2018

Machine Learning Methods for Septic Shock Prediction

Aiman A. Darwiche
Nova Southeastern University, ad1443@mynsu.nova.edu

Machine Learning Methods for Septic Shock Prediction

by

Aiman A. Darwiche

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Computer Science

College of Engineering and Computing
Nova Southeastern University

2018

We hereby certify that this dissertation, submitted by Aiman Darwiche, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Sumitra Mukherjee, Ph.D.
Chairperson of Dissertation Committee

July 10, 2018
Date

Michael J. Laszlo, Ph.D.
Dissertation Committee Member

July 10, 2018
Date

Francisco J. Mitropoulos, Ph.D.
Dissertation Committee Member

July 10 2018
Date

Approved:

Meline Kevorkian, Ed.D.
Interim Dean, College of Engineering and Computing

July 10, 2018
Date

College of Engineering and Computing
Nova Southeastern University

2018

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Machine Learning Methods for Septic Shock Prediction

by
Aiman A. Darwiche
July 2018

Sepsis is an organ dysfunction life-threatening disease that is caused by a dysregulated body response to infection. Sepsis is difficult to detect at an early stage, and when not detected early, is difficult to treat and results in high mortality rates. Developing improved methods for identifying patients in high risk of suffering septic shock has been the focus of much research in recent years. Building on this body of literature, this dissertation develops an improved method for septic shock prediction. Using the data from the MMIC-III database, an ensemble classifier is trained to identify high-risk patients. A robust prediction model is built by obtaining a risk score from fitting the Cox Hazard model on multiple input features. The score is added to the list of features and the Random Forest ensemble classifier is trained to produce the model. The Cox Enhanced Random Forest (CERF) proposed method is evaluated by comparing its predictive accuracy to those of extant methods.

Keywords: Sepsis, Septic Shock, Machine Learning, Prediction, Predictive Model, Classification, Ensemble Classifier

Acknowledgements

When I decided to pursue a Ph.D. in Computer Science, I predicted, without applying any machine learning techniques, that the journey was going to be rough, challenging, but at the same time fulfilling. The mere fact that I started pursuing this goal while maintaining a full-time job as a software developer, meant that the road to success would be an upward battle that required extra efforts, sleepless nights, and sacrifices from all family members. For that, I would like to express my deepest thanks to my wife and kids, my parents, my brother and my sisters, and my sister-in-law for their continued support and patience throughout the whole adventure.

The utmost thanks go to my advisor Dr. Sumitra Mukherjee for his tremendous support and guidance from the day we set out together. His wisdom and knowledge provided the much-needed light to maneuver through the turbulences of the challenging PhD river. I would like also to thank Dr. Michael Laszlo and Dr. Francisco Mitropoulos, the dissertation committee members, who provided precious input that incredibly enhanced the outcome of this work.

Table of Contents

Abstract iii
Acknowledgments iv
List of Tables vii
List of Figures ix

Chapters

1. Introduction 1
Background 1
Problem Statement 3
Goal 3
Research Question 4
Relevance and Significance 4
Issues 5
Definition of Terms 6
Summary 7

2. Literature Review 8
Introduction 8
Septic Shock Prediction 8
Ensemble Classifiers 12
    Types of ensemble classifiers 13
    Combining methods 14
Ensemble Classifier Usage in the Medical Field 16
Cox Proportional Hazards Model 17
Random Forest 19
Septic Shock Biomarkers 21

3. Methodology 25
Specific Research Method Employed 25
    1. Data Collection 25
    2. Features Selection 28
    3. Data Cleanup and Preparation 32
    4. Prediction Model 35
Summary 42

4. Results 44
Overview 44
Model Results 44
    1. Temperature, HR, RR, MAP, and SI Model 45
    2. Temperature, RR, MAP, Lactate, and WBC Model 46

3. Temperature, RR, Creatinine, Lactate, and WBC Model 47
4. Temperature, HR, Creatinine, Lactate, and WBC Model 48
5. HR, RR, MAP, SBP, and DBP Model 49
6. Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model 49
7. DBP and Albumin Model 50
8. MSI Model 51
9. ageSI, Age, SBP, and Gender Model 52
Selected Model 53
Model Validation 54
Model Comparison 55
Summary 56

5. Conclusions, Implications, Recommendations, and Summary 57
Overview 57
Conclusions 57
Implications 59
Recommendations 60
Summary 61

References 63

List of Tables

Tables

1. Features that may feed into classