who are at risk of developing an AHE within a future time span. To this objective, we first introduce the scope of prediction problems in the field of hypotension. Next, we formulate the problem of predicting events in a future time window, where related works in the area are also highlighted. Following this, we progress to describing the methodology for the experiment involving algorithm descriptions and dataset constructions. Finally, extensive discussions of prediction results and their clinical significance are reported.

3.1 Introduction

In the past few years, there has been a significant rise in patient monitoring devices aggregating large-scale patient data in intensive care units. Typically, most of this huge volume of data has remained underutilized, leading to slower progress in medical research. However, with increasing demand on healthcare organizations, there is now an urgent necessity to provide improved access and quality of care at lesser costs. As evidence obtained from modern data-driven techniques have contributed to significant advances in critical care patient diagnosis, such efforts have resulted in an improved understanding of diseases and guided appropriate medical interventions.

Appropriate clinical diagnosis of impending critical events is extremely important in an ICU, since rapid physiological changes cause critical patient instabilities that require immediate medical interventions. Conventional early warning monitoring systems turn out to be suboptimal in such cases. Existing systems embed a set of predefined clinical rules, which act on vital signs data, to raise an alarm reactively. Moreover, they are also known to generate a significant number of false alarms in ICUs (Pinsky 2007). In addition, the current systems do not account for the dynamic nature of complex physiological processes in a given time period. Hence, there exists a need for predictive technologies, which can act proactively for advanced medical decision-making in critical care units.

Hemodynamic monitoring is an essential mechanism in ICUs generating a significant amount of streaming blood pressure (BP) data. Acute hypotensive episodes (AHE) are defined as a sudden drop of patient blood pressure spanning over an extended time period. An AHE can lead to decreased tissue perfusion, which in turn can be a cause of multiple organ damages. Hemodynamic instabilities can be life-threatening to the concerned patients. On the other hand, if such instabilities are detected ahead of time to limit the effects of a life threatening event, then there are significant benefits associated with the outcomes.

The effectiveness of medical outcomes is generally assessed by the risk of

mortality and also involves the costs of treatment. For critical care patients, these factors tend to rise with time. Thus, the effectiveness of individual medical outcomes is strongly dependent on well-informed patient interventions. Proactive interventions are staged on the basis of clinical evidence of impending events. Such evidence needs to have two significant characteristics viz. predictive capability and clinical interpretability. The importance of clinical interpretability stems from the requirement of a clinician's enhanced degree of understanding of the patient's physiological condition. Such knowledge is fundamental for the selection of an optimal treatment plan.

A knowledge discovery based predictive system can meet this demand. Usually, such a predictive system takes into account time-based micro physiological events during a patient's ICU stay. It is able to make significant associations of interpretable clinical evidence to future hemodynamic behaviour. Accordingly, it has a strong potential for a reduction in operational costs, increase in efficiency, the development of novel goal directed treatments and scheduling of additional ICU services.

3.1.1 Aims of the study

The aim of this study is to identify discriminative hemodynamic sequential patterns via a novel data mining method for the risk stratification of ICU patients. These patterns are later utilized to distinguish hypotensive episodes from normotensive cases.

The informative sequential patterns are extracted from a large-scale patient population in the MIMIC-II critical care research database (Saeed et al. 2011). The MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care) database is a publicly available critical care data resource, encompassing a diverse and large population of ICU patients over the last 10 years. It comprises of high resolution temporal data including lab results, discharge notes, physiological trends and waveforms. The database has been widely used to support numerous research studies in the fields of epidemiology, clinical decision-rule improvement, and ICU alarm systems.

One important novelty of the current study is the application of a sequential contrast pattern mining strategy in the extraction of clinical episodes of arbitrary length, which are a characteristic of specific critical conditions like an AHE. The present study can thus meet the need to generate novel medical insights from the data of intensive care units and discover clinically relevant episodes separated by time windows.

3.1.2 Research contributions

Overall, our contributions made by this study include:

- the application of a contrast pattern mining technique in the field of critical care informatics
- a new method for generating predictive alerts for hypotensive episodes in an ICU, and
- validation of the method on data extracted from a large-scale deidentified critical care research database like the MIMIC-II.

3.2 Problem Definition

Acute hypotension is a clinical symptom showing a significant drop in mean arterial pressure (MAP) values for extended periods of time. The mean arterial pressure is often used in medicine as a popular measure of blood pressure, which can be derived from the systolic (SP) and diastolic pressure (DP) as given by equation 3.1.

$$MAP = \frac{2(DP) + SP}{3} \tag{3.1}$$

Although hypotension is not categorized as a disease state, it is considered to be a frequent ailment among the general population and especially among females. Owens et al (**?**) reported a prevalence of 49% hypotensive patients in a prevalence study of a general population cohort. Existing studies have

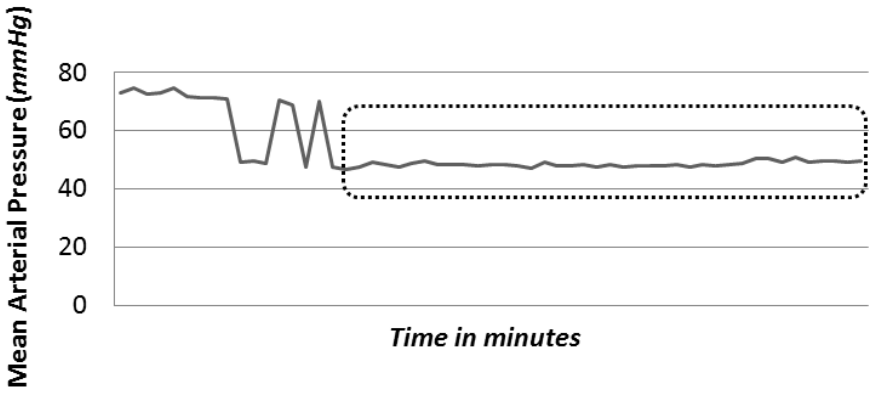

Figure 3.1: Acute Hypotensive Episode over a time period exceeding 30 minutes, when MAP $\leqslant$ 60 mmHg

indicated that hypotension is associated with morbidities stemming from dizziness and fatigue. Hypotensive subjects have previously demonstrated lower blood pressure, along with lower weights and had lesser likelihood of a family history of vascular disease or hypertension. However, in a diagnostic setting, actual prevalence can be dependent upon associated stress, anti-hypertensive medications and diuretics (Low 2008). Neurological diseases are also associated with an increasing likelihood of AHEs in an ICU. Depending on various definitions of hypotension, MAP values falling below the threshold range of 60-80mmHg for 30 minutes, could trigger an acute hypotensive episode. Figure 1 illustrates such a scenario, where MAP values sustain below 60 mmHg for a time period $\geq$ 30 minutes.

3.2.1 Formulation of the AHE prediction problem

Numerous studies report that hypotension could lead to critical events like acute kidney injury, severe sepsis, acute coronary syndrome and shock (Anderson 2011, Angus & Van der Poll 2013, Awad, Anderson, Gore, Goodman & Goldberg 2012, Mayaud et al. 2013). To enable prompt interventions, it is therefore important to predict an AHE ahead of time. Predicting an AHE can be formulated as a problem of classification of an admitted patient's

mean arterial pressure into a hypotensive or normotensive regime. The prediction of the mean arterial pressure in a future time window is central to the current study. An illustration of the AHE prediction problem is provided in Figure 2.

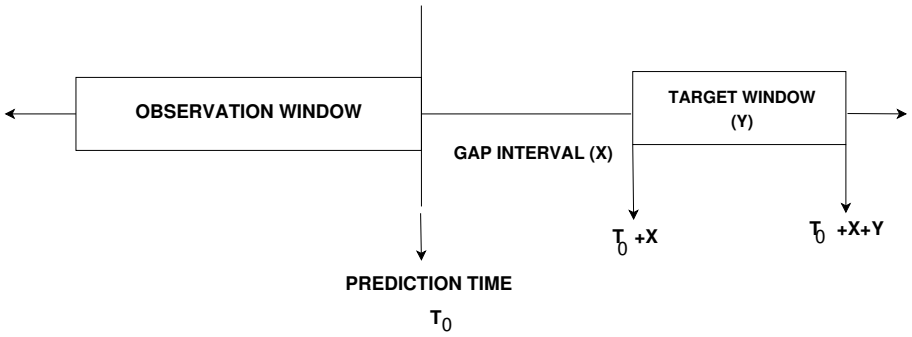

Figure 3.2: Observation and Target Windows with a Time Gap Interval

According to Figure 2, an user-defined MAP time series observation window of length 30 or 60 minutes, is provided as historical data. The time series observation window is subsequently utilized to predict the given MAP's class (hypotensive or normotensive) in a future target window of 30 minutes. Moreover, the observation and the target windows are separated by an user-defined gap interval of 60 and 120 minutes. The problem can be interpreted as that of performing an AHE prediction in a future time window, given the MAP observation data and a gap interval of one or two hours between the observation and the forecasting time windows.

3.2.2 Related works for prediction of hypotension

There have been a number of studies using pattern recognition techniques for the analysis of hypotensive behaviour. Wavelet-based similarity measures from blood pressure time series had been proposed to predict vasopressor onsets (Saeed & Mark 2006). Ghaffari et al (2010) have demonstrated the use of Hilbert-transform based techniques for predicting AHEs. In 2009, the Physionet AHE prediction challenge was instituted to advance the development

of state-of-the-art techniques (Moody & Lehman 2009), including neural networks, support vector machines and numerous statistical indices as features, for the prediction of AHE (Chen, Xu, Zhang & Mukkamala 2009, Henriques & Rocha 2009, Langley, King, Zheng, Bowers, Wang, Allen & Murray 2009). In some of these cases, historical time windows used for observations were considered as five minutes prior to the onset of an AHE. Accordingly, Wang et al (2013) have reported that medical pattern extraction was particularly challenging, owing to their longitudinal and sparse nature.

For longitudinal medical pattern extractions, Syed et al (2010) reported the development of motif mining methods, which were tested on long-range cardiovascular time series datasets. Moreover, Lee and Mark (2010) reported the extraction of hemodynamic patterns for hypotension through artificial neural networks.

For the area of predictive data mining for monitoring applications, previous research has reported the development of numerous pattern mining techniques. Typically, existing research tends to identify problems in either of two directions viz. short-term predictive modelling with the objective of generating daily alerts for physicians or long-term predictive modelling aimed at population level prognosis (Güiza, Van Eyck & Meyfroidt 2013). Monitoring systems help in capturing signals that can be used to identify time varying phenomenon, instead of traditional generation of alerts, which are known to generate a lot of false reports (Schmid, Goepfert & Reuter 2013). To overcome this weakness, intelligent noise removal methods are used as low pass filters which can aggregate high resolution signal frames and a number of good measurements (Nizami, Green & McGregor 2013). The processed input is then used for classification and regression problems, although the concerned method may or may not consider temporal aspects of the data. A wide range of ICU prediction tasks focus on the extraction of statistical features from medical time series and making them time-independent. For example, impending ICU physiological instability has been predicted by decision rules from time series data (Eshelman et al. 2008). Bayesian artificial

neural networks were employed for observation windows 15 and 30 minutes before hypotension for traumatic brain injury subjects resulting in 86% specificity and 41% sensitivity (Donald et al. 2012). Second day ICU discharges were predicted by gaussian processes (Meyfroidt et al. 2011). Celi et al (2008) also employed Bayesian networks to predict day-two fluid requirements for the study of patient inflammatory responses.

Apart from final prediction outcomes, medical decision makers also expect to discover insights relating to the processes employed on longitudinal patient records. Research on such data, begins with complex data transformation procedures by developing temporal abstractions to represent temporal relations between time intervals. Previous studies have reported the extraction of meaningful temporal patterns from a diabetes dataset (Moskovitch, Walsh, Hripcsak & Tatonetti 2014, Batal, Fradkin, Harrison, Moerchen & Hauskrecht 2012, Moskovitch & Shahar 2009). Prior to this, Tseng and Lee et al (2009) had reported temporal pattern-based classifiers for effective classification by sequences for atrial fibrillation datasets. Additionally temporal patterns were also used to predict the hospitalization of hemodialysis patients (Sacchi, Bellazzi, Larizza, Porreca & Magni 2005, Bellazzi, Larizza, Magni & Bellazzi 2005). A time-series knowledge mining method was used to discover frequent temporal patterns for patients who required mechanical ventilation for greater than 24 hours (Bellazzi, Ferrazzi & Sacchi 2011). Toma et al. (2007) utilized frequent temporal patterns to capture the evolution of organ failures status in a set of patients. Temporal history of patient event codes have also been reportedly used for mining frequent sequences of events to understand various illnesses (Patnaik, Butler, Ramakrishnan, Parida, Keller & Hanauer 2011). In this context, Perer et al. (2014) demonstrated the clinical usefulness of frequent temporal sequences by an interactive and visual analytics platform for mining sequences of ICD-9 codes to understand disease progression. Similar visual analytics platforms have been shown to have a greater clinical importance in the mining of medical event sequences having strong associations with specific disease outcomes (Gotz et al. 2014).

3.3 Methodology

The identification of sequential patterns is related to detecting subsequences contained within training sequences. According to the problem constraints, well-defined representative patterns may be grown, which display strong support in the concerned training sequences. Sequential pattern mining strategies can provide a useful alternative to mining interesting patterns of physiological time series data in an ICU in comparison to traditional scoring models, which may help discover significant insights in the form of important clinical episodes. In the following sections, we describe the various stages viz. data extraction, pre-processing and mining of sequential contrast patterns that are over-represented in the hypotensive training samples and under-represented in the normotensive samples.

3.3.1 Data extraction

The data of the study is a relevant subset of the MIMIC-II database, using a suitable data inclusion/exclusion criteria (Lee & Mark 2010). The MIMIC II is a large-scale intensive care unit database consisting of more than 30,000 patients with numerous patient variables, aggregated from patient health care records and physiological waveforms over a period of more than 10 years. The physiological time series waveforms data are organized into records, identified using unique patient identifiers. A specific patient identifier may correspond to multiple ICU stays. Thus, time series data for each ICU stay maintains a unique ICU stay identifier. The extracted subset of records also satisfied the following conditions, before extraction.

- The record had to be of an adult patient.
- Each patient time series constituted of minute-by-minute numeric samples, for at least the mean arterial blood pressure.
- Corresponding clinical records existed for the waveform records in MIMIC II.

As recommended by Lee and Mark (2010), we considered the following inclusion criteria, while compiling the data examples. As described in Figure 1, each data sample comprised of three time intervals as follows.

- a 30 or 60 minutes MAP observation window
- a 30 minutes target window
- a time interval gap of 60 or 120 minutes, which separates the observation and target windows.
- There exist seven categories for the ICD-9 code for hypotension (458.0 - 458.9) as shown in Table 3.1. Hypotensive records were selected by pattern matching over the higher level numerical classification of 458 in MIMIC-II.

Table 3.1: ICD-9 Classification of Hypotension

ICD-9 Code	Disease
458.0	Orthostatic hypotension
458.1	Chronic hypotension
458.2	Iatrogenic hypotension
458.21	Hypotension of hemodialysis
458.29	Other iatrogenic hypotension
458.8	Other specified hypotension
458.9	Hypotension unspecified

A target window was labelled either as normotensive (control) or hypotensive. The labelling of a target window as hypotensive (HE) was subject to satisfying a 30 minute period of time for which MAP was less than 60 mmHg and greater than 10 mmhg, for 90% of the time period. In contrast, a 30 minute window which did not satisfy the given HE definition as above was labelled as a normotensive (control) sample. Moreover, corresponding

to each target window, the extracted MAP observation windows were also verified to be within the 10-200 mmHg range.

Two data extraction mechanisms were considered viz. single and multiple modes. For single mode compilation, a single hypotensive or normotensive example was constructed from each separate patient waveform record. On the other hand, the multiple compilation mode considered a sliding window of 30 minutes, and all those examples were constructed, whenever satisfying the conditions for the observation and target windows.

In addition to the datasets extracted using the given inclusion criteria, hypotensive and normotensive datasets were also employed from the Physionet 2009 challenge (Moody & Lehman 2009). For the challenge datasets, their MIMIC II waveform signals were divided into two groups viz. H (hypotensive) and C (control) respectively. The groups H and C were further subdivided into H1, H2 and C1, C2. Each sub-group were defined to have the following properties.

- H1: Patients receiving pressor medication.
- H2: Patients not receiving pressor medication.
- C1: Patients with no acute hypotensive episodes during entire hospital stay.
- C2: Patients having AHE before or after the forecast window.

Accordingly, two challenge prediction tasks were constituted as follows.

- Event I: Patient risk classification between H1 and C1
- Event II: Patient risk classification between H and C

Moody and Lehman (2009) reported that the groups H1 and C1 indicated the extremes of AHE-associated risks. The described groups in Event I and II can also be termed as the target class definitions.

3.3.2 Data discretization

Physiological data often comprise of repetitive elements. To identify interesting patterns, a natural extension is to transform the real-valued physiological time series into string representations for mining symbolic discrete patterns (Pinsky 2007). Subsequently, we employed the symbolic aggregate approximation method (Lin et al. 2003) to segment the original MAP signal into discrete intervals and assigned an alphabetic label to each discrete region. This process transforms the continuous MAP data into a symbolic sequence, and enables the use of numerous pattern mining algorithms. The symbolic aggregate approximation (SAX) technique has emerged as a popular and efficient technique, producing an informative symbolization of large-scale time series data. Typically, SAX converts the continuous time series into a piecewise aggregate approximation (PAA) form (Lin et al. 2003). Later, the PAA series is converted to a symbolic sequence. Each MAP time series, before being discretized, undergoes a normalization process having a mean of 0 and variance 1. The SAX strategy selects breakpoints using a gaussian distribution, such that the discrete symbols are equiprobable in the time series. For example, to transform a normalized time series using five symbols, the discrete regions are specified by [$-\infty$, -0.84, -0.25, +0.25, +0.84, $+\infty$]. The symbolic representation adopted by SAX characterizes the inherent properties of the time series data. Consequently, an equiprobable distribution of symbols is maintained in the given time series (Lin et al. 2003).

In the process, SAX provides an effective discretization platform, which can be utilized to create efficient pattern mining and indexing algorithms for medical purposes. Figure 3 illustrates a visual representation of a real-valued time series being converted to a symbolic form, using five symbolic regions.

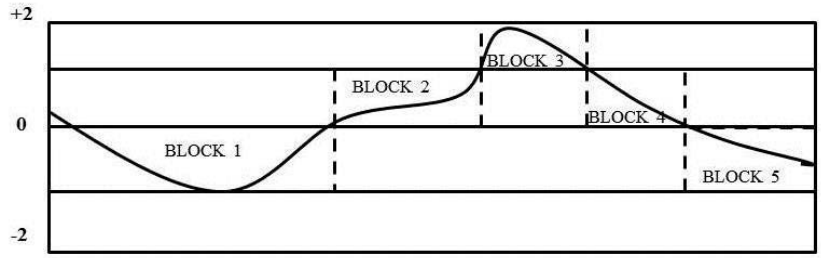

Figure 3.3: Discretization by Symbolic Aggregate Approximation using 4 symbols

3.3.3 Mining gap-constrained sequential contrast patterns

In studies related to binary or multi-class classification, the central objective is to develop a prediction model, which is capable of distinguishing an incoming signal using its inherent properties and assign a target label as the predicted outcome. Typically, in data mining problems, there exists a strong motivation to discover differentiable patterns' characteristic of disparate groups of data, that are used for prediction of records. Mining emerging patterns from distinctively labelled groups of relational data was initially introduced by Dong and Li (1999). However, the immediate application of emerging patterns to sequence databases was not possible owing to ordering of a sequence, and due to multiple occurrences of items in a sequence. Accordingly, the concept of emerging substrings was suggested (Chan, Kao, Yip & Tang 2003). Substrings are a special case of subsequences, where each consecutive symbol is separated by a gap interval of 0. Yet, an important aspect to note is that significant sequential episodes may not manifest as consecutive symbols existing in time-series symbolic sequences of interest. Thus, the identification of episodes having significant events ordered sequentially, while having arbitrary gap intervals between events, can be extremely useful. Towards this purpose, a number of algorithms have been reported (Xing, Pei & Keogh 2010). In the present study, we intend to discover gap-constrained

contrast subsequences from disparate groups of sequence data, using the principles of frequency support. In the following sections, the various definitions and processes associated with the extraction of gap-constrained sequential contrast patterns are described.

Sequential patterns

Let there be a set of distinct items denoted as I. I can also be called the alphabet set and $|I|$ is the size of the alphabet set. A sequence S defined over I may be denoted as $e_1 - e_2 - .... - e_n$, such that e_i belongs to I for $1 \leq i \leq n$. Accordingly, we consider univariate sequences where e_i represents a single item from I. A sequence $S' = e_{i1} - e_{i2} - ... - e_{im}$ is said to be contained in a sequence $S = e_1 - e_2 - e_3 - ... - e_n$, such that $1 \leq i_1 \leq i_2 \leq \ \leq i_m \leq n$. For example, a subseqeunce CD is contained in $CAAD$, but not DC. Hence, the order of the sequence S' is maintained in S, although items in S' are not consecutive in S. This indicates the existence of gap intervals between the items of S'.

Definition 3.3.1.1: (***Max-Prefix***) The max-prefix of the sequence $S = e_1 - e_2 - ... - e_k$ is given by $e_1 - e_2 - ... - e_{k-1}$. It constitutes the leading sequence of elements in S, without the final item of S.

Definition 3.3.1.2: (***Occurrence of a Subsequence***) Given the sequences, $S = e_1 - e_2 - .... - e_n$ and $S' = e_{i1} - e_{i2} - .... - e_{im}$, S' occurs in S if $1 \leq i_k \leq n$ and $e_k = e_{ik}$ for all $1 \leq k \leq m$, and $i_k \leq i_{k+1}$ for $1 \leq k \leq m$. For example, given sequences $S = XZXZYZY$ and subsequence $S' = XY$, there are four occurrences of S' in S at the positions - $\{1, 5\}, \{1, 7\}, \{3, 5\}$ and $\{3, 7\}$.

Definition 3.3.1.3: (***Satisfaction of Gap Constraints***) Consider a sequence $S = e_1 - e_2 ... - e_n$ and an occurrence $O = i_1, i_2, ..., i_m$ of a subsequence S', if $(i_{k+1} - i_k) \leq g + 1$, such that $|k| \in \{1, ..., m - 1\}$, then S' for the occurrence O, fulfills the gap constraint of g. Moreover, fulfilling the gap constraint once, in a given sequence serves the condition of gap-constraint satisfaction. For example, if $g = 2$, then XY is a subsequence of XZY, but not $XZZZY$. Now, let us consider $D = \{D_1, D_2, \ldots, D_n\}$ as a set of sequences in a

database, a sequential pattern P, and a gap-constraint of g, then the frequency of occurrences of P in D is given by $count_P(D, g)$, also known as the absolute frequency support of P in D. If there exists a frequency support threshold α and P satisfies a condition such as $count_P(D, g) \geq \alpha$, then P is is said to be frequent in D, with a gap constraint of g.

Definition 3.3.1.4: (***Gap constrained sequential contrast patterns***) Given two sets of sequence datasets D^+ (positive sequences) and D^-(negative sequences), two thresholds α and δ, and a maximum gap of g, a gap-constrained sequential contrast pattern P is required to satisfy the following conditions.

(1) Positive Support: $count_P(D^+, g) \geq \alpha$

(2) Negative Support: $count_P(D^-, g) \leq \delta$

Thus given D^+, D^-, α, δ and g, mining the gap-constrained sequential patterns involves finding the set of all such subsequences that fulfill the given conditions from (1) to (2).

Generation of candidate sequences

Towards finding the set of all gap-constrained contrast sequential patterns, we employ the ConSGapMiner algorithm (Ji, Bailey & Dong 2007), which was earlier used to extract minimal distinguishing subsequences (MDS) with user-defined gap constraints. The method utilizes the depth first search (DFS) technique for the generation of candidate sequences. This is done by growing a lexicographic sequence tree (LST) as shown in the example in Figure 3.4. Each node in the LST embeds a subsequence, along with its positive and negative frequency supports. In addition, each node is a max-prefix of its children.

Pruning non-minimal subsequences: After a sequence node is generated, if it satisfies the conditions (1) and (2), then the sequence node is not extended further. A supersequence of a potential contrast sequence is not minimal (Ji et al. 2007). Thus, restricting the growth of sequences by a minimality condition, helps in the reduction of redundant patterns.

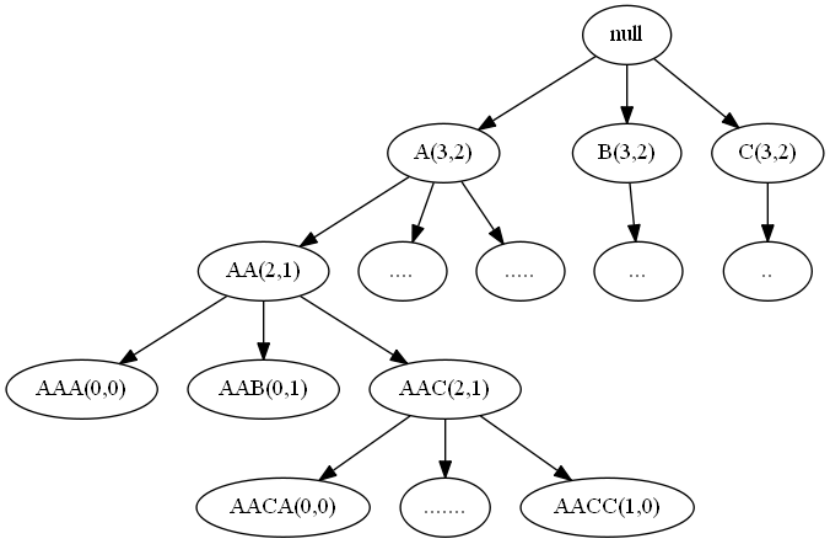

Figure 3.4: A Lexicographic Sequence Tree (LST) growing candidate sequences using 3 symbols as A, B, C

Pruning of infrequent subsequences: If a sequence node's positive frequency support is less than α (as specified in condition (1)) , then the concerned node need not be extended. This is because, supersequences of an infrequent max-prefix are also infrequent.

Gap constraint verification

For the verification of gap-constraint satisfaction, we employed a bitmap representation reported earlier for checking gap-constraints (Ayres et al. 2002). The bitmap process is explained by an example, as shown in Table 3.2. Let us consider verifying the gap constraint of XY in $XZXZY$, given maximum gap g is set to 2. In the first step, all the occurrences of X in the concerned sequence are set to 1 (as shown in X_{index}). These are position indices given by 1 and 3. Later, $(g + 1)$ index positions are set to 1 for each occurrences following X, separately as illustrated in rows 3 (given as 1_X) and 4 (given as 2_X). Following this, the bit vectors in rows 3 and 4 go through a logical OR operation, as given in row 5. Subsequently, a logical AND operation is

performed on the bit vectors in row 5 and for the occurrences of Y in row 6, to obtain a final bit vector, in row 7. An occurrence of 1 in the final bit vector (at row 7) indicates that the gap constraint of $g = 2$ was satisfied.

Table 3.2: Checking gap constraint satisfaction of XY in XZXZY

	X	Z	X	Z	Y
Index	1	2	3	4	5
X_{index}	1	0	1	0	0
1_X	0	1	1	1	0
2_X	0	0	0	1	1
$1_X(OR)2_X$	0	1	1	1	1
Y	0	0	0	0	1
AND	0	0	0	0	1

Finally, a post-processing step is applied such that any super-sequence of at least another shorter sub-sequence, is removed from the resulting set of contrast sequences. The algorithm for the generation of candidate sequences is provided by Algorithm 3.3.3.

3.4 Prediction Results

The sequential contrast pattern mining methodology was applied to both single-mode and multi-mode datasets, based on a clinical inclusion criteria, similar to principles used in (Lee & Mark 2010). From the MIMIC-II database, we extracted 254 segments (single mode) and 759 segments (multi-mode), which satisfied the criteria of hypotension. For the normotensive group, 275 segments were compiled for single mode whereas for multi-mode the exact number of segments varied from 13,3712 to 14,0006.

In addition, we also applied our techniques to the datasets provided by the Physionet 2009 AHE prediction challenge (Moody & Lehman 2009). In particular, the AHE challenge datasets had also been extracted from the MIMIC-II database in 2009. Our single-mode and multi-mode datasets tend

Algorithm 3.1 Generation of candidate sequences

$candGen(c, g, I, \delta, \alpha)$

1: *Require* : $c - sequence, g - maximum\ gap, I - alphabet, \alpha - maximum\ positive\ support, \delta - minimum\ negative\ support$
2: $ds \leftarrow \phi$ {*ds holds the distinguishing children of c*}
3: **for** $i \in I$ **do**
4: **if** $c + i$ is not a supersequence of any sequence in ds **then**
5: $nc \leftarrow c + i$
6: $supppos = SupportCount(nc, g, pos)$
7: $suppneg = SupportCount(nc, g, neg)$
8: $supppos \geqslant \alpha AND suppneg \leqslant \delta$
9: $ds \leftarrow ds \cup nc$
10: **else**
11: **if** $supppos \geq \alpha$ **then**
12: $candGen(c, g, I, \delta, \alpha)$
13: **end if**
14: **end if**
15: **end for**
16: $DS \leftarrow DS \cup ds$

to extend these datasets, since MIMIC-II has undergone multiple version updates, in the past 10 years. For the Physionet challenge, each of H1, H2, C1 and C2 groups consisted of 15 samples for training purposes. For test sets, Event I included 10 samples (H1=5, C1=5), while Event II had 40 (H=14, C=26). For the challenge data, an example training record like *a40439* contains a T_0 time-annotation, indicated as 18.30 on 04/09/2008 (T_0 was provided with each record). The time series data prior to T_0 is used for training purposes (treated as the observation window).

For the prediction of a record, a majority vote of contrast sequences is considered for the record to be treated as hypotensive. Single and multimode datasets extracted for the present study are available via https://github.com/s-ghosh/hypotension

3.4.1 Prediction performance on the two data sets

On the first data set, our 5-fold cross-validation classification results for both the single mode and multi-mode cases are summarized in Table 3.3 and 3.3. As can be noted, the classification results for the single mode executions are much better than multi-mode executions. This is because the single mode cross-validation accuracies are higher than multi-mode accuracies. A lower specificity in single mode executions can be attributed to the balanced nature of the single mode datasets. In contrast, the multi-mode datasets consist of a significantly higher percentage of instances, which are normotensive (for. e.g, 759 H to 140006 N). A sensitivity of 100% in our experiments, indicates that the sequential contrast method was able to predict all AHE instances correctly. Typically, the number of AHE instances are much fewer in comparison to non-AHE instances. As a result, the contrast pattern set generated due to the imbalance, can also consist of patterns which fulfill support conditions among non-AHE instances. Owing to this reason, contrast sets are highly capable of identifying positive instances. However, lower specificities reflect that a high percentage of false positives are also generated. Thus, our method demonstrates good performance when employed in the prediction of

an AHE. This means sequential contrast patterns are effective in detecting hypotensive behaviour. However, since similar blood pressure patterns also exist across both population groups, a lot of negative instances are incorrectly classified as hypotensive. Similar experiments on MIMIC-II by Rocha et al (Rocha, Paredes, De Carvalho & Henriques 2011) demonstrated a sensitivity of 82.8% and a specificity of 78.4%. In another study, Lee and Mark (Lee & Mark 2010) also demonstrated highest accuracies of 76% for single-mode and 86% for multi-mode datasets extracted from MIMIC-II. Moreover, increasing the size of the observation window does not result in significant improvements in performance. Also, increasing gap intervals from 60 to 120 minutes lead to a drop in performance. Specifically, the hypotensive (positive) segments were always predicted correctly in both the modes.

Generally, retrospective EHR based population comparison studies tend to have imbalanced datasets, where the count of positive instances is very small as compared to the negative instances. As a possible enhancement, contrast pattern sets can be post-processed using multi-objective optimization methods to obtain the most optimal combinations of contrast sequences for building models, which demonstrate better specificity, while reporting a higher classification performance.

Table 3.3: Single Mode Classification Performance with 10 symbols

	Gap Interval = 60 minutes		Gap Interval = 120 minutes	
	ObWin = 0.5 h	ObWin = 1 h	ObWin = 0.5 h	ObWin = 1 h
Sensitivity	100%	100%	100%	100%
Specificity	65.85%	68.29%	61.44%	62.19%
Accuracy	82.27%	83.54%	79.87%	80.37%

For the Physionet 2009 challenge dataset, the test prediction results are presented in Table 3.5. In Table 3.6, we provide a comparison of our results with the reported results from the Physionet 2009 challenge. As seen, models employing neural networks (GRNN, RPS-NN) and kernel methods like SVM

Table 3.4: Multi Mode Classification Performance with 15 symbols

	Gap Interval = 60 minutes		Gap Interval = 120 minutes	
	ObWin = 0.5 h	ObWin = 1 h	ObWin = 0.5 h	ObWin = 1 h
Sensitivity	100%	100%	100%	100%
Specificity	81.19%	80.76%	79.36%	74.79%
Accuracy	81.30%	80.88%	79.48%	74.94%

are heavily dependent on several parameters, and can have performances over wide ranges (Henriques & Rocha 2009, Mneimneh & Povinelli 2009, Jousset, Lemay & Vesin 2009). Most of the other methods employed rules based on simple averaging measures and still performed fairly (Chen et al. 2009, Fournier & Roy 2009). Moreover, hidden markov models (HMM) for hypotension had reported a cross-validation accuracy close to 97% (Singh, Tamminedi, Yosiphon, Ganguli & Yadegar 2010), which compares well with our cross-validation results too.

Table 3.5: Physionet 2009 AHE Test Prediction Classification Accuracies for events I and II given G=3

	Event I			Event II		
	S=3	S=4	S=5	S=3	S=4	S=5
L=8	5/10	7/10	7/10	23/40	23/40	32/40
L=9	5/10	7/10	9/10	23/40	25/40	33/40
L=10	5/10	7/10	10/10	25/40	32/40	36/40
L=11	5/10	7/10	10/10	25/40	32/40	36/40

3.4.2 Discussion

A comparison of our results with the reported results from the Physionet 2009 challenge demonstrates our competitive classification performances against

Table 3.6: A Comparison of classification methods employed for the AHE prediction problem. Sequential patterns report comparable accuracies against existing methods

Method	Event I	Event II
GRNN	10/10	37/40
5-min average of diastolic ABP	10/10	37/40
MAP averaging Rule	10/10	36/40
5-min average of ABP	10/10	36/40
Linear Regression	10/10	36/40
Median of MAP	10/10	34/40
NN with feature selection	9/10	32/40
SVM	10/10	30/40
RPS-NN	2/10	25/40
Sequential Contrast Patterns	10/10	36/40

those models employing neural networks (GRNN, RPS-NN), kernel methods like SVM, hidden markov models and various other statistical measures (Chen et al. 2009, **?**, **?**). Additionally, the effect of parameters like subsequence length (L), alphabet size (S) and maximum gap (G) are shown in Figure 5. As seen, the best performances were achieved using a maximum gap of 3, subsequence length of 10 and an alphabet of cardinality 5. A general trend is observed, where informative sequences could be extracted if the maximum gap constraint is iteratively increased. This has been demonstrated by Figure 5. A number of values were used incrementally for tuning and to reach the optimal value of 3. Increasing the gap threshold further does not improve the predictive performance of the algorithm. On the other hand, increasing gap threshold to higher values over 5 affected the computational run time of generating patterns.

As seen, classification performances tend to improve with an increase in gap sizes. At the same time, a very large gap size G, also means that

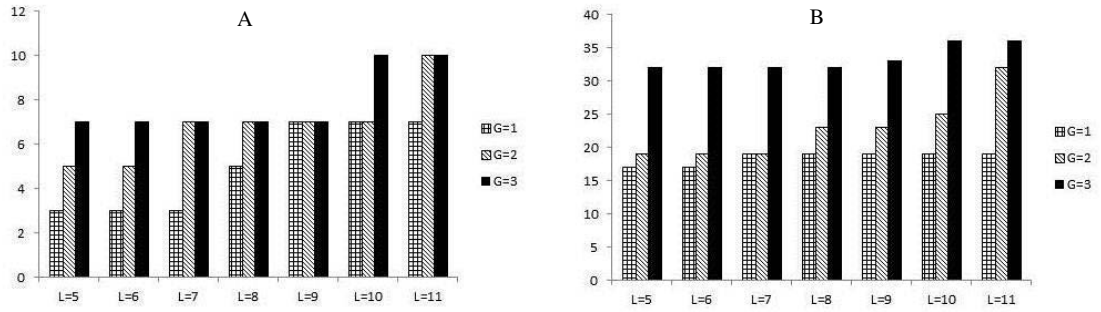

Figure 3.5: Effect of parameters L and G on the performance (A) For Event I, (B) For Event II

two consecutive symbols for a sequential pattern have occurred over a wide range, where the size was G. Extremely large gap sizes can impede a proper interpretation of contiguous events in a sequence. Typically, patient events occur over a time span, covering multiple days. Thus, sequences may be clinically useful and unique, when considered for shorter time windows in the original patient timeline, with multiple days. For larger cohorts, finding out an optimal gap is dependent on the resolution of the time series (i.e. the sampling frequency). Typically, for detecting differential blood pressure patterns, effective gap sizes can be decided based on their ability to capture clinically meaningful and informative episodes, spanning over shorter windows. In addition, increasing S provides more number of discrete cut points for MAP, and enables the algorithm to capture patterns which characterize more fluctuations in the blood pressure. Thus, for cases with S=5, the algorithm is able to find a more expressive pattern, than for S=3. Hence, selecting an alphabet size of 5 turned out to be an optimal choice, both in terms of the discretization of blood pressure range as well as keeping the algorithmic running costs within limits. This also contributes to making improved predictions. Thus, finding interesting sequences is highly dependent on the use of various parameters like the number of symbols, length of subsequence and gap sizes. Generally, the selection of appropriate parameter values like L (pattern length), G (gap size) and S (alphabet) tends to affect

the cardinality of the set of discovered patterns and the algorithmic running time. Thus, extracting minimally expressive shorter sequences allows the algorithm to restrict the running time as well as identify patterns, which are clinically important and appear in longer sequences.

In contrast to our method, the 5 minute averaging measures are statistical features obtained from a 5 minutes window prior to the immediate occurrence of an AHE. Thus, a major difference lies in the fact that our method considers a wider window of 30 and 60 minutes, prior to the onset of AHE (Chen et al. 2009). This also indicates that a method, which is effective in performing predictions using wider time windows may be more suitable in a real time scenario, in comparison to statistical measures obtained from a 5 minutes window (prior to AHE). In this context, better results from the 5 minutes timespan prior to an AHE, may be due to temporal proximity to the onset of an AHE. For methods employing neural networks, both GRNN and RPS-NN report 10/10, 2/10 (for Event 1) and 37/40, 25/40 (for Event II). These methods tend to be strongly dependent on the tuning of parameter, as was also discussed by the authors (Henriques & Rocha 2009). Additionally, recent experiments by Yapps et al (2017) reinforce that the use of feature selection algorithms on features constructed using blood pressure trends can be highly predictive.

The contrast mining method, on the other hand, helps to extract discretized sequential representations of the MAP time series, which provide the maximum support towards the occurrence of an AHE. These patterns are later useful, to not only predict an AHE for an unknown record, but for further clinical interpretation by domain experts. Our results indicate that sequential contrast patterns are capable of extracting informative symbolic episodes, which may be employed for both AHE risk prediction and understanding of hemodynamic behaviour towards effective analyses of sequential episodes, that may be indicative of medical symptoms.

3.5 Examples and Clinical Significance of Sequential Contrast Patterns

Acute hypotension is one of the most dangerous clinical conditions that frequently occurs in an ICU and can cause serious renal, cerebral and myocardial hypoxic damage. Existing medical interventions are reactive (i.e after an AHE has been triggered), for recommending treatment of underlying causes. In contrast, early bedside detection of AHEs can enable the development of life-saving interventions. Clinical interventions to treat AHE attempt to restore the physiological status of the body by targeting recommended BP values, increasing fluid and salt intakes, administration of vasoactive agents and so on (Shibao, Lipsitz & Biaggioni 2013, Takala 2010). The AHE definition considered in the current study, utilizes hypotension thresholds reported in previous studies (Lee & Mark 2010, Moody & Lehman 2009). Although ranges between 65-75 mmHg have also been reportedly used for defining hypotension, definitions for AHE time periods may also vary from 1 to 60 minutes, depending on the objective of the study. However, drops in blood pressure within smaller time spans (as indicated by monitoring systems), may not always indicate an AHE. Such changes may be due to monitoring errors or physiological changes caused by normal human activity. Hence, a larger time window of 30 minutes is a suitable definition for capturing AHE related information. Taking forward the suggested inclusion criteria for an AHE, we additionally employed the widely used ICD-9 code of hypotension to extract clinical records from MIMIC-II. The ICD-9 coding system describes a disease classification scheme used to monitor population group health situations for general epidemiological, health management purposes and clinical usage. The extracted datasets were sourced from the MIMIC-II repository, which tends to provide further credence to the study.

3.5.1 Sequential pattern examples

Our sequential pattern mining algorithm can discover simple-to-understand clinical symbolic subsequences. These subsequences can be treated as evidence while diagnosing for diseases. Even though methods such as neural networks and SVM demonstrate competitive prediction performances, they heavily dependent on non-linear kernel functions and parameters. But, our sequential pattern mining methods extract signatures of clinical episodes in the form of symbolic patterns.

Table 3.7: Representative Examples of Extracted AHE Sequential Patterns

..D..E..D..E..D..A..B..C..D..C..
..D..C..E..D..C..B..C..D..C..D..
..B..C..D..C..A..C..D..E..D..C..
..D..C..E..C..A..C..D..E..D..E..
..C..A..B..A..E..E..C..B..C..D..
..A..B..A..E..D..B..B..B..C..A..
..E..C..B..A..B..A..B..C..D..
..A..B..B..D..E..C..B..C..D..

In this study, we were able to mine a set of discretized sequential patterns like ABAEDBBBCA, which were prominent in acute hypotensive patients. Examples of representative sequential blood pressure patterns for hypotension are as reported in Table 3.7. For example, in the case of ABAEDBBBCA, the sequence indicates that the mean arterial pressure follows the given pattern trajectory among a majority of AHE patients. The given symbols indicate that the mean arterial pressure time series region was divided into 5 equiprobable regions (given by A, B, C, D, E) from 0 to 200 mmHg. The example pattern illustrates that the blood pressure time series followed a situation where majority of the AHE patients record

an episode of events represented by the MAP value in a particular sequential order of blood pressure regimes demonstrated symbolically as follows - $A \leq B \leq A \leq E \leq D \leq B \leq B \leq B \leq B \leq C \leq A$. Thus, each sequential pattern describes a train of clinical events, represented by the specific blood pressure regimes, categorised by discrete symbols.

3.5.2 Pattern visualization and clinical interpretation

Interpretive sequential representations can be extremely useful to clinicians for understanding the sequence of physiological states that a patient passes through, before developing a critical condition. Such interpretations can help establish potential combinations of observable physiological sequences, that precede AHE. Generally, the objective of clinical studies involves the estimation of causal relationships between selected clinical variables and disease specific laboratory test outcomes. Given temporal data for clinical variables, sequential patterns of specific clinical variables can aid in the interpretation of complex relationships between variables and patient specific outcomes. Towards this objective, general visual trends may be inferred from gap constrained sequences as shown in Figure 6. Thus, sequential patterns can have immense potential in the exploration of underlying clinical relationships to facilitate personalized treatments. Accordingly, similar studies have also claimed that the visual exploration of sequential and temporal patterns in clinical patient data can significantly aid in clinical decision making (Gotz et al. 2014).

Moreover, mining of complex contrast sequences in hypotensive patient groups can aid in the development of interesting clinical hypotheses such as the detection of a succession of clinical events prior to the onset of AHEs. Thus, extracting sequential contrast patterns can guide clinical decision-making towards the effective investigation of hypotensive events. In addition, the proposed methodology is flexible enough to also accommodate clinician-defined constraints.

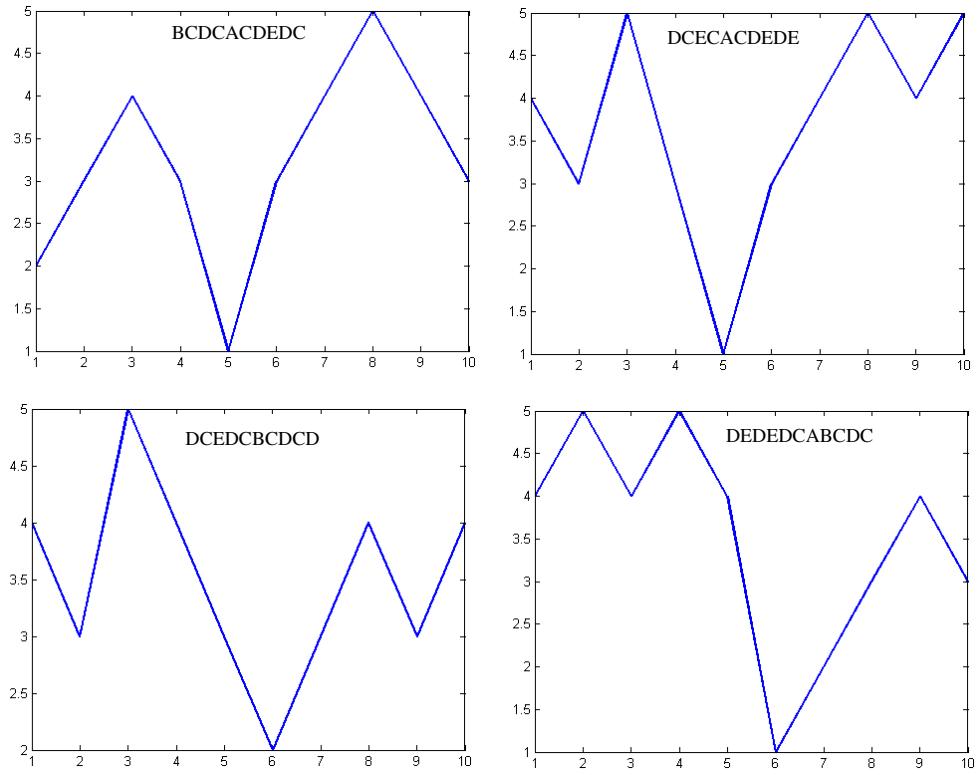

Figure 3.6: Inferring Visual Trends from Sequential Contrast Patterns Examples for AHE (A=1, B=2, C=3, D=4, E=5)

3.6 Conclusion

The current study investigated the application of a novel sequential contrast pattern mining methodology for predicting acute hypotensive episodes in an ICU. Our study demonstrates that research on the mining of informative sequential patterns can be of significant clinical value to concerned stakeholder in a clinical setting. In addition to demonstrating the classification performance, we also established the existence of gap-constrained symbolic subsequences, which have strong clinical interest to practitioners. Since the data encoded of a patient's journey is inherently temporal in nature, sequences have the capability to uncover numerous hidden patterns, which are otherwise not visible. As part of a knowledge discovery process, the con-

trast pattern mining method extracts patterns, which collectively help in the prediction of an AHE. A real-time application of the reported strategy can help derive significant sequential patterns of interest, which could be translated into a complex sequence of clinical events. A higher frequency of the occurrence of complex contrast sequences while comparing hypotensive and normotensive patient groups may be beneficial to a clinician to develop a clinical hypothesis relating to a succession of clinical events leading to an AHE. Extracting sequential patterns from hypotensive patient groups can inform medical decision-making towards the diagnosis and investigation of AHEs. Thus, significant patterns are a potential source for launching further data driven investigations validated by randomized clinical trials. Such patterns can also be employed in conjunction with multiple types of clinical features for the construction of accurate AHE prediction systems. In summary, the sequential contrast pattern mining approach described in this work well relates to the expectations of evidence-based medicine.

Chapter 4

Using Sequential Patterns as Classification Features for Accurate Prediction of ICU Events

Previously, pattern mining algorithms have been employed for extracting interesting rules in various clinical domains. However, the extracted rules are directly investigated by clinicians for diagnosing a disease. Towards this purpose, there is a need to develop advanced prediction models which integrate dynamic patterns to learn a patient's physiological condition. In this study, a sequential contrast patterns-based classification framework is presented for detecting critical patient events, like hypotension and septic shock. We build on top of work done in the previous chapter to use sequential contrast patterns, for conversion to two novel representations-(1) binary and frequency-based feature space and (2) ordered sequences of patterns, which conserve positional information of a pattern in a time series sequence.

4.1 Introduction

Patients in an intensive care unit (ICU) demonstrate numerous dynamic fluctuations in physiological responses, owing to underlying biological conditions. A cascade of multiple physiological episodes may be relevant determinants for identifying impending critical events. Typically, clinical determinants of medical events can range from disease severities, age, various comorbidities, drugs used and fluids taken, to physiological changes due to diagnostic treatment interventions (Latronico 2015). Conventional systems typically employ clinical scores like the Glasgow Coma Scale (GCS), Full Outline of Unresponsiveness (FOUR) and simple statistical models based on easily obtainable clinical covariates such as sex, age, artificial ventilation, hospital readmissions to assess patient risk of a disease (Vincent & Moreno 2010). However, scoring models based on simple indices have failed to capture dynamic symptoms in a fast evolving patient state. Therefore, it is important to consider large-scale sophisticated prognostic models which capture dynamic trends in patient behaviour, using informative clinical features (Jensen, Jensen & Brunak 2012). To design a system capable of making short term event predictions, advanced feature representations, which are episodic, sequential and temporal in nature, need to be considered. Learning complex feature representations in clinical scenarios can thus aid in the development of extremely powerful prediction models (Li, Li, Jia, Ramanathan & Zhang 2015). In this context, the electronic health record (EHR) of a patient is a valuable source of longitudinal data for exploratory mining of complex or latent features to assist with clinical decision-making. However, it is an extremely challenging task to find an optimal set of informative features from heterogeneous clinical records. Thus, advanced learning models that capture complex variations as feature representations in patient hemodynamic conditions are extremely important for staging early and effective life-saving medical interventions.

Conventional methodologies have been proposed to determine the relationships between a risk factor and a clinical outcome using statistically significant regression models such as logistic regression, linear regression, and

Cox regression (Bender 2009). Alternative sophisticated clinical machine learning models have focused on decision trees, association rules and artificial neural networks (Ha 2011, Ordonez & Zhao 2011). Feature selection algorithms are also employed to select optimal clinical feature sets, prior to being fed as input data to classifier. Informative feature sets have been employed in the construction of Bayesian networks, and features were considered to be conditionally independent to each other (Li, Shi & Satz 2008). It can be noted that the above described methods tend to employ handcrafted feature sets due to their easy identifiability and visibility for human users. However, such approaches are relatively time-consuming, and incomplete.

Yet from a data mining viewpoint, there have been limited recent studies exploring the integration of pattern mining and machine learning for ICU prediction problems (Batal, Valizadegan, Cooper & Hauskrecht 2013, Moskovitch et al. 2014). Enabling such integrations in clinical contexts is a difficult problem owing to the unstructured and longitudinal nature of medical records (Wang, Lee, Hu, Sun, Ebadollahi & Laine 2013).

The current study builds on foundations from Chapter 3, and proposes a pattern based classification framework, where sequential contrast patterns are employed as features using different feature space transformations from multivariate physiological time series to predict short term ICU events. Each pattern mapping method creates a new feature space involving binary valued attributes and frequency based information, to predict ICU events like the onset of a future acute hypotensive episode (AHE) and patient mortality. Generally, the integration of these methods involve employing patterns to define features for building classification models, which are easily interpretable by clinicians.

The detailed contributions of the current study are:

- The use of sequential contrast patterns to transform the original patient sequence data to a feature set, using two pattern mapping approaches,
- A comprehensive modeling approach is adopted by comparing two types of pattern-based models for predicting a test instance, and

- Two applications on critical ICU event prediction are reported using the integrated sequential patterns based modeling framework.

Our study demonstrates that the integration of contrast sequences using learning models like support vector machine (SVM) and Naive Bayes can achieve improved performances, in comparison to traditional models operating on simple clinical and statistical features. Thus, we systematically investigate the use of sequential contrast patterns to construct a feature space of patterns, which is used to build better classification models in clinical settings.

4.2 Related Work

Mining various kinds of patterns such as itemsets, sequences and graphs have remained a focus area of data mining research for a long time. Depending on the significance and value such patterns can add in their respective domains, they have been extensively studied for constructing high performance rules in decision-making systems (Cheng et al. 2007). Typically, extracted patterns have strong associations with class sensitive datasets, since they capture the underlying dynamic behaviour of a specific sub population, in the given class. This makes the use of patterns very suitable as potential variables or features, while building a robust classifier.

Methods described above are limited to the application of statistical models to determine associations between a risk factor and a disease outcome. However, real-world healthcare data is intrinsically too complex and massive to be limited to finding pair-wise associations. Thus, the identification of novel sequential and temporal patterns turns out to be a crucial advancement towards the development of state-of-the-art clinical informatics tools and techniques. Typically, such methods aim to determine statistically relevant patterns from discrete sequences of items. In this context, Klema et al (2008) identified frequent sequential patterns from a longitudinal dataset to map atherosclerisis risk factors to health outcomes. The mined patterns

were later used to create classification rules for predicting cardiovascular risk. Baralis et al (2010) employed the patient examination histories to derive significant closed sequential patterns to derive standard clinical workflows as well as workflow deviations, that were not compliant. Moreover, Berlingerio et al (2007) demonstrated further expressiveness in medical sequential patterns by mining event sequences along with the most frequently elapsed time intervals, between these events. Patnaik et al (2011) reported the extraction of sequential coding patterns from EHR data and followed up with the derivation of partial orders from the extracted sequences for generalizing patterns. Moreover, the LEGO approach of using automatically induced patterns as features in model construction was previously reported (Knobbe, Crémilleux, Fürnkranz & Scholz 2008).

Due to the longitudinal EHR's intrinsic temporal nature, Sachi et al (2007) proposed a method for mining temporal association rules from time series variables monitored during hemodialysis sessions. These temporal rules were mined based on prior definitions of temporal abstractions of interest. Later, researchers (Moskovitch et al. 2014) studied the problem of mining frequently occurring temporal patterns in abstracted EHR data and used Hoppner's representation (Höppner & Peter 2014) to define complex time-interval patterns for diabetic patients. A method for mining minimal time-interval patterns (Batal et al. 2013, Batal, Cooper, Fradkin, Harrison Jr, Moerchen & Hauskrecht 2016) that are most useful for predicting patients who are at risk of developing heparin induced thrombocytopenia (HIT), a life threatening condition that may develop in patients treated with heparin was later proposed. Among other, temporal methods, Wang et al (2013) proposed a non-negative matrix factorization framework using a convolutional approach for temporal pattern discovery in EHR data. This approach models each patient's record as an image matrix, where the x-axis corresponds to the time stamps and the y-axis corresponds to the event types.

Recent research by Dafe et al (2013) strongly reflected on the importance of capturing sequential relationships among discrete events for build-

ing robust sequential classifiers. The application of sequential patterns to create a feature space for learning models has also been reported (Fradkin & Mörchen 2015). However, the direct use of simple learning models on signal data makes them vulnerable to noise and tends to use statistical features that aggregate information based on windowing methods. Accordingly, such a process fails to capture interesting sequence based features. Moreover, the auto-integration of informative sequential patterns while creating learning models for ICU event prediction, largely remains an open area of research.

4.3 Methodology

In this section, we describe the detailed steps of the proposed ICU event prediction framework. We recapitulate from the previous chapter, with a brief description of the symbolic discretisation approach adopted for the given time series datasets. Next, the concepts related to the automatic construction of learning models using sequential contrast patterns are provided. Finally, we describe the integration of sequential contrast patterns with support vector machines and naive Bayes methods for predicting the class label of an unknown patient sequence (the test data instance) for classification purposes. The novelty of our integrated approach lies in the exploitation of contrast sequential patterns, within the given patient sequences, as features to build robust models. Towards this purpose, we demonstrate two feature construction approaches viz. existence of a given pattern in a sequence and pattern frequency. Later, the construction of two predictive models using differential sets of patterns (or features) is reported while predicting a given unknown instance.

Thus, we systematically investigate the use of sequential contrast patterns to construct a feature space of patterns, which is used to build better classification models in clinical settings. In addition, there is an inherent interpretable value in identifying discriminative sequential patterns for disease specific sub-populations.

4.3.1 Data discretisation

To facilitate the processing of pattern mining algorithms on temporal patient data comprising real-valued continuous representations, we employ the symbolic aggregate approximation (SAX) (described in Chapter 3) (Lin, Keogh, Wei & Lonardi 2007, **?**). SAX transforms a real-valued time series into a piecewise aggregate approximation (PAA) representation, which is converted to a symbolic string. As claimed by the authors, the advantages of SAX involve that of dimensionality reduction and lower bounding. As a result, due to the nature of physiological time series generated over a number of days, and their importance in determining critical conditions, SAX provides a proper platform to create efficient indexing and pattern mining algorithms for medical purposes.

4.3.2 Mining sequential contrast patterns

The discovery of sequential patterns is associated with the mining of transactional data to identify significant ordered sequences of items. Existing research demonstrates numerous applications of sequential pattern extraction in various domains (Mooney & Roddick 2013, Shen, Wang & Han 2014). Among many such applications, elegant sequential pattern discovery solutions are required in the context of timestamped sequences. Thus, given a set of well-defined training sequences, representative sequential patterns can be derived indicating high frequency supports in the corresponding training dataset. Among these studies, contrast pattern mining in supervised classification problems was initially addressed in the context of Emerging patterns (Dong & Li 1999). Earlier studies have reported the extraction of distinguishing sequential patterns (Ji et al. 2007). Typically, a distinguishing sequential pattern is defined as a subsequence, which satisfies the multiple algorithmic preconditions of user-defined maximum and minimum frequency supports for two differently labelled groups in a given dataset. Relevant definitions to sequential contrast patterns have been reported in Chapter 3.

For discovering the set of all contrast sequential patterns, we make use of the ConSGapMiner technique (Ji et al. 2007), proposed earlier for the extraction of minimal distinguishing subsequences (MDS), where gap constraints are defined by the user. The method employs the depth first search (DFS) technique for generating the set of candidate contrast sequences. Towards this purpose, a lexicographic sequence tree (LST) is grown (Ji et al. 2007). Further details on the technique have been described by Ji et al (Ji et al. 2007).

Thus, a sequential contrast pattern is accepted if it satisfies the user-defined support constraints. The ConsGapMiner approach allows us to restrict the generation of redundant patterns, making it computationally efficient.

4.3.3 Integrating sequential patterns for model construction

The use of frequent patterns for classification purposes have earlier been adopted for various applications (Li, Han & Pei 2001). In these cases, classifiers were mainly based on mining association rules in a supervised setting, also known as classification rule mining. This was followed by a selection of important rules by ranking them. Later, construction of a feature space using frequent patterns was utilized for discriminative pattern mining (Cheng, Yan, Han & Philip 2008). In the current context, we utilize sequential contrast patterns as features to build classification models in two distinct ways, as described next.

In conventional methods, if a test data instance satisfies one of the discovered patterns, then that instance is interpreted as satisfying a rule based on the corresponding pattern. In the context of sequence based training data, order information between contiguous elements can be exploited for robust classification or prediction of sequences for supervised learning applications. These patterns are known as sequential patterns, where informative sequences are derived using frequency measures like absolute or relative fre-

quency support, within the training data (Li et al. 2001). Later, the extracted set of sequential patterns are used to correlate a given test sequence with an outcome. Towards this purpose, the existence of individual sequential patterns is tested to make an outcome prediction or test instance classification.

Consider the set of sequential patterns as $P = \{Pt_k\}$, where $k = 1, 2, ..., m$, for a binary class labelled dataset $D = \{X_i\}$ such that $i = 1, 2, ..., n$, X_i represents a specific sample or data point. Now, if a sequential pattern Pt_k is present in a given sample X_i, then the binary valued feature corresponding to the given pattern is set as 1. The absence of Pt_k in a sample is encoded as a 0 for the corresponding feature. Hence, the set of sequential patterns and the input dataset is utilized to generate a transformed dataset having $|P|$ binary features and $|D|$ samples.

An alternative approach is also used to create a feature space by employing the relative support of patterns in differently labelled groups in the dataset. Here, we consider the relative support of a sequential pattern Pt_k for populating a feature value, provided the corresponding pattern is present in the given instance. Thus, if the corresponding pattern is absent from the given instance, the feature value is set to a 0. So if a sequential pattern Pt_k is present in a given instance X_i, then the corresponding feature is set to the frequency support $sup(Pt_k)$.

A simple example is used in Figure 4.1 to demonstrate the two ideas of feature space construction.

After obtaining a set of sequential patterns given by $P = \{AB, AC, AD\}$, each of these patterns is converted to a feature in the transformed dataset. In the first case, for example, the feature vector corresponding to the sample $ABCB$ is given by $< 1, 1, 0 >$ due to the presence of the patterns AB and AC in the given instance. The binary valued vector indicates the presence or absence of patterns at the corresponding feature positions. In the second case, $ABCB$ is transformed to $< 2, 1, 0 >$. This is because AB and AC are both present in the given instance and have a frequency support of 2 and 1 for the positive class. Finally, the transformed binary valued dataset is provided

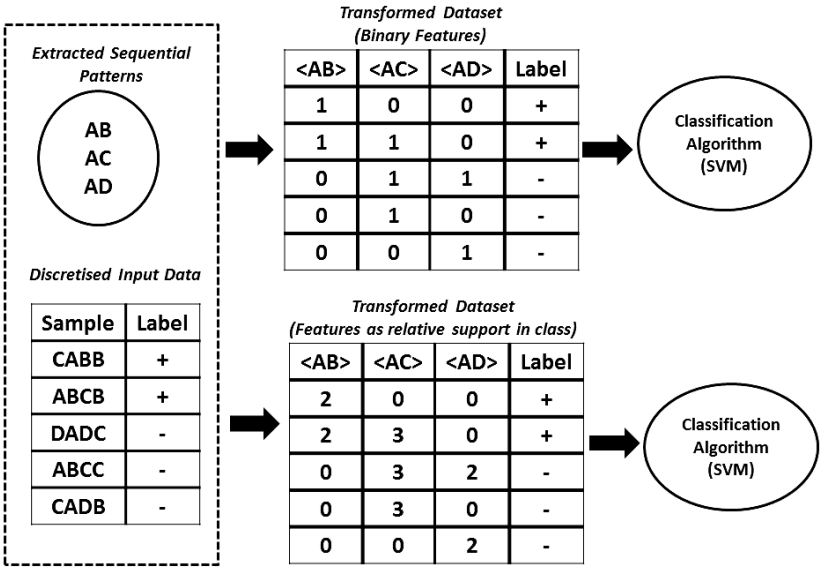

Figure 4.1: Transforming sequential patterns to binary or frequency based features.

as input to a classification algorithm such as SVM (support vector machines) (Cristianini & Shawe-Taylor 2000), Random Forests (Breiman 2001), and Naive Bayes (NB) (Friedman, Geiger & Goldszmidt 1997) for constructing a robust prediction model.

For the construction of a robust model, we further employ two types of models to predict a given test instance. Generally, sequential contrast patterns are extracted, while constraining frequency supports in the positive and negative groups. This means the extracted set of contrast patterns is mostly favourable or predictive of the positive class. In other words, the discovered patterns have a greater propensity to identify a positive test instance than a negative instance. In contrast, if the support constraints are reversed for a training set, then the generated set of patterns are more predictive of the negative class.

In our approach, for a single iteration, we thus obtain two groups of contrast patterns for feature transformations viz. contrast patterns favoring the positive group and patterns favoring the negative class. As described earlier,

this can be achieved by reversing the support constraints during the mining operation. Consequently, for each iteration, two sets of contrast patterns are obtained, i.e positive and negative. Subsequently, two pattern sets allow us to build two different prediction models to predict a given test instance.

Finally, the prediction of a given test instance is achieved by a comparison of probability estimates from the positive and negative model. In this case, the model reporting a higher likelihood estimate is used to assign the corresponding class label, to the unknown test instance.

4.4 Results and Discussions

The MIMIC-II database is a publicly accessible resource, subject to an appropriate NIH certification, which consists of >30,000 ICU patient records and has been traditionally employed for demonstrating the performance of novel algorithms for critical care applications. The patient records include numerous clinical variables such as laboratory test values, physiological measures, textual notes, medication records and physiological waveform signals, mapped to each patient identifier with a unique value.

For our current experiments, we employed two case studies in clinical events prediction. These are related to the prediction of an acute hypotension event (AHE) and hospital mortality of a patient in ICUs.

4.4.1 Dataset description

An acute hypotensive episode (AHE) is defined as a period of 30 minutes or greater, when 90% of the mean arterial pressure (MAP) readings are less than 60 mmHg. AHE datasets were used from the Physionet 2009 challenge directory and the study was approved by appropriate institutional review boards. They consisted of 2 major groups of patients viz. H and C, where H indicates the occurrence of an AHE in the forecast window and C indicates no occurrences of AHE in the forecast window. The groups H and C were further subdivided into H1, H2 and C1, C2. H1 describes patients who

received pressor medication (15 samples). H2 reported patients not receiving pressor medication (15 samples). C1 indicated patients having no AHE during complete hospital stay (15 samples) and C2 provided patients having AHE before or after the forecast window (15 samples).

In our experiments, we only consider the MAP (mean arterial pressure) time series for each patient record. The prediction tasks consisted of the following two events -

- Event I: AHE Risk classification of test patients between H1 and C1 (10 samples)
- Event II: AHE Risk classification between H and C (40 samples)

Event 1 thus helps in the prognosis of pressor medication resistant AHE. On the other hand, event 2 is aimed towards developing AHE predictors for patients at risk.

Additionally, we also extracted extended AHE datasets using the following clinical inclusion/exclusion criteria from the MIMIC-II database.

- a 30 to 60 minutes observation window
- a 30 minutes forecast interval where the ICU event occurs
- a time interval gap of 60 to 120 minutes separating the observation and forecast windows.
- ICD-9 code for hypotension (458.0 - 458.9).

The present problem is formulated as described by Figure 4.2. I_O indicates the observation window for a record and the class label is decided by the predicted occurrence of an ICU event in I_F. Generally, a time lag called T_X is considered since we make an event prediction for the future time window 1 or 2 hours ahead.

Using the MIMIC-II database, we originally compiled 254 records for AHE and 274 segments for normotensive records for the extended datasets.

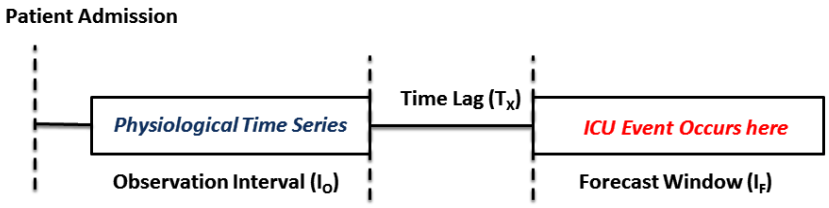

Figure 4.2: The ICU Event Prediction Problem

Among these the training data consisted of 370 records (178 H and 192 C) and the test dataset comprised 158 records (76 H and 82 C).

For the second case study, we focus on mortality, prior to hospital discharge, as a patient outcome for developing sequential patterns based predictive models. Traditionally, mortality prediction modelling has been carried out using simplified models such as the Simplified Acute Physiology Score (SAPS), which do not deliver sufficient precision for event predictions at the individual patient level. The dataset employed was obtained from the Physionet 2012 challenge, which focused on the problem of patient-specific mortality prediction (Silva, Moody, Scott, Celi & Mark 2012).

Towards this purpose, we used a dataset extracted from MIMIC II, comprising of 4,000 patient stays in the ICU lasting at least 48 hours. The datasets were formatted as time-stamped measurements for 37 distinct variables like urine output, white blood cell count, temperature, blood pressure etc. Among these, we employed the mean arterial blood pressure (MAP) of the patient and the respiration rate of a patient, for mining sequential contrast patterns. Moreover, the in-hospital death of a patient was adopted as the prime hospital mortality outcome variable for classification. Thus, a positive outcome indicated that the patient died in the hospital before discharge.

4.4.2 Classification results

A number of simulations were performed using various parameters like subsequence length, alphabet size and maximum gap. A 2-fold cross validation

(CV) was performed using the larger training dataset consisting of 60 samples (30 Hs and 30 Cs), thus reporting a CV accuracy of 94.9%. Optimal performances were obtained using a maximum gap of 3, subsequence length of 10 and alphabet of cardinality 5.

Table 4.1: Physionet AHE 2009 Test Prediction Results

	5 minute MAP averages	GRNN	SVM	Patterns SVM_B	Patterns SVM_F
AHE- Event I	100%	100%	100%	100%	100%
AHE-Event II	83%	92.5%	75%	92.50%	92.50%

As shown in Table 4.1, statistically computed averages of 5 minute MAP windows did not perform very well (Chen et al. 2009). However, good results were obtained when statistical averages of 5 minute diastolic arterial blood pressure windows were considered. Henriques and Rocha et al (Henriques & Rocha 2009) discussed the high performances of generalized regression neural networks (GRNN) and acknowledged the final model's dependency on parameter tuning. In comparison, we demonstrate the use of a pattern-based SVM model on the given test sets, which performed equivalently as the GRNN results. As noted, feature sets using binary values as well pattern frequency demonstrate similar performances.

To augment the Physionet 2009 results with further large scale studies, we also performed similar experiments with the AHE extended datasets. Our results for the extended datasets are shown in Table 4.2. Here, we include 4 datasets for each combination of observation interval (I_0) with a time lag (T_x). Our comparisons show that considering a model based on sequential patterns, which are transformed to features, have the ability to provide significant performance improvements. Thus, test predictions using pattern-based models demonstrate highest accuracies in the range of 83-85.8%, while predicting ICU events like acute hypotension.

In addition to using sequential patterns for predicting AHE events, we

Table 4.2: MIMIC-II Hypotension Test Prediction Results

	SVM	Pattern-based SVM (Binary)	Pattern-based SVM (Frequency)
AHE-Extended-I ($I_0 = 30mins, T_x = 60mins$)	71.30%	83.54%	83.54%
AHE-Extended-II ($I_0 = 60mins, T_x = 60mins$)	71.30%	85.80%	85.80%
AHE-Extended-III ($I_0 = 30mins, T_x = 120mins$)	72.70%	82.20%	82.20%
AHE-Extended-IV ($I_0 = 60mins, T_x = 120mins$)	70.10%	81.00%	81.00%

also employed a similar approach to predict patient mortality. In order to estimate the mortality prediction performance of our models, a 5 fold cross validation process was employed. Each iteration involved redeveloping the model for each of the randomly sampled folds of training data, followed by evaluating the predictive performance on the out of sample subset. Finally, the performance of the method was tested using the sensitivity measure. In this context, sensitivity was the standard metric employed to compare model performances for the Physionet 2012 mortality prediction challenge (Silva et al. 2012). Our results using MAP and respiration rate, are as given in Table 4.3.

For the baseline SVM, Random Forests and Naive Bayes models, the feature set was constructed using mean and median measures for sliding windows of size 10, over the given time series for each patient. Our experiments using sequential contrast patterns as features with the same baseline classifiers like SVM, Random Forests, and Naive Bayes, demonstrated improved sensitivity performance. Particularly, the use of a Random Forests classifier using 100 trees showed the best performance when used with frequency based features.

As demonstrated by the described studies, sequential contrast patterns

Table 4.3: 5-fold cross validated performances for Mortality Prediction

	Mean Arterial Pressure	Respiratory Rate
SVM	0.41	0.34
NB	0.45	0.39
$Patterns(SVM)_{Binary}$	0.47	0.44
$Patterns(SVM)_{Frequency}$	0.47	0.44
$Patterns(NB)_{Binary}$	0.51	0.44
$Patterns(NB)_{Frequency}$	0.51	0.44
$Patterns(RF)_{Binary}$	0.51	0.44
$Patterns(RF)_{Frequency}$	0.51	0.53

based learning models report better accuracies in comparison to using simple statistical features or just simple learning models. Moreover, winner results in Physionet 2009 challenge demonstrate similar performances with neural network based multimodels (Henriques & Rocha 2009). A similar large scale study carried out by Lee and Mark(Lee & Mark 2010), employed neural networks (NN) on hypotensive datasets extracted from MIMIC-II. Our results indicate performances similar to the NN results.

For the mortality prediction challenge, our results tend to demonstrate better performances in comparison to previous methods based on logistic regression (Bera & Nayak 2012), two-layer neural networks (Xia, Daley, Petrie & Zhao 2012) and a clinical scoring system like the SAPS-I (Silva et al. 2012).

Despite, the use of complex NN based methods in the models reported for both case studies, these are considered to be very complex for interpretation in the context of clinical scenarios. In contrast, our models are based on features, which are symbolic clinical sequences and easily interpretable.

The important aspect of our study is its reliance of constructing features from patterns for model development. Our approach clearly demonstrates the importance of learning good features that capture dynamic behaviour. We demonstrate that useful and robust clinical models can be constructed

when feature learning methods are integrated with learning models.

The experimental results for two ICU event prediction problems, indicate that automatic integration of sequential contrast patterns as features with learning models provide improved patient risk predictions. We note that for the same physiological variables, baseline models like SVM and Naive Bayes report lower performances. However, the CVA performance tends to be higher when models are built using sequential patterns as features.

Feature construction approaches based on sequential contrast patterns tend to inherently capture a patient's signal trajectory towards a critical ICU event. Thus, discrete sequences encoding significant physiological episodes after ICU admission, are able to capture a dynamically evolving patient state. In this context, using discrete episodes to learn interesting features and then using them in conjunction with learning models to predict ICU events, hold significant value both in terms of clinical interpretation of episodes as well as for the construction of robust prediction models. In addition, sequential contrast patterns are also easily interpretable , since the most significant features also correspond to investigating a set of episodic sequences, which may have significance from a clinical perspective.

An important consideration of the approach also involves employing the propensity of patterns to favour a positive or a negative class. Pattern mining techniques tend to extract sequences based on guidance by user defined support constraints. Hence, a better prediction is achieved, if models based on oppositional pattern sets are compared. Thus, the pattern-based model having greater propensity towards a given instance has a higher likelihood estimate.

Hence, our current results show that integrating sequential contrast patterns using classification models, allows us to capture interactions among discrete patterns of physiological variables, which are useful for predicting labels for patient sequences.

4.5 Conclusion

The present study investigated the effectiveness of an integrated pattern mining framework, where sequential patterns were used to build a pattern or feature space. Thus, each pattern was encoded as a binary valued and frequency based attribute in the transformed dataset. Subsequently, the dataset was provided as input to learning models like SVM, Random Forests and Naive Bayes for building robust sequential patterns-based classifiers. To demonstrate the effectiveness of pattern sequences as features, we compared our method with the traditional SVM, Random Forests, and Naive Bayes counterparts using generalised statistical features. Our results demonstrate that the learning models, which account for additional information in the form of sequence contrast patterns, tend to perform well in comparison to models not exploiting such information. This is because sequential patterns are able to capture dynamic behavior of a physiological signal as compared to statistical features which provide a static snapshot. Specially, our approach demonstrates the importance of training ICU classifier models using informative sequential patterns, in addition to conventional clinical measures. Hence, the recommended framework employing sequential contrast patterns and their feature transformations, can provide proactive ICU care systems a novel clinical pattern discovery platform for the improvement of patient outcomes. Finally, finding relevant sequences of symbolic events that predict ICU conditions can contribute to the investigation and development of cause-effect hypotheses, which are good candidates for investigations in clinical settings.

Chapter 5

Septic Shock Prediction for ICU Patients via Coupled HMM Walking on Sequential Contrast Patterns

In this chapter, we build on top of sequential contrast pattern mining foundations described in the previous chapters, and introduce the novel transformation of patient time series to multivariate "time series of sequential contrast patterns". These multivariate pattern sequences are integrated with sequential model like coupled HMMs to report predictive performances on an important problem of septic shock prediction.

5.1 Introduction

Septic shock is a critical complication arising from an infection, such that a systematic inflammatory response syndrome (SIRs) is triggered in the human body. Due to SIRs, tiny blood clots are formed. These clots block the oxygen and nutrients from reaching vital organs, leading to acute organs dysfunction and death.

Generally, sepsis treatment accounts for 10% of all ICU admissions (Angus & Van der Poll 2013). Over 19 million cases have been extrapolated to report worldwide incidence of SIRs (Rivers, Nguyen, Havstad, Ressler, Muzzin, Knoblich, Peterson & Tomlanovich 2001). Hospitalization due to SIRs, has overtaken those for myocardial infarction (commonly known as heart attack) (Yeh, Sidney, Chandra, Sorel, Selby & Go 2010). Currently, sepsis is the most expensive medical condition to be treated in hospitals and cost more than $20 billion in 2011, in US hospitals. Reportedly, these costs have been increasing by 11.9% annually (Torio & Andrews 2006).

It should be noted that the survival outcomes of sepsis treatments greatly depend on the early recognition of sepsis stages. Thus, discovering potential biomarkers for sepsis and septic shock, is an active area of research and a substantial literature of methods have been reported. Traditionally, SIRs is diagnosed using laboratory tests to determine the presence of factors like bacteria, low platelet counts, electrolyte imbalance etc. Complex patient health scoring systems like the Acute Physiology and Chronic Health Evaluation (APACHE II), and the recently developed targeted scoring systems, are employed to direct early interventions for sepsis (Rivers et al. 2001, Henry, Hager, Pronovost & Saria 2015). Past studies have employed features like patient demographics, heart rate variability, hypotension levels, and patient medical history, at the time of ICU admission to develop machine learning models using multivariate logistic regression, multilayered perceptrons, decision trees, principal component analysis and support vector machines (Capp & et al 2015, Lukaszewski & et al 2008, Gwadry-Sridhar, Lewden, Mequanint & Bauer 2009, Tang, Middleton, Savkin, Chan, Bishop & Lovell 2010).

A septic shock is identified by the occurrence of a hypotensive event (an extended drop in blood pressure), despite of a prior fluid resuscitation treatment. The mortality risk can increase dramatically, when patients progress from a sepsis situation to a septic shock. Therefore, the accurate identification of patients at risk of septic shock during the critical “golden hours” (Rivers et al. 2001) is crucial for the improvement of the traditional treat-

ment protocols in the current clinical care implementations. To this purpose, the direct use of machine learning models using static variables (commonly applied in current severity scoring systems), are not suitable for short-term predictions (for e.g., within 2 hours) of fast-evolving critical events in ICU settings. This is because, accurate forecasting of critical events, require dynamic temporal patient data. Subsequently, recent studies of temporal pattern mining methods for outcome prediction using Electronic Health Records have generated significant interest in the field of medical informatics and event predictions (Moskovitch & Shahar 2015, Batal et al. 2012, Yang, McAuley, Leskovec, LePendu & Shah 2014, Sacchi, Dagliati & Bellazzi 2015). However, the previous methods involved the use of highly curated datasets, involving invasively collected clinical variables and comparatively smaller population samples, in comparison to the requirements of large-scale clinical studies. Typically, simple machine learning models do not scale well in performance for large-scale databases of ICU patients.

For clinical research, randomised controlled trials (RCTs), are costly undertakings, requiring immense time and resources. In comparison, large-scale retrospective data driven studies can complement the mainstream clinical research, by providing effective testbeds for the development of interesting computational algorithms involving time-to-event prediction models using dynamic physiological data.

In the current study, we exploit the potential of commonly observed physiological measurement data for the early prediction of septic shock. The data was obtained from the measurements of the Mean Arterial Pressure (MAP), the Heart Rate (HR) and the Respiratory Rate (RR) in the MIMIC-II database (Saeed et al. 2011). Our approach discovers sequential contrast patterns from these physiological measurements, and then transforms the original training data into a time series of patterns. Later, we apply a coupled Hidden Markov Model (CHMM) to these *time series of patterns* for constructing the septic shock classifier. Later, for a given test sample (new patient), the classifier can estimate the probability of septic shock, happen-

ing in a future time window, after a half or one hour. Additionally, these sequential contrast patterns contained in a patient sample, can help to provide valuable insights about the physiological fluctuations that lead to septic shock events.

5.1.1 Contributions

The detailed contributions of this study are:

- A multi-variate contrast patterns mining based sequential modeling approach, in the form of a wrapper, is employed for ICU time series,
- The extracted contrast patterns are used to encode the dicretized training data, by creating a novel ordered sequence of contrast patterns for each patient, i.e as a time series of contrast patterns,
- A Coupled HMM is used to couple multiple channels of pattern sequences, for the prediction of high-risk septic shock patients, in a future time window

This study indicates that the integration of multi-channel contrast sequential patterns using CHMMs, can achieve accuracies competitive to earlier models. More importantly, our integrated approach makes it possible to simultaneously extract patterns that record dynamic patient information, and use these contrast patterns as inputs to sequential learning models for large-scale physiological data sets, from major healthcare database providers.

5.2 Related Work

In the past, numerous studies have been reported in the field of biomedical event prediction. An overview of recent research advances related to the use of pattern-based classification approaches for medical events prediction are reported, along with previous statistical modeling studies for septic shock prediction.

5.2.1 Previous studies in septic shock prediction

For the early prediction of septic shock, a number of previous studies have employed multivariate logistic regression models (Shavdia 2007, Hug 2009, Carrara, Baselli & Ferrario 2015). Thiel et al. (2010) performed regression tree analysis for multiple populations of greater than 13000 patients, for early prediction of septic shock risk among non-ICU patients. Decision trees were also employed by Gwadry-Sridhar et al. (2009) for 20 clinical variables, achieving nearly 100% predictive accuracy. Among the soft computing techniques, numerous wrapper-based feature selection and preprocessing methods, namely Zero-Order-Hold,and missing-value imputation techniques have been employed along with particle swarm optimization, fuzzy models, and neural networks, to improve septic shock classification performance (Vieira, Mendonça, Farinha & Sousa 2013, Ho, Lee & Ghosh 2012, Fialho, Celi, Cismondi, Vieira, Reti, Sousa, Finkelstein et al. 2013). Selecting appropriate clinical features turns out to an important concern for predicting cases of septic shock. Accordingly, Lukaszewski et al. (2008) demonstrated the efficacy of using blood sample measures and the expression levels of miRNAs, for learning a multilayered perceptron model to forecast septic risk and achieved 83% predictive accuracy. Additionally, Tang et al. (2010) employed principal component analysis (PCA) in combination with a non-linear support vector machine (SVM) on high resolution temporal physiological waveform datasets to achieve an 84% accuracy for predicting sepsis onset among 28 patients.

5.2.2 Pattern-based classification models for predicting biomedical events

Mining various kinds of patterns such as itemsets, and sequences have remained a focus area of data mining research, for a long time. In classification problems, discriminative patterns have strong associations with class sensitive datasets, making them suitable for use as potential variables or features, while building a robust classifier (Cheng et al. 2007).

Methods described in section 2.1, are limited to the application of statistical models to determine associations between a risk factor and a disease outcome. However, real-world healthcare data is intrinsically too complex and massive to be limited to finding pair-wise associations. Thus, the identification of novel sequential and temporal patterns turns out to be a crucial advancement towards the development of state-of-the-art clinical informatics tools and techniques.

In this context, Klema et al. (2008) identified frequent sequential patterns from a longitudinal dataset to map atherosclerisis risk factors to health outcomes. The mined patterns were later used to create classification rules for predicting cardiovascular risk. Baralis et al. (2010) employed the patient examination histories to derive significant closed sequential patterns to derive standard clinical workflows as well as workflow deviations, that were not compliant. Moreover, Berlingerio et al. (2007) demonstrated further expressiveness in medical sequential patterns by mining event sequences along with the most frequently elapsed time intervals, between these events. Patnaik et al. (2011) reported the extraction of sequential coding patterns from EHR data and followed up with the derivation of partial orders from the extracted sequences for generalizing patterns.

Due to the longitudinal EHR's intrinsic temporal nature, Sachi et al. (2007) proposed a method for mining temporal association rules from time series variables monitored during hemodialysis sessions. These temporal rules were mined based on prior definitions of temporal abstractions of interest. Toma et al. (2010) proposed logistic regression models for mortality prediction which integrated frequent temporal episodes constructed from patient time series of organ failure scores. Later, Moskovitch and Shahar (2014) studied the problem of mining frequently occurring temporal patterns in abstracted EHR data and used Allen's interval algebra representations (Höppner & Peter 2014) to define complex time-interval patterns for diabetic patients. Batal et al (2013) proposed a method for mining minimal time-interval patterns that are useful for predicting patients who are at risk

of developing heparin induced thrombocytopenia (HIT), a life threatening condition that may develop in patients treated with heparin. Among other temporal methods, Wang et al (2013) proposed a non-negative matrix factorization framework using a convolutional approach for temporal pattern discovery in EHR data. This approach models each patient's record as an image matrix, where the x-axis corresponds to the time stamps and the y-axis corresponds to the event types. Recently, Peek and Abu Hanna (2012) reported about past uses of time-to-event prediction methods for obtaining more fine-grained prognostic information in comparison to static data. In particular, the authors highlighted temporal modeling studies using hierarchical bayesian networks to predict organ failure (Peelen, de Keizer, de Jonge, Bosman, Abu-Hanna & Peek 2010), frequent temporal sequences to predict mortality(Toma, Bosman, Siebes, Peek & Abu-Hanna 2010), and temporal bootstraps to explore disease progression (Li, Swift & Tucker 2013). Here, time-to-event prediction models are associated with the estimation of the amount of time that passes prior to the occurrence of a clinical event.

Recent research by Dafe et al (2015) strongly reflected on the importance of capturing sequential relationships among discrete events for building robust sequential classifiers. The application of sequential patterns to create a feature space for learning models has also been reported by Fradkin et al (2015).

As described in section 2.1, the direct use of learning models on raw physiological data make them vulnerable to noise and tends to use statistical features that aggregate information based on windowing methods. Accordingly, such processes fail to capture interesting sequence based features. Moreover, the auto-integration of informative sequential patterns while creating learning models for ICU event prediction, remains an open area of research.

5.3 Materials and Methods

In this section, the detailed steps of the proposed septic shock prediction approach are presented. Initially, a brief description of the data discretization technique for the continuous time series data is provided. This is followed by the relevant definitions and concepts related to the extraction of sequential contrast patterns from the waveform datasets of two differently labelled groups of patients is discussed. Finally, we describe the integration of sequential contrast patterns using coupled hidden markov models (CHMMs) for predicting the class label of an unknown patient sequence (the test data instance) for classification purposes.

The novelty of our integrated approach lies in the exploitation of position information of sequential patterns (also described as the offset of a pattern), within a given patient sequence.

5.3.1 Discretisation of continuous time series

For discovering informative sequential patterns, an initial step requires the transformation of real-valued timestamped data to discretized representations (Syed, Stultz, Kellis, Indyk & Guttag 2010). This is a necessary step for the effective application of pattern discovery methods, since they operate on symbolic data types. Subsequently, the symbolic aggregate approximation (SAX) method (Lin et al. 2003) can be used to transform a time series signal into a discrete sequence, where a symbol is assigned to discrete intervals within the signal amplitude range.The SAX technique has emerged as a leading discretisation method, which has demonstrated its efficiency in numerous data mining applications by producing informative symbolic representations of large-scale time series data. SAX converts the given time series to a piecewise aggregate approximation (PAA) representation (Lin et al. 2003), which is later converted to a symbolic sequence. As described by Lin et al (2003), SAX characterizes the inherent properties of a time series data. Thus, an equiprobable distribution of symbols is obtained for the corresponding time

series (Lin et al. 2003). Algorithmic details on SAX discretization can be obtained in (Lin et al. 2003).

Following the discretization of time stamped data, data mining algorithms can be employed for discovering sequential patterns from disparate populations of sequence datasets. Previously, the discovery of emerging patterns from differently labelled groups of data was described by Dong and Li (1999). Emerging patterns are described as itemsets, which are constrained by user-defined frequency supports in differently labelled populations (or classes). Thus, given a dataset consisting of two classes, emerging patterns can be discovered, which frequently appear in the positive class compared to less frequency support in the negative class. Emerging patterns was later extended to identify emerging substrings in (Chan et al. 2003). Substrings are categorised as a special case of subsequences, where symbols in a substring have a gap interval of 0. However, sequential patterns of interest may not always be composed of consecutive symbols, within a given symbolic sequence. Accordingly, numerous algorithms have been reported for realizing gap intervals between symbols in a sequential pattern (Xing et al. 2010, Ghosh, Feng, Nguyen & Li 2014, Ghosh, Feng, Nguyen & Li 2016).

In following sections, we initially describe the extraction of gap-constrained subsequences from differently labelled groups of training sequence data, based on our prior work related to the mining of sequential contrast patterns for the acute hypotension problem (Ghosh et al. 2016).

Sequential patterns

The discovery of sequential patterns is associated with the mining of transactional data to extract frequently occurring ordered sequences of items.

Let us consider a set of distinct items represented by I. A sequence S defined over I, may be written as $e_1 - e_2 - \cdots - e_n$, given that $e_p \in I$, such that $1 \leq p \leq n$. A sequence $S' = e_{p_1} - e_{p_2} - \cdots - e_{p_m}$ exists within another sequence $S = e_1 - e_2 - e_3 - ... - e_n$, such that $1 \leq p_1 \leq p_2 \leq \ \leq p_m \leq n$. For example, a subsequence XY is contained in $XAAY$, but not YX. Hence, the

sequence order of items in S' is maintained within S, however the individual items in S' are not necessarily consecutive in S.

Moreover, given the sequences, $S = e_1 - e_2 - \cdots - e_n$ and $S' = e_{p_1} - e_{p_2} - \cdots - e_m$, S' occurs in S if $1 \leq p_k \leq n$ and $e_k = e_{p_k}$ for all $1 \leq k \leq m$, and $p_k \leq p_{k+1}$ for $1 \leq k \leq m$. For example, given sequences $S = ACACBCB$ and subsequence $S' = AB$, S' occurs four times in S, at the positions given by $\{1, 5\}, \{1, 7\}, \{3, 5\}$ and $\{3, 7\}$.

For satisfying the condition of gap constraints between symbols, let there exist a sequence $S = e_1 - e_2 - \cdots - e_n$ and the occurrence information as $O = p_1, p_2, \ldots, p_m$ of a subsequence S'. If $(p_{k+1} - p_k) \leq g + 1$, then it is said that S' satisfies the gap constraint of g. Typically, a singular occurrence of a sequence with gaps, within a training data instance, is a necessary condition for satisfying the gap-constraint requirement, for that sequence within the instance. For example, if $g = 3$, then AB is a subsequence of $ACCB$, but not $ACCCCB$.

Mining sequential contrast patterns

Emerging patterns (EP) are described as itemsets, which are constrained by user-defined frequency support conditions in different classes (Dong & Li 1999). This means that for a dataset consisting of two classes, patterns satisfying the condition of high frequency support in the positive class and low frequency support in the negative class are known as emerging patterns. Thus, an EP having high support in one class and low support in the contrasting class is considered to be a discriminative pattern that is able to contrast between the two opposite classes. Accordingly, the strength of such a pattern is expressed by the ratio of frequency supports in both classes (also known as the growth rate of EP). Here, we begin with the identification of gap-constrained subsequences from differently labelled groups of training sequence data, based on the principles of frequency support.

Let there be $D = \{D_1, D_2, \ldots, D_n\}$ representing a set of training instances, S_P - a sequential pattern, and g is the gap-constraint. The cardinal-

ity of occurrences of S_P in D is given by $count_{S_P}(D, g)$, also known as the absolute frequency support of S_P within D. Suppose, there exists a user-defined cardinality threshold of α and S_P satisfies $count_{S_P}(D, g) \geq \alpha$, then S_P is a frequent sequential pattern in D, having a gap constraint of g.

Extending the above description, given two differently labelled sequence datasets D^+ (positive sequences) and D^-(negative sequences), we can maintain two cardinality thresholds α and β, and a maximum gap of g, where a sequential contrast pattern S_P needs to satisfy the conditions, as below.

(1) Positive Support: $count_{S_P}(D^+, g) \geq \alpha$

(2) Negative Support: $count_{S_P}(D^-, g) \leq \beta$

Thus, given D^+, D^-, α, β and g, mining of sequential contrast patterns consists of discovering all gap-constrained sub-sequences as sequential patterns, which satisfy (1) and (2).

The rationale behind the extraction of contrast patterns is associated with the growth rate of a pattern, which can be described as the ratio of a given pattern's support in D^+ over D^- (Dong & Li 1999). The growth rate of a pattern is intuitive from a clinical applications perspective. This is because the traditional objective in clinical trials, is oriented towards finding differences between the intervention and control population of patients. Thus, discovering patterns based on differences in their supports in the intervention and control populations, allow us to find sequential patterns that can explain the difference between two populations of data. Specifically, the use of α and β to compute a growth rate of a pattern, is similar to the odds ratio, which is an intuitive measure to clinicians for finding association between an exposure and an outcome. In this context, given that a particular clinical event has occurred, we find the odds of a patient having a specific sequential pattern.

Generating candidate contrast sequences

For discovering the set of all contrast sequential patterns, we make use of the ConSGapMiner technique (Ji et al. 2007), proposed earlier for the extraction of minimal distinguishing subsequences (MDS), where gap constraints are defined by the user. The method employs the depth first search (DFS) technique to generate the set of candidate contrast sequences. To this purpose, a lexicographic sequence tree (LST) is grown. In our case, an LST is a tree where each node contains a subsequence (refer Figure 1), with its positive and negative frequency supports. Typically, a child node is grown by extending the parent node's sequence, using a new item (or symbol) (Ji et al. 2007, Ghosh et al. 2016).

After a sequence node is generated, if it satisfies the conditions (1) and (2), then the sequence node is not extended further. This is because a supersequence of a potential sequential pattern that satisfies conditions (1) and (2), is not minimal (Ji et al. 2007). Hence, in order to reduce the generation of redundant patterns as well as minimize tree depth, the growth of sequences is restricted by a minimality condition.

Moreover, if a sequence node's positive frequency support is lesser than α (as specified in condition (1)) , then the concerned node is not extended further. This is because a supersequence of the current node is also infrequent (Ji et al. 2007). Later, gap-constraint satisfaction is verified by the application of a bitmap representation reported earlier for checking gap-constraints (Ayres et al. 2002, Ghosh et al. 2016). Finally, a post-processing step is also applied so that any supersequence of at least another shorter subsequence, is removed from the resulting set of contrast sequences.

An example of a LST is shown in Figure 1. Here, node *XXZ(2,1)* represents the sequence *XXZ* with 2 as positive and 1 as negative supports. A child sequence may be grown by extending the parent sequence with a unique symbol from the alphabet, based on a certain lexicographic order. Thus, given the present LST, whose alphabet is defined as I = X, Y, Z, XXZ has three children nodes as XXZX, XXZY and XXZZ. Subsequently each

nodes supports are computed from the positive (D^+) and negative (D^-) classes.

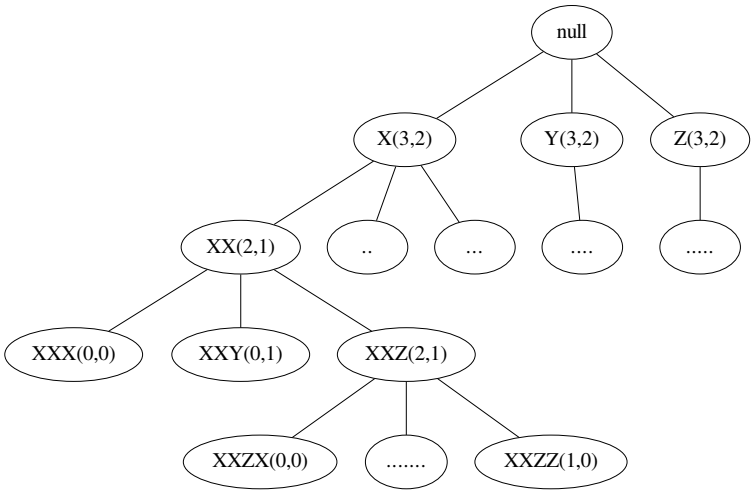

Figure 5.1: A Lexicographic Sequence Tree (LST) growing candidate sequences using 3 symbols as X, Y, Z

5.3.2 Discretised timestamped instance to sequential contrast patterns

In conventional methods, if a test data instance satisfies one of the discovered patterns, then that instance is interpreted as satisfying a rule based on the corresponding pattern. In the context of sequence based training data, order information between contiguous elements can be exploited for robust classification or prediction of sequences for supervised learning applications. These patterns are known as sequential patterns, where informative sequences are derived using frequency measures like absolute or relative frequency support, within the training data (Li et al. 2001). Later, the extracted set of sequential patterns are used to correlate a given test sequence with an outcome. To this purpose, the existence of individual sequential patterns is tested to make an outcome prediction or test instance classification.

However, a sequential pattern can also have a strong interpretive value

associated with its position information (described by the offset of a pattern) within a given discretised data instance. This means that an ordered set of patterns, occurring at different offset positions within an instance, is relevant for predicting an outcome for the given instance. Using offset values of the extracted sequential patterns in a discretised timestamped instance, allows us to transform the data instance to a meaningful episode consisting of consecutive sequential contrast patterns. As described in section 5.3.1, the set of contrast patterns is obtained from a simple and flexible sequential contrast mining technique. Following the extraction, the training dataset is transformed to a dataset of meaningful episodes, constructed by ordering sequential patterns based on their position, within an original training sequence. Sequences of patterns are then provided as input to a hidden markov model, which is an appropriate sequential learning method for exploiting a set of observations ordered in time.

Let us consider, $P = \{P_1, P_2, \ldots, P_n\}$ as a set of contrast sequences obtained from the D^+ and D^- training sequences, as described previously. Subsequently, a discretised instance of a training dataset, is transformed to a sequence of items or patterns from P. This is carried out by using a sliding window to incrementally move through the original discrete sequence. A sliding window of length equivalent to the longest item (pattern) in P is selected for our purpose. For each iteration of the sliding window through the sequence, the existence of item P_i (a sequential pattern) is tested in the corresponding segment of the sequence. This can be illustrated using Figure 5.2.

Let us consider $P = \{P_1, P_2, P_3\}$ as the set of sequential contrast patterns. In the first iteration of the sliding window, P_1 and P_2 are identified. This is followed by the detection of P_2, P_3 and P_1, in the second iteration. To determine the order information between two patterns, we employ the rule $pos[P_j]_1 < pos[P_k]_1$, where $pos[P_x]_1$ gives the position of the first symbol of a pattern P_x within a given sequence. Here $x, j, k \in \mathbb{N}$ and ≤ 3.

The above encoding procedure is repeated for each of the training se-

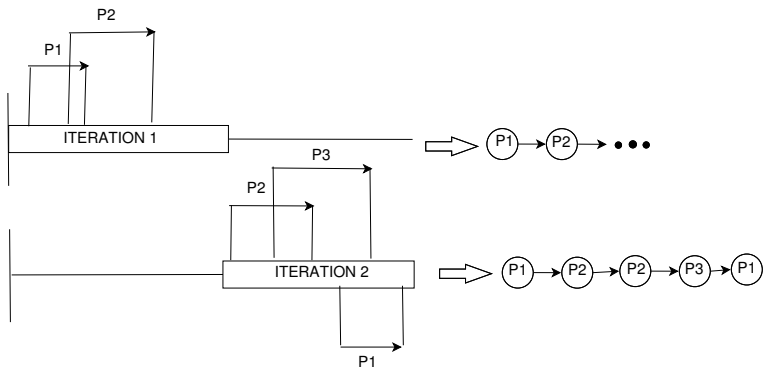

Figure 5.2: Encoding and transformation of a data instance to an ordered sequence of patterns

quences to obtain a transformed dataset, where each original sequence is thus encoded using an ordered series of patterns P_i. The transformed set of sequences is subsequently provided to an HMM (and CHMM) for learning its model parameters. Later, in the prediction phase, an unlabelled discrete test sequence is transformed to a pattern sequence using P (the contrast pattern set), which is provided as an input to the learned HMM for obtaining a probability likelihood estimate for the corresponding test sequence. Finally, the class label of the sequence is predicted to be positive, if the likelihood estimate is higher than a user-defined threshold.

The above process of transforming a single discrete sequence to an informative episode of patterns can be readily extended for multiple time series variables. For a multivariate sequence, a data instance is composed of multiple sequences, each representing a specific time series variable. For each of the given variables, we extract a set of sequential contrast patterns. Subsequently, the transformation of the multivariate training dataset is performed by encoding each variable sequence (of a data instance) using its corresponding contrast pattern set.

Coupled hidden markov models

CHMM extends the conventional form of HMM to multiple observation sequences or channels. Existing studies have employed CHMM in applications such as speech recognition, activity recognition, anomalous trading activities, medical events, disease interactions and fault diagnosis (Zhou, Chen, Dong, Wang & Yuan 2016, Audhkhasi, Osoba & Kosko 2013, Cao, Ou & Philip 2012, Masoudi, Montazeri, Shamsollahi, Ge, Beuchee, Pladys & Hernández 2013). In the current study, CHMM is used to integrate and model interactions between multiple physiological variables, each represented by a sequence of discrete observations. Accordingly, multiple HMMs are aggregated by enabling transitions between the discrete hidden states for each HMM. The topological structure of a CHMM is shown in Figure 5.3, where for example, two variables with corresponding channels are integrated.

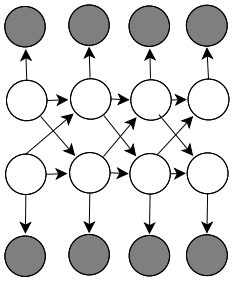

Figure 5.3: Topology of a two-channel CHMM.

Let us consider a generalised CHMM model with C parallel channels from $\{1, \ldots, C\}$. The set of states is given by $S^c = \{s_1^c, s_2^c, \ldots, s_{I_K}^c\}$, where I_K is the number of states and $c \in \{1, \ldots, C\}$. The set of observations is represented by $V^c = \{v_1^c, v_2^c, v_3^c, \ldots, v_{J_c}^c\}$, where J_c is the number of discrete observations. The state transition distribution $A^c = \{a^c_{i_1 i_2 \ldots i_C k_c}\}$, based on the generalised markov property, where each hidden node has C parent nodes (corresponding to C channels) from the previous time point, is given by

$$P(q_{t+1}^c = s_{k_c}^c | q_t^1 = s_{i_1}^1, \ldots, q_t^c = s_{i_C}^C) = a^c_{i_1 i_2 \ldots i_C k_c} \tag{5.1}$$

where $\sum_{k_c=1}^{I_K} a^c_{i_1 i_2 \ldots i_C k_c} = 1$.

The emission probability distribution in state s^c_i is given by $B^c = \{b^c_i(k)\}$, such that

$$b^c_i(k) = P(O^c_t = v^c_k | q^c_t = s^c_i) \tag{5.2}$$

where $\sum_{k=1}^{J_c} b^c_i(k) = 1$.

In equation (2), the identifiers c, i and k indicate a channel, a state and an observation, respectively.

The initial state distribution $\pi^c = \{\pi^c_i\}$ is represented as

$$\pi^c_i = P(q^c_1 = s^c_i) \tag{5.3}$$

where $\sum_{i=1}^{I_K} \pi^c_i = 1$.

Accordingly, each channel is described by the following HMM notation of parameters

$$\lambda^c = (A^c, B^c, \pi^c) \tag{5.4}$$

The final CHMM model can thus be denoted by

$$\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^C) \tag{5.5}$$

Similarly as conventional HMM, the three specific research areas for a CHMM include, (1) the classification of observation sequences, (2) inferring the sequence of hidden states which maximizes the sequence likelihood estimate, and (3) learning the parameters of the CHMM.

For classification, if we have C channels corresponding to C observation sequences, such that $o^c = o^c_1, o^c_2, o^c_3, \ldots, o^c_T$, we need to compute the probability of the given C sequences denoted by $P(o^1, o^2, \ldots, o^C | \lambda^1, \lambda^2, \ldots, \lambda^C)$. For inferring the hidden state sequence, given C channels, the final CHMM needs to determine the sequence of hidden states: $q^c = q^c_1, q^c_2, \ldots, q^c_T$ for each channel $c = 1, 2, \ldots, C$, such that the likelihood estimate is maximized for the given observation sequences. Finally, for model estimation, given C observation sequences $o^c = o^c_1, o^c_2, o^c_3, \ldots, o^c_T$ for each of the C channels, we need to optimize optimize the parameters of the CHMM model to maximize $P(o^l, o^2, \ldots, o^C | \lambda^1, \lambda^2, \ldots, \lambda^C)$.

Previously, various algorithms have been employed to solve the CHMM problem (Zhong & Ghosh 2002, Kristjansson, Frey & Huang 2000). For our implementations, we adopted the procedure described by Rezek et al (Rezek & Roberts 2000). Here, the CHMM with C channels was modified to construct a single channel large HMM. In this large single channel CHMM, each state is viewed as a cartesian product of states from the C channels and is given by $s = (s^1_{i_1}, s^2_{i_2}, s^3_{i_3}, \ldots, s^C_{i_C})$. Note that $s^C_{i_C}$ represents a discrete state from the C^{th} channel and $i_C \in s^C_1, \ldots, s^C_{I_k}$. Thus, $s^C_{i_C}$ a member of the set S^c for $c \in \{1, 2, \ldots, C\}$.

The above formulation leads to a total of $N = \pi^C_{k=1} I_k$ possible states for the HMM at every time instance. Accordingly, an $A = NXN$ matrix is formed, where each element denotes the probability of state transition from one state s to another state in the given HMM. Note that each state consists of C ordered components. According to this procedure, an observation for a given time step is a $CX1$ vector give by v. Here, $v = \{v^1_{k_1}, v^2_{k_2}, v^3_{k_3}, \ldots, v^C_{k_C}\}$, where $v^C_{k_C} \in V^c$, such that $c \in \{1, \ldots, C\}$. Thus, we have $M = \pi^C_{c=1} J_c$ possible observations, at a given time instance. Subsequently, an NXM matrix B can be defined to represent the observation probabilities of the final CHMM. This large HMM can now adopt the general structure given by $\lambda = \{\pi, A, B\}$.

Based on the above transformations, the aforementioned CHMM problems for model estimation and classification become the same as a single-channel HMM. To this purpose, we employ the generalised forward-backward algorithm for solving the classification problem (Rabiner 1989). For model estimation, we use the expectation-maximization algorithm (also known as the Baum-Welch method) to maximize $P(O|\lambda)$ to adjust model parameters for HMM (Rezek & Roberts 2000, Rabiner 1989).

5.3.3 Illustrative examples of CHMM walking on sequential patterns

To demonstrate the sequential patterns based CHMM technique, we consider two simple examples: (1) a single channel patterns based HMM (SCP-HMM) and (2) multi-channel patterns based CHMM (MCP-CHMM).

Single channel patterns based HMM (SCP-HMM)

In the following example, let us consider a set consisting of the patient mean arterial pressures (MAPs) for positive (D^+) class labels. Let the set of sequential patterns extracted after the contrast mining process be denoted by $P = \{P_j^i | j = 1 \dots n, i = 1 \dots m\}$, where i encodes the channel and j encodes the pattern, as shown in Figure 4. Due to the nature of contrast mining, the patterns listed in P have stronger support in D^+ than D^-. Thus, each pattern is encoded using a symbol P_j^i, where i indicates the number index of variables and j indicates the number index of patterns.

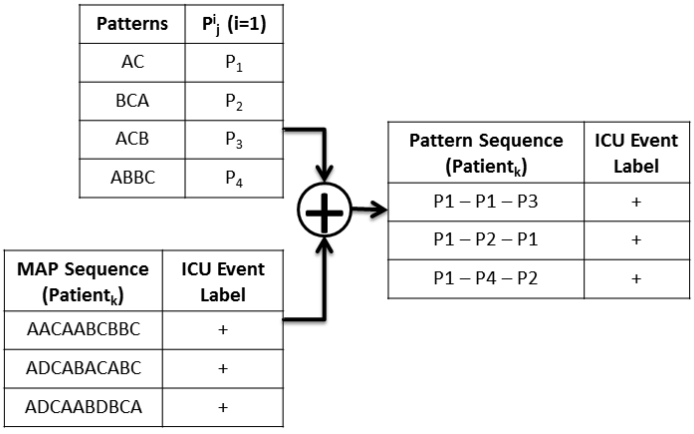

Figure 5.4: Encoding patient sequences using extracted patterns. P_j^i denotes a sequential pattern. Here, *i=1* indicates a single channel or variable. A patient MAP sequence such as $AACAABCBBC$ is converted to $P_1 - P_1 - P_3$. Finally, a new training set of pattern sequences is obtained.

Table 5.1: A indicates the state transition function for discrete states S_1 and S_2

State Transition Function ($\boldsymbol{A}$)	S_1	S_2
S_1	0.5	0.5
S_2	0.7	0.3

Table 5.2: B denotes the emission probability distribution for 2 states and 4 pattern observations

Emission Distribution ($\boldsymbol{B}$)	P_1	P_2	P_3	P_4
S_1	0.5	0.2	0.2	0.1
S_2	0.7	0.1	0.05	0.15

For an HMM with two discrete states S_1 and S_2, let us have the state transition and pattern emission probabilities are shown in Table 5.1 and Table 5.2. In Figure 5.5, the state transition diagram is illustrated with output emissions and their probabilities.

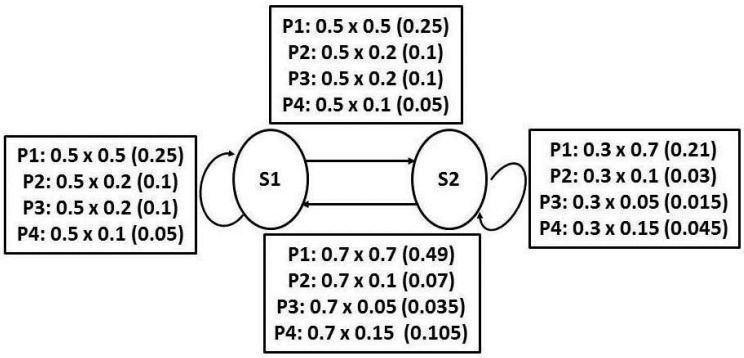

Figure 5.5: State transition diagram with output emissions and their probabilities

Based on the described HMM model, a pattern sequence $P_1 - P_1 - P_3$ (as shown in Figure 5.4) is computed to have a likelihood estimate of 0.72. Accordingly, we classify this instance to be a positive pattern sequence, since its probability likelihood is greater than the threshold of 0.5. Maximum

likelihood measures for each of the other pattern sequences are also estimated in a similar manner.

Coupled HMM for multichannel pattern sequences (MCP-CHMM)

For the example shown in Figure 5.4, we consider a single channel MAP sequence denoted by i in P_j^i. Thus, in the context of parallel channels, each variable like HR and RR can have their corresponding set of patterns denoted by P_j^2 and P_j^3, for the example in Figure 5.4.

Therefore, for a given patient instance having three sets of sequential patterns given by P_j^i, each variable (i.e MAP, HR or RR) sequence for a patient is converted to an ordered sequence of patterns P_j^i, where the channel $i = 1 \ldots 3$. Given the CHMM formulation, each discrete state for a particular channel now becomes a function of three states, based on the markov property. Thus, the state transition and emission probability functions can be realized, by mapping a permutation of three unique states (corresponding to each channel). This can be illustrated by the directed graph (DAG) shown as per Figure 5.6, where a single edge from the previous state in each channel enters the next state of another channel.

Here, S_j^i is a discrete hidden state for the channel i and j is the index of a state. Thus, figure 5.6 illustrates that the emission of a contrast pattern is probabilistically estimated by a a discrete state for that channel, which depends on three states at time t_{m-1}. Here, t_m indicates the current iteration at m for time t.

5.4 Evaluation

Our experimental plan begins with a description of the septic shock event prediction problem, followed by a brief description of the MIMIC-II database (our primary source for data collection). Next, we describe the clinical inclusion and exclusion criteria for the selection of patients. For baseline estimations, we employed SVM and HMM models on the continuous time se-

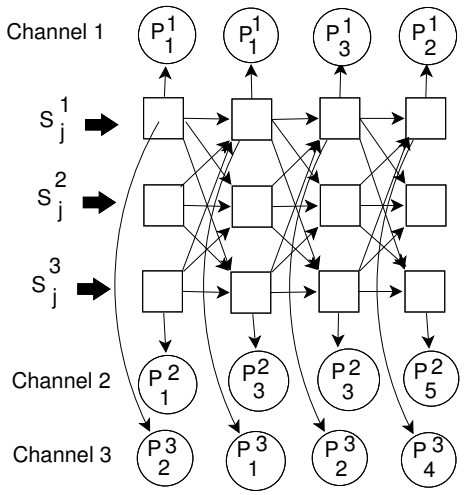

Figure 5.6: A coupled HMM topology for 3 channels. Here, P_j^i denotes a sequential pattern. Here, i indicates a channel and j corresponds to a specific pattern-id for a variable.

ries data for the given patients, using MAP (mean arterial pressure), HR (heart rate), and RR (respiratory rate). The baseline methods are denoted by SVM-MAP, HMM-MAP, HMM-HR, HMM-RR. Single channel patterns based HMM, for the three physiological variables are denoted by SCP-HMM-MAP, SCP-HMM-HR, SCP-HMM-RR. Finally, coupled HMM is employed, for both the multivariate continuous times series (CHMM) and multi-channel patterns (MCP-CHMM).

5.4.1 The septic shock prediction problem

Sepsis can be defined as a life-threatening condition occurring, due to a systemic inflammatory response syndrome (SIRs) triggered to fight an infection. Under such circumstances, SIRs is diagnosed, if two or more of the following critera are satisfied, namely abnormal body temperature (i.e >38 C or <36 C), higher heart rate (>90 beats per minute), respiratory rate >20 per minute, and abnormal white blood cell counts. In later stages, septic shock can be characterised by a systolic blood pressure (SBP) <90 mmHg, despite

of a fluid resuscitation treatment of >600 mL, one hour before (Bone, Balk, Cerra, Dellinger, Fein, Knaus, Schein & Sibbald 1992, Shavdia 2007, Ho, Lee & Ghosh 2014).

The problem of septic shock prediction can be simply illustrated by Figure 1.

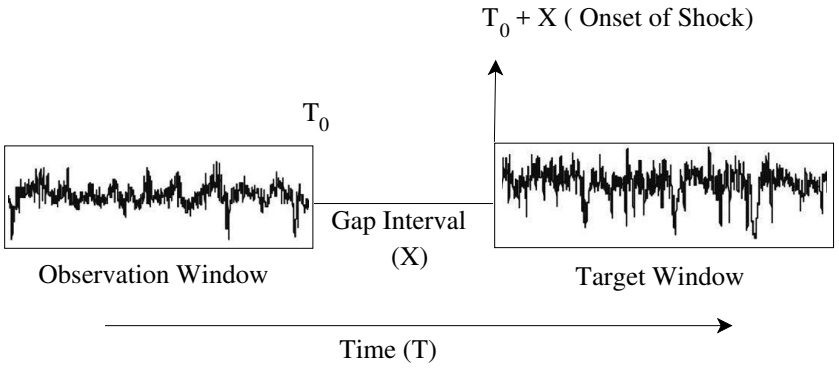

Figure 5.7: Observation and Target Windows with a Time Gap Interval

For this problem, we are given a test sample (e.g., a new patient) whose waveform data of a user-defined window of length 60 or 90 minutes have been observed and recorded till the time point T_0, the goal is to predict whether a septic shock will happen to this patient or not at a future target window of 30 minutes (namely, at the time window from $T_0 + X$ to $T_0 + X + 30$) through an HMM classifier. Usually, the observation and the target windows are separated by a user-defined gap interval X of 30 and 60 minutes. The classifier is constructed using a set of training data. In this work, the classifier (prediction model) is constructed on three non-invasive channels of waveform signals of the patients in the training set. The three channels of waveform data are the commonly measured MAP, HR, and RR for every patient. This research problem is important because it is an early prediction of septic shock at a future time window with a gap interval of a half or one hour between the observation and the forecasting time window.

5.4.2 The MIMIC II database

The MAP, HR, and RR waveform data used by this study were downloaded from the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) database which is a publicly available resource developed to support research in clinical decision support and critical care medicine (Saeed et al. 2011). MIMIC-II version 2.6 consists of clinical and waveform records for more than 30,000 ICU patients collected between 2001 and 2007. The electronic health database along with the waveform records, include numerous patient variables such as high resolution time-stamped physiological waveforms (e.g. blood pressure, heart rate etc.) and clinical variables (e.g. fluid input and output, laboratory tests, patient discharge notes etc.).

5.4.3 Selection of patients

As the clinical inclusion criteria, our current study considered adults (i.e >18 years of age) from the MIMIC II database. Each patient consisted of at least one hour of observations for mean aerterial pressure (MAP), heart rate (HR), and respiration rate (RR).

ICD-9 codings were employed to identify septic patients (995.91 or 995.92). Patients with septic shock were identified by examining their clinical chart records. The time of shock onset was determined using criteria used in (Shavdia 2007). Following from Shavdia et al (Shavdia 2007), we define a hypotension observation as any time point where systolic blood pressure (SBP) was <90 mmHg. Consecutive hypotension observations were then aggregated to define a hypotension region. Total fluid intake for one hour prior to the first hypotension observation was then calculated. Any hypotension region that registered a total fluid intake >600 mL was classified as septic shock, with onset defined as the start time of the hypotensive region. Such a definition for shock onset follows the standard definition from (Shavdia 2007). For our experiments, we only considered the first detection of a septic shock onset to construct our observation periods. Towards this purpose, a total

of 1,310 patients were diagnosed with sepsis or severe sepsis in MIMIC-II. Among these, 209 patients were diagnosed with a septic shock condition, given our inclusion criteria.

5.5 Prediction Results

The MIMIC-II database is a publicly accessible resource, subject to an appropriate NIH certification, which consists of >30000 ICU patient records and has been traditionally employed for demonstrating the performance of novel algorithms on benchmarked datasets for critical care applications. The patient records include numerous clinical variables such as laboratory test values, physiological measures, textual notes, medication records and physiological waveform signals, mapped to each patient identifier by a unique value.

5.5.1 Four data sets extracted from MIMIC-II

The sequential contrast patterns CHMM framework was applied to multiple septic shock datasets, based on the data descriptions provided in section 5.4.3. Accordingly, the total number of samples with sepsis (ICD9 code - 995.91 or 995.92) were found to be 1310. Among these, the number of patients which moved over to a septic shock condition (identified by ICD9 code 785.52) were found to be 209. Thus, our main patient dataset consisted of 209 positive instances and 1101 negative instances. Further, the MIMIC variables extracted for use were mean arterial pressure (MAP), heart rate (HR) and respiratory Rate (RR) for each of the extracted records.

Later, 4 datasets were constructed based on a combination of two factors as given below

- where the gap interval is 30 or 60 minutes, and
- where the observation window is 60 or 90 minutes. See 5.4.1.

Thus, we have 4 datasets, where each record is defined by a 30 or 60 minutes gap interval, following a 60 or 90 minutes observation window. The allotted time windows are standard references associated with short term ICU prediction problems and is similar to (Ho et al. 2012, Ho et al. 2014). For our experiments, we only considered the first detection of a septic shock onset to construct our observation periods. Generally, gap-time intervals and observation windows sizes used for the problem are standard and motivated from prior research (Ho et al. 2014, Lee & Mark 2010) for similar studies.

5.5.2 Cross-validation classification results on the four data sets

A number of previous studies have been carried out for predicting the risk of sepsis and septic shock. These studies largely focus on pre-selected sets of clinical patient features. As these features significantly differ from one study to another, there does not exist any accepted gold standard which we could adapt for evaluating the performances of the models. For this work, we employed multiple rounds of 5-fold cross validation to assess our models' performance.

For each of these four datasets, the 5-fold cross validation was performed for three rounds. At each round, the 5 different folds were randomly selected as a test set to obtain the corresponding 5-fold cross validation classification accuracy (CVA). In each round, we also used the records' observation window to train the model of support vector machines (SVMs), single-variable hidden markov models and coupled hidden markov models. Our three rounds of 5-fold cross validation results for each model are presented at Table 5.3 to 5.6 for the four datasets.

In detail, each Table (5.3 to 5.6) records the 5 fold cross validation classification accuracy performance among 9 different types of variable and learning model combinations. These include a machine learning SVM for the MAP variable for estimating baseline performance followed by single channel HMM models for each of HR, RR and MAP respectively. These models are then

Table 5.3: A comparison of different models using 5-fold cross validation classification accuracy at $t_{gap} = 60$ mins and $t_{obs} = 60$ mins

	Round 1	Round 2	Round 3
SVM-MAP	77.2	82.1	78.3
HMM-MAP	84.3	83.7	84.2
HMM-HR	75.1	82	81.1
HMM-RR	74.4	80.1	77.9
SCP-HMM-MAP	85.1	82.2	85
SCP-HMM-HR	80.2	79	81.1
SCP-HMM-RR	79.1	80.1	77.9
CHMM	84.3	83.7	85
MCP-CHMM	**85.1**	**87.1**	**85.4**

compared to the HMMs of sequential contrast patterns for single variables (HR, BP and RR). Finally, we consider CHMM models using both the continuous multivariate and discretised sequential contrast patterns. Also, each of the four tables progressively reports the CVA performances with different combinations of observation window length and gap interval (for each of the four dataset, respectively).

Finally, 5 fold cross-validation accuracy results using repeated resampling of each of the 4 different datasets for CHMM and MCP-CHMM are reported in Table 5.7, for 5 separate rounds. Variances across multiple rounds for each dataset is also shown.

5.5.3 Predicting coupled discrete sequences using HMMs: An illustrative case study

A case study is used to demonstrate the prediction of a specific multivariate test sequence using our proposed CHMM framework. The given multivariate instance is composed of three variables, namely the mean arterial pressure, heart rate and respiratory rate. Initially the sequences of continuous time

Table 5.4: A comparison of different models using 5-fold cross validation classification accuracy at $t_{gap} = 30$ mins and $t_{obs} = 60$ mins

	Round 1	Round 2	Round 3
SVM-MAP	77.2	82.4	77.1
HMM-MAP	84.7	83.7	84.2
HMM-HR	75.5	82.0	81.1
HMM-RR	74.4	81.0	77.9
SCP-HMM-MAP	85.5	82.7	83
SCP-HMM-HR	80.2	79.1	81.1
SCP-HMM-RR	79.1	80.1	76.9
CHMM	85.0	84.7	85.3
MCP-CHMM	**86**	**87.1**	**84.8**

Table 5.5: A comparison of different models using 5-fold cross validation classification accuracy at $t_{gap} = 30$ mins and $t_{obs} = 90$ mins

	Round 1	Round 2	Round 3
SVM-MAPP	77.2	82.1	78.3
HMM-MAP	84.3	83.7	84.2
HMM-HR	75.1	82	81.1
HMM-RR	74.4	80.1	77.9
SCP-HMM-MAP	85.1	82.2	85
SCP-HMM-HR	80.2	79	81.1
SCP-HMM-RR	79.1	80.1	77.9
CHMM	84.3	83.7	85
MCP-CHMM	**85.7**	**85.2**	**85**

Table 5.6: A comparison of different models using 5-fold cross validation classification accuracy at $t_{gap} = 60$ mins and $t_{obs} = 90$ mins

	Round 1	Round 2	Round 3
SVM-MAPP	77.2	82.1	78.3
HMM-MAP	84.3	83.7	84.2
HMM-HR	75.1	82	81.1
HMM-RR	74.4	80.1	77.9
SCP-HMM-MAP	85.1	82.2	85
SCP-HMM-HR	80.2	79	81.1
SCP-HMM-RR	79.1	80.1	77.9
CHMM	84.3	83.7	85
MCP-CHMM	**85.1**	**85.5**	**85**

Table 5.7: 5-fold cross validation classification accuracy on CHMM and MCP-CHMM for 5 rounds of repeated re-sampling. g - gap interval size, o - observation window size

Method	Gap, Obs.	R_1	R_2	R_3	R_4	R_5	σ^2
CHMM	g=30,o=60	84.3	83.7	85.0	84.7	85.1	0.26
	g=30,o=90	85.0	84.7	85.3	85.3	85.0	0.05
	g=60,o=60	84.3	83.7	85.0	83.7	79.3	4.03
	g=60,o=90	84.3	83.7	85.0	85.0	85.7	0.46
MCP-CHMM	g=30,o=60	85.1	87.1	85.4	84.7	85.0	0.72
	g=30,o=90	86.0	87.1	84.8	85.1	85.1	0.70
	g=60,o=60	85.1	85.2	85.0	84.8	85.1	0.01
	g=60,o=90	85.7	85.5	85.0	85.1	85.0	0.08

series data were converted to their discretised representations. Thus, each observation test sequence consists of 60 time points for each variable, where the time-stamped value is converted to a discrete symbol belonging to a set of pre-defined symbols. Each of these discretised sequences was then truncated internally, by reducing a consecutive sequence of 3 similar symbols to 1 symbol. This is because long runs of similar symbols lead to significantly more computational time and have less interpretative value in a clinical context. Some of the sequential contrast patterns that appeared in the given discretized training signals are as shown in Table 5.9. For each of the discretised signals, we employed a sliding window to move through the given sequence. Subsequently, the existence of patterns in the consecutive passes of the window was used to build a pattern sequence, as shown in Table 5.8. Each pattern is uniquely encoded for each variableś contrast pattern set. For example, the M_1 pattern uniquely identifies the sequence 6<7<6 only for instances of the mean arterial pressure. In Table 5.9, we list some of the prominent contrast sequences, which can be found within the variable sequences of Table 5.8. Note that Table 5.9 is a subset of the larger set of contrast patterns. In addition to the contrast patterns, we also consider a "dont-care" pattern denoted by X. Multiples of "dont-care" patterns are inserted at the end of a pattern sequence, so that every variableś pattern sequence has the same length.

Table 5.8: A multivariate (MAP, HR, RR) discrete patient sequence composed of an ordered series of contrast patterns

max width=

Discretised Sequence	Variable Name	Sequence of Patterns
6-6-6-6-7-6-6-6-6-6-6-6-6-6-7-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5	MAP	M_1 - M_14 - M_1 - M_14 - M_5 - M_9 - X - X
4-5-5-5-5-5-4-6-6-6-6-6-5-6-6-6-6-6-6-6-5-6-6-6-6-6-6-4-6-6-6-6-5-6-6-6-6-6-6-5-6-5-5-5-5-6-6-6-6-6-6-6-6-6-6-6-6-6-6	HR	H_7 - H_2 - H_9 - H_9 - H_8 - H_2 - H_9 - H_9
6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6-8-8-8-6-6-6-6-8-8-8-8-6-6-6-6-6-6-6-6-6-6-6-6-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5-5	RR	R_7 - R_7 - R_8 - R_13 - X - X - X - X

Table 5.9: Visualizing contrast sequence patterns matching the three variables MAP, HR and RR

MAP patterns	Pattern-id	HR patterns	Pattern-id	RR patterns	Pattern-id
6 <7 <6	M_1	4 <6 <6	H_2	6 <6 <8 <6	R_6
6 <5 <5 <5	M_5	5 <4 <6	H_7	6 <6 <8 <8	R_7
6 <6 <5 <5	M_9	6 <4 <6	H_8	6 <6 <6 <8	R_8
7 <6 <6 <6	M_14	6 <5 <6	H_9	8 <8 <6 <6	R_13

The likelihood for this discrete multivariate test sequence was estimated at the level of 0.71 by CHMM. As we assumed the likelihood threshold for differentiating between a positive and negative classification as 0.5, the given multivariate test sequence was predicted to be a positive case, i.e the given patient multivariate sequence was classified as 'having a higher risk for the occurrence of a septic shock'. It is also worth noting that converting multivariate discrete sequences to a multivariate time series of contrast patterns, allows an HMM to exploit the order (or offset) information among the patterns, which are crucial for making a robust HMM based prediction.

5.5.4 Discussion

The experimental results have evidenced that integrating sequential contrast patterns with CHMM models can help provide an assessment of septic shock risk. The current study mainly explores the use of a novel data transformation technique by chaining discrete sequential contrast patterns to represent a data instance, before input to any sequence modelling algorithm (for e.g. HMM). To this purpose, for evaluation, the cross validation accuracy results after chaining of discrete patterns (MCP-HMM) were compared to baseline models of HMM and SVM on the selected variables like MAP, HR and RR. Thus, the single variable MAP HMM and SVM models were used as baseline models for evaluation.

We note that for the same physiological variables, baseline models like SVM and single channel HMMs using continuous variables, report standard

performances within the range of 77-84% CVA. However, for the single variable HMMs, the CVA performance tends to be higher for MAP. This is due to MAP being the primary physiological signal used to decide the onset of a septic shock. The discrete sequence single channel HMMs generally post similar CVA performances, with minor variations across the training groups. Here, it is noted that the coupled model using continuous MAP posts a higher 5-fold CVA performance than the single variable HMM MAP. In addition, a paired-t test is carried out to compare the classification performance between the single variable MAP HMM and the MCP-CHMM. To this purpose, the classification accuracies for each of the three rounds given in Table 5.3 to 5.6 corresponding to the respective gap-intervals and observation windows were compared for the single variable MAP HMM and the MCP-CHMM. A p-value of $0.0014 \leq 0.05$ for the t-statistic was obtained. Accordingly, the null hypothesis stating that the single variable MAP HMM and MCP-CHMM perform equivalently is rejected, since the difference in performance is statistically significant.

A one-way ANOVA test was also carried out between the 4 different datasets (corresponding to the 4 different combinations of gap-interval size and the observation window sizes). Here, the null hypothesis states that there is no significant difference across the 4 different groups (or executions). Accordingly, the F-test statistic was obtained as $3.34 \leq 5.14$ (F-critical value). As a result, the null hypothesis is not rejected. The ANOVA results are shown in Table 5.10.

Symbolic pattern driven analysis of arterial blood pressure data is useful for understanding of physiological functions such as in the prediction of events like septic shock. However, there does not exist enough clarity on the non-stationarity characteristics of blood pressure datasets. For our study, it is assumed that transforms like the symbolic aggregate approximation (SAX), followed by transformation using SCPs in smaller windows, generates a locally stationary sequence. The SAX transform uses a sliding window on time-series data to generate symbolic patterns which are used to

Table 5.10: One way ANOVA Test on the 4 datasets (groups) corresponding to gap interval and observation window

Source of Variation	SS	df	MS	F	P-value	F-crit
Between Groups	2.83	2	1.41	3.34	0.10	5.14
Within Groups	1.36	3	0.45	1.07	0.428	4.75
Error	2.54	6	0.42			
Total	6.73	11				

decompose a non-stationary data.

Additionally, the continuous CHMM does not necessarily improve upon the discrete single channel HMM using MAP. Moreover, it can be seen that a coupled HMM, which considers contrast sequences from multiple physiological variables, tends to have marginally better CVAs than both continuous coupled models as well as single channel discrete HMM models. Our simulations also demonstrate that varying the gap interval size (within the range of 3 to 5) can affect the prediction performance, and that increasing the size of the observation window do not improve the performance of the training models.

Interestingly, we note that HMM models that were trained using sequences of contrast patterns generally outperformed models which used raw continuous signals only. This suggests that a patient's signal trajectory towards sepsis-related complications and significant episodes after ICU admission, can help determine a dynamically evolving patient state. Being able to use a set of discrete episodes to construct a meaningful observation sequence and then using sequential learning models to predict ICU events like septic shock, hold greater value both in terms of clinical interpretation of episodes as well as in the construction of robust prediction models.

This has been demonstrated in earlier studies for temporal data mining on medical datasets, where temporal abstractions (TA) and Allens temporal relations like "before", "equals" etc., between various TA states are used. Previous studies on temporal data mining including Sacchi et al (2007, 2015)

employed Allens temporal relations on TAs (for e.g. trend abstractions employing piecewise linear approximations). Thus, temporal relations were defined between abstracted temporal states or events. Complex temporal relations were also used to reconstruct a single multivariate abstracted sequence corresponding to a data instance. Later, frequent patterns and association rule mining algorithms were employed to extract interesting temporal rules.

On the other hand, the current study was applied on a homogeneous dataset of MAP, HR and RR time series. Generally, for the application of temporal relations, defining a temporal abstraction is necessary. In this regard, if a sequential contrast pattern is regarded as a temporal abstraction, then temporal relations can be applied. Our study differs from previous studies, in exploring the use of a gapped sequential contrast pattern (SCP) as a temporal abstraction (in comparison to trend abstractions). Accordingly, the chaining technique applies Allens temporal relations of before and overlap to reconstruct and transform the original data instance into an ordered set of SCPs. Finally, our study explores the use of a coupled HMM to integrate multivariate SCP sequences to construct a supervised prediction model. Thus, the current study employs temporal relations to construct a sequence of multivariate SCPs before employing a CHMM for better performance. Typically, the primary goal of previous temporal pattern mining studies had been the extraction of interesting temporal rules using frequent mining techniques. In contrast, our study focuses on the integration of sequential contrast patterns for use in multivariate sequential learning models (such as CHMMs) for predicting septic shock.

As described, pattern based CHMM models outperform simple measures like APACHE-III, SVM models, neighborhood-based imputation techniques described in (Ho et al. 2014). Previously, Ho et al (2014) had demonstrated the application of forward and backward selection strategies using EWS feature matrices to obtain accuracies in the range of 72-78% on a similar septic shock dataset from MIMIC-II. The authors used the ICD 785.52 code to select septic shock patients (based on ICD9 coding) from the septic patients

population, whereas the current study applied the conditions of systolic blood pressure ≤ 90 mmHg and fluid input ≥ 600 mL to select septic shock patient segments from the septic patients population. Our results at 60 minutes of gap interval using discretized patterns and CHMM, also post comparatively similar performances. The results also demonstrate that the integration of coupled HMM with discrete sequential patterns provide better performance, in comparison to using HMM models on continuous variables. Further, it can be said that sequential contrast patterns have interpretive significance, such that a sequence of patterns, when used to describe a variable sequence encodes it into a set of episodes in a sequence. This sequence of episodes clearly allows the CHMM model to perform well in comparison to the direct use of models on continuous time series data.

Results show that our models can predict septic shock events using time series of contrast patterns, which have comparative performances as earlier models. However, one must note that the application of complex septic shock models and the acceptable detection rates in actual practice have been limited to the use of traditional clinical measures like APACHE-III. Integrating sequential patterns using a CHMM, allow us to capture interactions among discrete patterns of physiological variables, which are useful for predicting labels for patient sequences. For this study, our models were trained using a set of contrast patterns favouring positive instances i.e patients having septic shock. However, there may be certain sequences where the CHMM probabilities are marginally greater or lesser than the user-defined threshold, to be labelled as a positive instance. Therefore, it is necessary to also explore models which can deal with predicting instances on the fringe regions of a probability threshold. In the context of pattern-based classification methods, different types of interestingness measures can be used to select patterns which are more representative of a patient. Currently, the growth rate (which includes the use of positive and negative support) of a pattern is used to extract a candidate pattern. Complex interestingness measures can be developed, using the support, cohesiveness (i.e. the closeness of items

in a pattern), coverage of the pattern etc. The interestingness measure of a pattern is important since setting the pattern growth rate threshold too high results in not generating enough contrast patterns, while setting the threshold too low leads to generating useless patterns. An additional step can involve the use of variable selection algorithms to determine an optimal discriminate pattern set to characterise a data instance. Additionally, ensemble pattern-based classification models can be created for representative samples from the given population for addressing learning bias. Accordingly, these cases can be difficult to detect and require further studies.

To address issues related to clustering of patient data, random effect models can be used to account for correlation within a cluster. From a pattern mining viewpoint, this means each unique patient cluster consists of a subset of patterns (i.e a specific pattern subset is most frequent in this cluster). If the cluster is represented as a tree, such that the root node (or a unique patient) consists of the maximum number of patterns in the subset, then other nodes (other patients) in the cluster having a smaller subset of these patterns can be labelled as child nodes of the root or form sub-trees. To this purpose, the maximality condition and the length of a pattern can be used as a condition for mining patterns. However, mining of longer sequential contrast patterns also requires lowering the positive frequency support threshold. This may limit the formation of large multilevel pattern trees (representing patient clusters). Thus, future clinical data mining work can adopt multiobjective pattern selection techniques such that the size of a patient cluster is reduced, while also maximizing the pattern coverage of the patient dataset. Additionally, to account for effects within a cluster, multi-level models can be explored to predict both the disease outcome of a patient (i.e the class label) as well as the cluster label (or expectation value of a cluster).

Moreover, dimensionality reduction and variable selection strategies could be explored on the discrete sequential patterns space to compare pattern classification models using methods like logistic regression. In the current study, the extracted gapped SCPs imposed a gap in the range of 3 to 5 between two

consecutive elements in an SCP. Future studies can address finding optimal gap constraints between consecutive symbolic items for sequential patterns in clinical domains. Specially, the use of optimal gaps for clinical patterns can inform studies related to the correlation of clinical events separated over a time scale.

5.6 Conclusion

In this study, we have presented a novel integrated framework, consisting of sequential contrast patterns with coupled hidden markov models (CHMM) to predict ICU events like the onset of a septic shock. The method involves the determination of contrast sequences from differentially labelled multivariate patient populations and the method then employs a generalised coupled modelling process for multiple channels of time series of contrast patterns. In turn, the CHMM model allows us to account for interactions among patterns from different channels or variables. To verify the effectiveness of pattern sequences, we compared our method with the traditional SVM and continuous single variable HMM counterparts. These methods were all tested using datasets extracted from the MIMIC-II database. Our results demonstrate that the learning models, which account for position or order information among sequential patterns, tend to perform well in comparison to models not exploiting such information. Hence, the current study describes the integration of meta-information about patterns and intermediate relationships, such as sequence ordering, to improve the performance of sequential learning models.

Thus, the current study demonstrates the importance of training ICU classifier models using informative sequential patterns, in addition to conventional clinical measures. Accordingly, the use of sequential patterns to encode discretised sequences, allows easier handling of large scale noisy data commonly encountered in modern clinical studies. Hence, the recommended septic shock prediction framework employing discrete sequential patterns,

can provide ICU care systems a novel clinical pattern discovery platform to improve patient outcomes.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The mining and extraction of interpretable and highly predictive sequential patterns is an essential task in data mining and machine learning. In the context of critical care, sequence mining focuses on extracting patterns that have strong clinical value and impact. Such patterns turn out to be extremely useful for medical decision-making and deriving important clinical insights into disease progressions.

In this dissertation, we studied about the mining of sequential contrastive patterns in the supervised clinical setting. The objective of thse studies were to find sequential patterns that were discriminative (i.e able to capture differentiating physiological behaviour among patient subpopulations) and then employ such patterns to derive univariate and multivariate contrast prediction models for predicting important clinical events in critical care. To this purpose, we have presented multiple methods for integrative mining of contrast patterns, relevant data transformations to connect with generalised pattern based classification models.

The inherent value of contrast mining in clinical knowledge discovery lies in the application of the growth rate of a pattern. The growth rate of a pattern is intuitive from a clinical applications perspective. This is because the

traditional objective in clinical trials, is oriented towards finding differences between the intervention and control population of patients. Thus, finding patterns based on differences in their supports in the intervention and control populations, allow us to find sequential patterns that can explain the difference between two populations of patient data. This is closely similar to the concept of the odds ratio, which is a popular measure to clinicians for finding association between an exposure and an outcome. Thus, given that a particular clinical event has occurred, mining of contrast sequential patterns prior to the event's occurrence provides us with the odds of a patient displaying a complication at a later point in time. To this purpose, there is a need for exploring novel methods in contrast pattern mining that can be associated with desirable treatment outcomes for patients in critical care.

Moreover, our experience shows that classification problems in clinical care settings suffer from imbalanced datasets. This means for various types of medical treatments and conditions the intervention population (i.e the positive class) tends to be skewed in terms of the control population. The challenge here turns out to be designing interestingness measures like pattern growth rates to select relevant clinical patterns, while reducing redundant patterns.

In Chapter 3, we applied a systematic methodology using a flexible sequential contrast mining algorithm on a discretised dataset, targeting the prediction of acute hypotension. In numerous studies prior to our work, simple and easily available patient indices and statistical measures had been used to tackle this problem. Our work demonstrates that sequential contrast patterns when extracted from discretised physiological variables of a patient, turn out to be highly predictive of a hypotensive state. In addition to demonstrating the classification performance, we also established the existence of gap-constrained symbolic subsequences, which could be translated into a complex sequence of clinical symptoms. Hence, these patterns are a potential source for launching further data driven investigations validated by randomized clinical trials and can enable clinicians to develop complex med-

ical hypotheses by investigating association of patterns to patient response for a specific treatment.

In Chapter 4, we recommend a number of large-scale short term critical event prediction models using binary and frequency based feature transformations of sequential patterns for predicting events including patient mortality. From a clinical data mining viewpoint, the integration of pattern mining and standard machine learning algorithms for ICU prediction problems is relatively nascent. Our work systematically investigates the integration of sequential contrast patterns with classification models by two mapping techniques in clinical settings.

Chapter 5 extends the concepts described in earlier chapters to propose a methodology for concatenating contrast patterns depending on their order of occurrence within a patient sequence. Here, we presented a novel integrated framework, consisting of sequential contrast patterns with coupled hidden markov models (CHMM) to predict the onset of a septic shock event. The method involves the determination of contrast sequences from differentially labelled multivariate patient populations and then employing a generalized coupled modelling process for multiple channels of time series of contrast patterns. To verify the effectiveness of pattern sequences, we compared our method with the traditional SVM and continuous single variable HMM counterparts. Our results indicate that the learning models, which account for position or order information among sequential patterns, tend to perform well in comparison to models not exploiting such information. thus, this study describes the integration of meta-information about patterns and intermediate relationships, such as sequence ordering, to improve the performance of sequential learning models. Moreover, the use of sequential patterns to encode discretised sequences, allows easier handling of large scale noisy data commonly encountered in modern clinical studies. Hence, the recommended septic shock prediction framework employing discrete sequential patterns, can provide ICU care systems a novel clinical pattern discovery platform to improve patient outcomes.

6.2 Future Work

In future research, we would like to explore the potential directions as described below, which can have tremendous value both from a clinical as well as theoretical perspective.

(i). **Patient phenotype discovery using emerging patterns**: In the applications described in this thesis, we have focused on using contrast patterns to classify critical events. However, there has been relatively less research carried out in mining discriminative approximate sequential patterns which can be used to query databases to identify patient cohorts. To this purpose, the application of emerging patterns can be significant in the process of clinical phenotype discovery.

(ii). **Extracting relationships between emerging patterns**: Emerging physiological patterns can be extracted and the relationships between these patterns can be learned using graphical models. Currently, the use of temporal abstractions is a way to encode relationships between clinical events. However, these abstractions tend to be predefined temporal relations. There is a need for models that are able to extract relationships between patient physiological patterns which can help to explain the interactions between the variables.

(iii). **Contrastive Temporal and Time-Interval Patterns**: The models explored in this thesis focused on extracting contrast sequential patterns. Contrast mining techniques could be further extended using temporal abstractions between a sequence of events.

(iv). **Mining medication pathways using sequential contrast patterns**: Sequential pattern mining is a useful data mining technique for identifying temporal relationships between medications. Mining temporal relationships are useful for making predictions about which medication a prescriber is likely to choose next when treating a progressive disease such as diabetes.

(v). **Contrast patterns and Clinical Text Mining**: The clinical discharge notes of a patient record the long term progression of a patient involving observed symptoms, treatments administered, allergic drug reactions and so on. In the context of clinical text mining, contrast patterns can be employed to determine changes in patient progression. To this purpose, mining of clinical concepts and semantic relations that contrast between patient populations holds significant value for knowledge discovery in clinical domains.

(vi). **Discretisation of Clinical Time Series and Mining of Emerging Patterns**: Generally, sequential pattern mining algorithms re-quire a time series dicretisation process prior to mining of patterns. Depending on how the discretisation of medical time series is carried out, the quality of extracted sequential patterns can differ and vary in predictive capability. There is significant opportunity to propose methods in this area so that optimal predictive results are obtained while increasing the explainability of extracted patterns.

Bibliography

Agrawal, R., I. T. & Swami, A. (1993), Mining association rules between sets of items in large databases, *in* 'Proceedings of the international conference on Management Of Data (SIGMOD)'.

Agrawal, R., Imieliński, T. & Swami, A. (1993), Mining association rules between sets of items in large databases, *in* 'Acm sigmod record', Vol. 22, ACM, pp. 207–216.

Agrawal, R. & Shafer, J. C. (1996), 'Parallel mining of association rules', *IEEE Transactions on knowledge and Data Engineering* **8**(6), 962–969.

Agrawal, R. & Srikant, R. (1995), Mining sequential patterns, *in* 'Data Engineering, 1995. Proceedings of the Eleventh International Conference on', IEEE, pp. 3–14.

Alhammady, H. (2007), Mining streaming emerging patterns from streaming data, *in* 'Computer Systems and Applications, 2007. AICCSA'07. IEEE/ACS International Conference on', IEEE, pp. 432–436.

Anderson, R. J. (2011), 'Plumbing the depths of blood pressure: hypotensive hemorrhage and acute kidney injury', *Critical care medicine* **39**(9), 2196–2197.

Andruszkiewicz, P. (2011), Lazy approach to privacy preserving classification with emerging patterns, *in* 'Emerging intelligent technologies in industry', Springer, pp. 253–268.

Angus, D. C. & Van der Poll, T. (2013), 'Severe sepsis and septic shock', *New England Journal of Medicine* **369**(9), 840–851.

Audhkhasi, K., Osoba, O. & Kosko, B. (2013), Noisy hidden markov models for speech recognition, *in* 'Neural Networks (IJCNN), The 2013 International Joint Conference on', IEEE, pp. 1–6.

Awad, H. H., Anderson, F. A., Gore, J. M., Goodman, S. G. & Goldberg, R. J. (2012), 'Cardiogenic shock complicating acute coronary syndromes: insights from the global registry of acute coronary events', *American heart journal* **163**(6), 963–971.

Ayres, J., Flannick, J., Gehrke, J. & Yiu, T. (2002), Sequential pattern mining using a bitmap representation, *in* 'Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 429–435.

Bailey, J., Manoukian, T. & Ramamohanarao, K. (2002), Fast algorithms for mining emerging patterns, *in* 'European Conference on Principles of Data Mining and Knowledge Discovery', Springer, pp. 39–50.

Bailey, J., Manoukian, T. & Ramamohanarao, K. (2003), A fast algorithm for computing hypergraph transversals and its application in mining emerging patterns, *in* 'Data Mining, 2003. ICDM 2003. Third IEEE International Conference on', IEEE, pp. 485–488.

Batal, I., Cooper, G. F., Fradkin, D., Harrison Jr, J., Moerchen, F. & Hauskrecht, M. (2016), 'An efficient pattern mining approach for event detection in multivariate temporal data', *Knowledge and information systems* **46**(1), 115–150.

Batal, I., Fradkin, D., Harrison, J., Moerchen, F. & Hauskrecht, M. (2012), Mining recent temporal patterns for event detection in multivariate time series data, *in* 'Proceedings of the 18th ACM SIGKDD international

conference on Knowledge discovery and data mining', ACM, pp. 280–288.

Batal, I., Sacchi, L., Bellazzi, R. & Hauskrecht, M. (2009), 'Multivariate time series classification with temporal abstractions'.

Batal, I., Valizadegan, H., Cooper, G. F. & Hauskrecht, M. (2011), A pattern mining approach for classifying multivariate temporal data, *in* 'Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on', IEEE, pp. 358–365.

Batal, I., Valizadegan, H., Cooper, G. F. & Hauskrecht, M. (2013), 'A temporal pattern mining approach for classifying electronic health record data', *ACM Transactions on Intelligent Systems and Technology (TIST)* **4**(4), 63.

Bayardo Jr, R. J. (1998), 'Efficiently mining long patterns from databases', *ACM Sigmod Record* **27**(2), 85–93.

Bellazzi, R., Ferrazzi, F. & Sacchi, L. (2011), 'Predictive data mining in clinical medicine: a focus on selected methods and applications', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1**(5), 416–430.

Bellazzi, R., Larizza, C., Magni, P. & Bellazzi, R. (2005), 'Temporal data mining for the quality assessment of hemodialysis services', *Artificial intelligence in medicine* **34**(1), 25–39.

Bender, R. (2009), 'Introduction to the use of regression models in epidemiology', *Cancer Epidemiology* pp. 179–195.

Bera, D. & Nayak, M. M. (2012), Mortality risk assessment for icu patients using logistic regression, *in* 'Computing in Cardiology (CinC), 2012', IEEE, pp. 493–496.

Bohensky, M. A., Jolley, D., Pilcher, D. V., Sundararajan, V., Evans, S. & Brand, C. A. (2012), 'Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post–hospital discharge', *Journal of critical care* **27**(4), 422–e11.

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., Schein, R. M. & Sibbald, W. J. (1992), 'Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis', *Chest* **101**(6), 1644–1655.

Boulesteix, A.-L., Tutz, G. & Strimmer, K. (2003), 'A cart-based approach to discover emerging patterns in microarray data', *Bioinformatics* **19**(18), 2465–2472.

Brinkman, S., Abu-Hanna, A., van der Veen, A., de Jonge, E. & de Keizer, N. F. (2012), 'A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units', *Critical care medicine* **40**(2), 373–378.

Cao, L., Ou, Y. & Philip, S. Y. (2012), 'Coupled behavior analysis with applications', *IEEE Transactions on Knowledge and Data Engineering* **24**(8), 1378–1392.

Capp, R. & et al, H. (2015), 'Predictors of patients who present to the emergency department with sepsis and progress to septic shock between 4 and 48 hours of emergency department arrival', *Critical care medicine* **43**(5), 983–988.

Carrara, M., Baselli, G. & Ferrario, M. (2015), Mortality prediction in septic shock patients: Towards new personalized models in critical care, *in* 'Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE', IEEE, pp. 2792–2795.

Casanova, I. J., Campos, M., Juarez, J. M., Fernandez-Fernandez-Arroyo, A. & Lorente, J. A. (2015), Using multivariate sequential patterns to improve survival prediction in intensive care burn unit, *in* 'Conference on Artificial Intelligence in Medicine in Europe', Springer, pp. 277–286.

Ceci, M., Appice, A., Caruso, C. & Malerba, D. (2008), Discovering emerging patterns for anomaly detection in network connection data, *in* 'International Symposium on Methodologies for Intelligent Systems', Springer, pp. 179–188.

Celi, L. A., Christian, L. H., Alterovitz, G. & Szolovits, P. (2008), 'An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study', *Critical Care* **12**(6), R151.

Celi, L. A. G., Tang, R. J., Villarroel, M. C., Davidzon, G. A., Lester, W. T. & Chueh, H. C. (2011), 'A clinical database-driven approach to decision support: Predicting mortality among patients with acute kidney injury', *Journal of healthcare engineering* **2**(1), 97–110.

Celi, L. A., Galvin, S., Davidzon, G., Lee, J., Scott, D. & Mark, R. (2012), 'A database-driven decision support system: customized mortality prediction', *Journal of personalized medicine* **2**(4), 138–148.

Chan, S., Kao, B., Yip, C. L. & Tang, M. (2003), Mining emerging substrings, *in* 'Database Systems for Advanced Applications, 2003.(DASFAA 2003). Proceedings. Eighth International Conference on', IEEE, pp. 119–126.

Chen, X., Xu, D., Zhang, G. & Mukkamala, R. (2009), Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform, *in* 'Computers in Cardiology, 2009', IEEE, pp. 545–548.

Chen, Y.-T. (2007), 'Moment-based copula tests for financial returns', *Journal of Business & Economic Statistics* **25**(4), 377–397.

Cheng, H., Yan, X., Han, J. & Hsu, C.-W. (2007), Discriminative frequent pattern analysis for effective classification, *in* 'Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on', IEEE, pp. 716–725.

Cheng, H., Yan, X., Han, J. & Philip, S. Y. (2008), Direct discriminative pattern mining for effective classification, *in* 'Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on', IEEE, pp. 169–178.

Cheung, D. W., Han, J., Ng, V. T., Fu, A. W. & Fu, Y. (1996), A fast distributed algorithm for mining association rules, *in* 'Parallel and Distributed Information Systems, 1996., Fourth International Conference on', IEEE, pp. 31–42.

Chiu, D.-Y., Wu, Y.-H. & Chen, A. L. (2004), An efficient algorithm for mining frequent sequences by a new strategy without support counting, *in* 'Data Engineering, 2004. Proceedings. 20th International Conference on', IEEE, pp. 375–386.

Cristianini, N. & Shawe-Taylor, J. (2000), *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press.

Donald, R., Howells, T., Piper, I., Chambers, I., Citerio, G., Enblad, P., Gregson, B., Kiening, K., Mattern, J., Nilsson, P. et al. (2012), *Early warning of EUSIG-defined hypotensive events using a Bayesian Artificial Neural Network*, Springer.

Dong, G. & Li, J. (1999), Efficient mining of emerging patterns: Discovering trends and differences, *in* 'Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 43–52.

Eshelman, L. J., Lee, K., Frassica, J. J., Zong, W., Nielsen, L. & Saeed, M. (2008), Development and evaluation of predictive alerts for hemodynamic instability in icu patients., *in* 'AMIA'.

Exarchos, T. P., Tsipouras, M. G., Papaloukas, C. & Fotiadis, D. I. (2009), 'An optimized sequential pattern matching methodology for sequence classification', *Knowledge and information systems* **19**(2), 249–264.

Fan, H., Fan, M., Ramamohanarao, K. & Liu, M. (2006), Further improving emerging pattern based classifiers via bagging, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 91–96.

Fan, H. & Kotagiri, R. (2002), An efficient single-scan algorithm for mining essential jumping emerging patterns for classification, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 456–462.

Fan, H. & Ramamohanarao, K. (2003), A bayesian approach to use emerging patterns for classification, *in* 'Proceedings of the 14th Australasian database conference-Volume 17', Australian Computer Society, Inc., pp. 39–48.

Fan, H. & Ramamohanarao, K. (2006), 'Fast discovery and the generalization of strong jumping emerging patterns for building compact and accurate classifiers', *IEEE Transactions on Knowledge and Data Engineering* **18**(6), 721–737.

Fayyad, U. & Irani, K. (1993), 'Multi-interval discretization of continuous-valued attributes for classification learning'.

Fialho, A., Celi, L., Cismondi, F., Vieira, S., Reti, S., Sousa, J., Finkelstein, S. et al. (2013), 'Disease-based modeling to predict fluid response in intensive care units', *Methods of information in medicine* **52**(6), 494–502.

Fischer, J., Mäkinen, V. & Välimäki, N. (2008), Space efficient string mining under frequency constraints, *in* 'Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on', IEEE, pp. 193–202.

Fournier, P. & Roy, J. (2009), Acute hypotension episode prediction using information divergence for feature selection, and non-parametric methods for classification, *in* 'Computers in Cardiology, 2009', IEEE, pp. 625–628.

Fradkin, D. & Mörchen, F. (2015), 'Mining sequential patterns for classification', *Knowledge and Information Systems* **45**(3), 731–749.

Friedman, N., Geiger, D. & Goldszmidt, M. (1997), 'Bayesian network classifiers', *Machine learning* **29**(2-3), 131–163.

Gamberger, D. & Lavrac, N. (2002), 'Expert-guided subgroup discovery: Methodology and application', *Journal of Artificial Intelligence Research* **17**, 501–527.

García-Borroto, M., Martínez-Trinidad, J. F. & Carrasco-Ochoa, J. A. (2010), A new emerging pattern mining algorithm and its application in supervised classification, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 150–157.

García-Borroto, M., Martínez-Trinidad, J. F. & Carrasco-Ochoa, J. A. (2011), 'Fuzzy emerging patterns for classifying hard domains', *Knowledge and Information Systems* **28**(2), 473–489.

García-Borroto, M., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A., Medina-Pérez, M. A. & Ruiz-Shulcloper, J. (2010), 'Lcmine: An efficient algorithm for mining discriminative regularities and its application in supervised classification', *Pattern Recognition* **43**(9), 3025–3034.

Gavrishchaka, V. V. & Bykov, V. (2007), Market-neutral portfolio of trading strategies as universal indicator of market micro-regimes: from rare-event forecasting to single-example learning of emerging patterns, *in*

'Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on', IEEE, pp. 215–215.

Ghosh, S., Feng, M., Nguyen, H. & Li, J. (2014), Risk prediction for acute hypotensive patients by using gap constrained sequential contrast patterns, *in* 'AMIA Annual Symposium Proceedings', Vol. 2014, American Medical Informatics Association, p. 1748.

Ghosh, S., Feng, M., Nguyen, H. & Li, J. (2016), 'Hypotension risk prediction via sequential contrast patterns of icu blood pressure', *IEEE journal of biomedical and health informatics* **20**(5), 1416–1426.

Gotz, D., Wang, F. & Perer, A. (2014), 'A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data', *Journal of biomedical informatics* **48**, 148–159.

Gu, T., Wu, Z., Tao, X., Pung, H. K. & Lu, J. (2009), epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition, *in* 'Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on', IEEE, pp. 1–9.

Güiza, F., Van Eyck, J. & Meyfroidt, G. (2013), 'Predictive data mining on monitoring data from the intensive care unit', *Journal of clinical monitoring and computing* **27**(4), 449–453.

Gwadry-Sridhar, F., Lewden, B., Mequanint, S. & Bauer, M. (2009), Comparison of analytic approaches for determining variables-a case study in predicting the likelihood of sepsis., *in* 'HEALTHINF', pp. 90–96.

Ha, S. H. (2011), 'Medical domain knowledge and associative classification rules in diagnosis', *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)* **2**(1), 60–73.

Hämäläinen, W. (2010), 'Statapriori: an efficient algorithm for searching statistically significant association rules', *Knowledge and information systems* **23**(3), 373–399.

Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U. & Hsu, M.-C. (2000), Freespan: frequent pattern-projected sequential pattern mining, *in* 'Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 355–359.

Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. & Hsu, M. (2001), Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth, *in* 'proceedings of the 17th international conference on data engineering', pp. 215–224.

Han, J., Pei, J. & Yin, Y. (2000), Mining frequent patterns without candidate generation, *in* 'ACM Sigmod Record', Vol. 29, ACM, pp. 1–12.

Han, J., Pei, J., Yin, Y. & Mao, R. (2004), 'Mining frequent patterns without candidate generation: A frequent-pattern tree approach', *Data mining and knowledge discovery* **8**(1), 53–87.

Hauskrecht, M., Valko, M., Batal, I., Clermont, G., Visweswaran, S. & Cooper, G. (2010), Conditional outlier detection for clinical alerting, *in* 'AMIA annual symposium proceedings', Vol. 2010, pp. 286–90.

Henriques, J. & Rocha, T. (2009), Prediction of acute hypotensive episodes using neural network multi-models, *in* 'Computers in Cardiology, 2009', IEEE, pp. 549–552.

Henry, K. E., Hager, D. N., Pronovost, P. J. & Saria, S. (2015), 'A targeted real-time early warning score (trewscore) for septic shock', *Science Translational Medicine* **7**(299), 299ra122–299ra122.

Ho, J. C., Lee, C. H. & Ghosh, J. (2012), Imputation-enhanced prediction of septic shock in icu patients, *in* 'Proceedings of the ACM SIGKDD Workshop on Health Informatics', pp. 21–27.

Ho, J. C., Lee, C. H. & Ghosh, J. (2014), 'Septic shock prediction for patients with missing data', *ACM Transactions on Management Information Systems (TMIS)* **5**(1), 1.

Höppner, F. (2003), Knowledge discovery from sequential data, PhD thesis, PhD thesis, Technical University Braunschweig, Germany.

Höppner, F. & Peter, S. (2014), 'Temporal interval pattern languages to characterize time flow', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **4**(3), 196–212.

Hug, C. W. (2009), Detecting hazardous intensive care patient episodes using real-time mortality models, PhD thesis, Citeseer.

Hug, C. W., Clifford, G. D. & Reisner, A. T. (2011), 'Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension', *Critical care medicine* **39**(5), 1006.

Jensen, P. B., Jensen, L. J. & Brunak, S. (2012), 'Mining electronic health records: towards better research applications and clinical care', *Nature Reviews Genetics* **13**(6), 395–405.

Ji, X., Bailey, J. & Dong, G. (2007), 'Mining minimal distinguishing subsequence patterns with gap constraints', *Knowledge and Information Systems* **11**(3), 259–286.

Jousset, F., Lemay, M. & Vesin, J. (2009), Computers in cardiology/physionet challenge 2009: Predicting acute hypotensive episodes, *in* 'Computers in Cardiology, 2009', IEEE, pp. 637–640.

Kim, S., Kim, W. & Park, R. W. (2011), 'A comparison of intensive care unit mortality prediction models through the use of data mining techniques', *Healthcare informatics research* **17**(4), 232–243.

Klösgen, W. & May, M. (2002), 'Spatial subgroup mining integrated in an object-relational spatial database', *Principles of Data Mining and Knowledge Discovery* pp. 323–344.

Knobbe, A., Crémilleux, B., Fürnkranz, J. & Scholz, M. (2008), 'From local patterns to global models: the lego approach to data mining', *LeGo* **8**, 1–16.

Kobyliński, Ł. & Walczak, K. (2008), Jumping emerging patterns with occurrence count in image classification, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 904–909.

Kristjansson, T. T., Frey, B. J. & Huang, T. S. (2000), Event-coupled hidden markov models, *in* 'Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on', Vol. 1, IEEE, pp. 385–388.

Langley, P., King, S., Zheng, D., Bowers, E., Wang, K., Allen, J. & Murray, A. (2009), Predicting acute hypotensive episodes from mean arterial pressure, *in* 'Computers in Cardiology, 2009', IEEE, pp. 553–556.

Latronico, N. (2015), 'Prediction is very difficult, especially about the future', *Critical care medicine* **43**(2), 505–506.

Lee, J. & Mark, R. G. (2010), 'An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care', *Biomedical engineering online* **9**(1), 62.

Li, H., Li, X., Jia, X., Ramanathan, M. & Zhang, A. (2015), Bone disease prediction and phenotype discovery using feature representation over electronic health records, *in* 'Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics', ACM, pp. 212–221.

Li, J., Dong, G., Ramamohanarao, K. & Wong, L. (2004), 'Deeps: A new instance-based lazy discovery and classification system', *Machine learning* **54**(2), 99–124.

Li, J., Ramamohanarao, K. & Dong, G. (2000), The space of jumping emerging patterns and its incremental maintenance algorithms., *in* 'ICML', pp. 551–558.

Li, J., Shi, J. & Satz, D. (2008), 'Modeling and analysis of disease and risk factors through learning bayesian networks from observational data'.

Li, L., McCann, J., Pollard, N. S. & Faloutsos, C. (2009), Dynammo: Mining and summarization of coevolving sequences with missing values, *in* 'Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 507–516.

Li, W., Han, J. & Pei, J. (2001), Cmar: Accurate and efficient classification based on multiple class-association rules, *in* 'Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on', IEEE, pp. 369–376.

Li-wei, H. L., Nemati, S., Adams, R. P., Moody, G., Malhotra, A. & Mark, R. G. (2013), Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series, *in* 'Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE', IEEE, pp. 7072–7075.

Li, Y., Swift, S. & Tucker, A. (2013), 'Modelling and analysing the dynamics of disease progression from cross-sectional studies', *Journal of biomedical informatics* **46**(2), 266–274.

Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003), A symbolic representation of time series, with implications for streaming algorithms, *in* 'Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery', ACM, pp. 2–11.

Lin, J., Keogh, E., Wei, L. & Lonardi, S. (2007), 'Experiencing sax: a novel symbolic representation of time series', *Data Mining and knowledge discovery* **15**(2), 107–144.

Loekito, E. & Bailey, J. (2009), Using highly expressive contrast patterns for classification-is it worthwhile?, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 483–490.

Low, P. A. (2008), 'Prevalence of orthostatic hypotension', *Clinical Autonomic Research* **18**(1), 8–13.

Lowe, H. J., Ferris, T. A., Hernandez, P. M., Weber, S. C. et al. (2009), Stride-an integrated standards-based translational research informatics platform., *in* 'AMIA'.

Lukaszewski, R. A. & et al, Y. (2008), 'Presymptomatic prediction of sepsis in intensive care unit patients', *Clinical and Vaccine Immunology* **15**(7), 1089–1094.

Mabroukeh, N. R. & Ezeife, C. I. (2010), 'A taxonomy of sequential pattern mining algorithms', *ACM Computing Surveys (CSUR)* **43**(1), 3.

Malhotra, K., Hobson, T. C., Valkova, S., Pullum, L. L. & Ramanathan, A. (2015), Sequential pattern mining of electronic healthcare reimbursement claims: Experiences and challenges in uncovering how patients are treated by physicians, *in* 'Big Data (Big Data), 2015 IEEE International Conference on', IEEE, pp. 2670–2679.

Mannila, H., Toivonen, H. & Verkamo, A. I. (1997), 'Discovery of frequent episodes in event sequences', *Data mining and knowledge discovery* **1**(3), 259–289.

Masoudi, S., Montazeri, N., Shamsollahi, M., Ge, D., Beuchee, A., Pladys, P. & Hernández, A. I. (2013), Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden markov model, *in* 'Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on', IEEE, pp. 000243–000248.

Mayaud, L., Lai, P. S., Clifford, G. D., Tarassenko, L., Celi, L. A. G. & Annane, D. (2013), 'Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension', *Critical care medicine* **41**(4), 954.

McGregor, C., Catley, C., Padbury, J. & James, A. (2013), 'Late onset neonatal sepsis detection in newborn infants via multiple physiological streams', *Journal of critical care* **28**(1), e11.

Meyfroidt, G., Güiza, F., Cottem, D., De Becker, W., Van Loon, K., Aerts, J.-M., Berckmans, D., Ramon, J., Bruynooghe, M. & Van den Berghe, G. (2011), 'Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a gaussian processes model', *BMC medical informatics and decision making* **11**(1), 64.

Minato, S.-i. (1993), Zero-suppressed bdds for set manipulation in combinatorial problems, *in* 'Proceedings of the 30th international Design Automation Conference', ACM, pp. 272–277.

Mneimneh, M. & Povinelli, R. (2009), A rule-based approach for the prediction of acute hypotensive episodes, *in* 'Computers in Cardiology, 2009', IEEE, pp. 557–560.

Moerchen, F. (2006), Algorithms for time series knowledge mining, *in* 'Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 668–673.

Moody, G. B. & Lehman, L. (2009), Predicting acute hypotensive episodes: The 10th annual physionet/computers in cardiology challenge, *in* 'Computers in Cardiology, 2009', IEEE, pp. 541–544.

Mooney, C. H. & Roddick, J. F. (2013), 'Sequential pattern mining–approaches and algorithms', *ACM Computing Surveys (CSUR)* **45**(2), 19.

Mörchen, F. & Fradkin, D. (2010), Robust mining of time intervals with semi-interval partial order patterns, *in* 'Proceedings of the 2010 SIAM International Conference on Data Mining', SIAM, pp. 315–326.

Moskovitch, R. & Shahar, Y. (2009), Medical temporal-knowledge discovery via temporal abstraction, *in* 'AMIA annual symposium proceedings', Vol. 2009, American Medical Informatics Association, p. 452.

Moskovitch, R. & Shahar, Y. (2015), 'Classification-driven temporal discretization of multivariate time series', *Data Mining and Knowledge Discovery* **29**(4), 871–913.

Moskovitch, R., Walsh, C., Hripcsak, G. & Tatonetti, N. (2014), Prediction of biomedical events via time intervals mining, *in* 'NYC, USA: ACM KDD Workshop on Connected Health in Big Data Era'.

Muyeba, M. K., Khan, M. S., Warnars, S. & Keane, J. (2011), A framework to mine high-level emerging patterns by attribute-oriented induction, *in* 'International Conference on Intelligent Data Engineering and Automated Learning', Springer, pp. 170–177.

Nemati, S., Li-wei, H. L. & Adams, R. P. (2013), Learning outcome-discriminative dynamics in multivariate physiological cohort time series, *in* 'Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE', IEEE, pp. 7104–7107.

Nizami, S., Green, J. R. & McGregor, C. (2013), 'Implementation of artifact detection in critical care: a methodological review', *IEEE reviews in biomedical engineering* **6**, 127–142.

Ordonez, C. & Zhao, K. (2011), 'Evaluating association rules and decision trees to predict multiple target attributes', *Intelligent Data Analysis* **15**(2), 173–192.

Papapetrou, P., Kollios, G., Sclaroff, S. & Gunopulos, D. (2005), Discovering frequent arrangements of temporal intervals, *in* 'Data Mining, Fifth IEEE International Conference on', IEEE, pp. 8–pp.

Park, J. S., Chen, M.-S. & Yu, P. S. (1995), *An effective hash-based algorithm for mining association rules*, Vol. 24, ACM.

Pasquier, N., Pasquier, C., Brisson, L. & Collard, M. (2008), 'Mining gene expression data using domain knowledge', *International Journal of Software and Informatics (IJSI)* **2**(2), 215–231.

Patnaik, D., Butler, P., Ramakrishnan, N., Parida, L., Keller, B. J. & Hanauer, D. A. (2011), Experiences with mining temporal event sequences from electronic medical records: initial successes and some challenges, *in* 'Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 360–368.

Peelen, L., de Keizer, N. F., de Jonge, E., Bosman, R.-J., Abu-Hanna, A. & Peek, N. (2010), 'Using hierarchical dynamic bayesian networks to investigate dynamics of organ failure in patients in the intensive care unit', *Journal of biomedical informatics* **43**(2), 273–286.

Pei, J., Han, J. & Wang, W. (2007), 'Constraint-based sequential pattern mining: the pattern-growth methods', *Journal of Intelligent Information Systems* **28**(2), 133–160.

Pinsky, M. R. (2007), 'Hemodynamic evaluation and monitoring in the icu', *CHEST Journal* **132**(6), 2020–2029.

Rabiner, L. R. (1989), 'A tutorial on hidden markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.

Ramamohanarao, K. & Fan, H. (2007), 'Patterns based classifiers', *World Wide Web* **10**(1), 71–83.

Rezek, I. & Roberts, S. J. (2000), Estimation of coupled hidden markov models with application to biosignal interaction modelling, *in* 'Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop', Vol. 2, IEEE, pp. 804–813.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., Peterson, E. & Tomlanovich, M. (2001), 'Early goal-directed therapy in

the treatment of severe sepsis and septic shock', *New England Journal of Medicine* **345**(19), 1368–1377.

Rocha, T., Paredes, S., De Carvalho, P. & Henriques, J. (2011), 'Prediction of acute hypotensive episodes by means of neural network multi-models', *Computers in biology and medicine* **41**(10), 881–890.

Rosenberg, A. L. (2002), 'Recent innovations in intensive care unit risk-prediction models', *Current opinion in critical care* **8**(4), 321–330.

Sacchi, L., Bellazzi, R., Larizza, C., Porreca, R. & Magni, P. (2005), Learning rules with complex temporal patterns in biomedical domains, *in* 'Conference on Artificial Intelligence in Medicine in Europe', Springer, pp. 23–32.

Sacchi, L., Dagliati, A. & Bellazzi, R. (2015), 'Analyzing complex patients temporal histories: new frontiers in temporal data mining', *Data Mining in Clinical Medicine* pp. 89–105.

Saeed, M. & Mark, R. G. (2006), A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations., *in* 'AMIA'.

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L.-W., Moody, G., Heldt, T., Kyaw, T. H., Moody, B. & Mark, R. G. (2011), 'Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database', *Critical care medicine* **39**(5), 952.

Savasere, A., Omiecinski, E. R. & Navathe, S. B. (1995), An efficient algorithm for mining association rules in large databases, Technical report, Georgia Institute of Technology.

Scalzo, F. & Hu, X. (2013), 'Semi-supervised detection of intracranial pressure alarms using waveform dynamics', *Physiological measurement* **34**(4), 465.

Schafer, J. L. & Graham, J. W. (2002), 'Missing data: our view of the state of the art.', *Psychological methods* **7**(2), 147.

Schmid, F., Goepfert, M. S. & Reuter, D. A. (2013), 'Patient monitoring alarms in the icu and in the operating room', *Critical care* **17**(2), 216.

Shavdia, D. (2007), Septic shock: Providing early warnings through multivariate logistic regression models, PhD thesis, Massachusetts Institute of Technology.

Shen, W., Wang, J. & Han, J. (2014), Sequential pattern mining, *in* 'Frequent Pattern Mining', Springer, pp. 261–282.

Shibao, C., Lipsitz, L. A. & Biaggioni, I. (2013), 'Ash position paper: evaluation and treatment of orthostatic hypotension', *The Journal of Clinical Hypertension* **15**(3), 147–153.

Silva, I., Moody, G., Scott, D. J., Celi, L. A. & Mark, R. G. (2012), Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012, *in* 'Computing in Cardiology (CinC), 2012', IEEE, pp. 245–248.

Singh, A., Tamminedi, T., Yosiphon, G., Ganguli, A. & Yadegar, J. (2010), Hidden markov models for modeling blood pressure data to predict acute hypotension, *in* 'Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on', IEEE, pp. 550–553.

Skapura, N. & Dong, G. (2015), Distribution skew-based binning: Towards mining highly discriminative patterns from eeg/emg time series, *in* 'Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on', IEEE, pp. 1–6.

Srikant, R. & Agrawal, R. (1996), Mining sequential patterns: Generalizations and performance improvements, *in* 'International Conference on Extending Database Technology', Springer, pp. 1–17.

Stow, P. J., Hart, G. K., Higlett, T., George, C., Herkes, R., McWilliam, D., Bellomo, R., Committee, A. D. M. et al. (2006), 'Development and implementation of a high-quality clinical database: the australian and new zealand intensive care society adult patient database', *Journal of critical care* **21**(2), 133–141.

Syed, Z., Stultz, C., Kellis, M., Indyk, P. & Guttag, J. (2010), 'Motif discovery in physiological datasets: a methodology for inferring predictive elements', *ACM Transactions on Knowledge Discovery from Data (TKDD)* **4**(1), 2.

Takala, J. (2010), 'Should we target blood pressure in sepsis?', *Critical care medicine* **38**(10), S613–S619.

Tang, C. H., Middleton, P. M., Savkin, A. V., Chan, G. S., Bishop, S. & Lovell, N. H. (2010), 'Non-invasive classification of severe sepsis and systemic inflammatory response syndrome using a nonlinear support vector machine: a preliminary study', *Physiological measurement* **31**(6), 775.

Terlecki, P. & Walczak, K. (2008*a*), Efficient discovery of top-k minimal jumping emerging patterns, *in* 'International Conference on Rough Sets and Current Trends in Computing', Springer, pp. 438–447.

Terlecki, P. & Walczak, K. (2008*b*), Local projection in jumping emerging patterns discovery in transaction databases, *in* 'Pacific-Asia Conference on Knowledge Discovery and Data Mining', Springer, pp. 723–730.

Toivonen, H. et al. (1996), Sampling large databases for association rules, *in* 'VLDB', Vol. 96, pp. 134–145.

Toma, T., Bosman, R.-J., Siebes, A., Peek, N. & Abu-Hanna, A. (2010), 'Learning predictive models that use pattern discoverya bootstrap evaluative approach applied in organ functioning sequences', *Journal of biomedical informatics* **43**(4), 578–586.

Torio, C. M. & Andrews, R. M. (2006), 'National inpatient hospital costs: the most expensive conditions by payer, 2011: statistical brief# 160'.

Tseng, V. S. & Lee, C.-H. (2005), Cbs: A new classification method by using sequential patterns, *in* 'Proceedings of the 2005 SIAM International Conference on Data Mining', SIAM, pp. 596–600.

Van Looy, S., Verplancke, T., Benoit, D., Hoste, E., Van Maele, G., De Turck, F. & Decruyenaere, J. (2007), 'A novel approach for prediction of tacrolimus blood concentration in liver transplantation patients in the intensive care unit through support vector regression', *Critical Care* **11**(4), R83.

Vieira, S. M., Mendonça, L. F., Farinha, G. J. & Sousa, J. M. (2013), 'Modified binary pso for feature selection using svm applied to mortality prediction of septic patients', *Applied Soft Computing* **13**(8), 3494–3504.

Vincent, J.-L. & Moreno, R. (2010), 'Clinical review: scoring systems in the critically ill', *Critical care* **14**(2), 207.

Wang, F., Lee, N., Hu, J., Sun, J., Ebadollahi, S. & Laine, A. F. (2013), 'A framework for mining signatures from event sequences and its applications in healthcare data', *IEEE transactions on pattern analysis and machine intelligence* **35**(2), 272–285.

Wang, J. & Han, J. (2004), Bide: Efficient mining of frequent closed sequences, *in* 'Data Engineering, 2004. Proceedings. 20th International Conference on', IEEE, pp. 79–90.

Wang, Z., Fan, H. & Ramamohanarao, K. (2004), Exploiting maximal emerging patterns for classification, *in* 'Australasian Joint Conference on Artificial Intelligence', Springer, pp. 1062–1068.

Winarko, E. & Roddick, J. F. (2007), 'Armada–an algorithm for discovering richer relative temporal association rules from interval-based data', *Data & Knowledge Engineering* **63**(1), 76–90.

Wong, D., Clifton, D. A. & Tarassenko, L. (2012), Probabilistic detection of vital sign abnormality with gaussian process regression, *in* 'Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on', IEEE, pp. 187–192.

Xia, H., Daley, B. J., Petrie, A. & Zhao, X. (2012), A neural network model for mortality prediction in icu, *in* 'Computing in Cardiology (CinC), 2012', IEEE, pp. 261–264.

Xing, Z., Pei, J. & Keogh, E. (2010), 'A brief survey on sequence classification', *ACM Sigkdd Explorations Newsletter* **12**(1), 40–48.

Xu, W., Guan, C., Siong, C. E., Ranganatha, S., Thulasidas, M. & Wu, J. (2004), High accuracy classification of eeg signal, *in* 'Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on', Vol. 2, IEEE, pp. 391–394.

Yan, X., Han, J. & Afshar, R. (2003), Clospan: Mining: Closed sequential patterns in large datasets, *in* 'Proceedings of the 2003 SIAM International Conference on Data Mining', SIAM, pp. 166–177.

Yang, J., McAuley, J., Leskovec, J., LePendu, P. & Shah, N. (2014), Finding progression stages in time-evolving event sequences, *in* 'Proceedings of the 23rd international conference on World wide web', ACM, pp. 783–794.

Yang, Z., Kitsuregawa, M. & Wang, Y. (2006), Paid: Mining sequential patterns by passed item deduction in large databases, *in* 'Database Engineering and Applications Symposium, 2006. IDEAS'06. 10th International', IEEE, pp. 113–120.

Yang, Z., Wang, Y. & Kitsuregawa, M. (2007), Lapin: effective sequential pattern mining algorithms by last position induction for dense databases, *in* 'International Conference on Database systems for advanced applications', Springer, pp. 1020–1023.

Yeh, R. W., Sidney, S., Chandra, M., Sorel, M., Selby, J. V. & Go, A. S. (2010), 'Population trends in the incidence and outcomes of acute myocardial infarction', *New England Journal of Medicine* **362**(23), 2155–2165.

Zaki, M. J. (2001), 'Spade: An efficient algorithm for mining frequent sequences', *Machine learning* **42**(1-2), 31–60.

Zaki, M. J. & Hsiao, C.-J. (2005), 'Efficient algorithms for mining closed itemsets and their lattice structure', *IEEE transactions on knowledge and data engineering* **17**(4), 462–478.

Zhang, X., Dong, G. et al. (2000), Information-based classification by aggregating emerging patterns, *in* 'International Conference on Intelligent Data Engineering and Automated Learning', Springer, pp. 48–53.

Zhang, Y., Silvers, C. T. & Randolph, A. G. (2007), Real-time evaluation of patient monitoring algorithms for critical care at the bedside, *in* 'Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE', IEEE, pp. 2783–2786.

Zhang, Y. & Szolovits, P. (2008), 'Patient-specific learning in real time for adaptive monitoring in critical care', *Journal of biomedical informatics* **41**(3), 452–460.

Zhong, S. & Ghosh, J. (2002), Hmms and coupled hmms for multi-channel eeg classification, *in* 'Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on', Vol. 2, IEEE, pp. 1154–1159.

Zhou, H., Chen, J., Dong, G., Wang, H. & Yuan, H. (2016), 'Bearing fault recognition method based on neighbourhood component analysis and coupled hidden markov model', *Mechanical Systems and Signal Processing* **66**, 568–581.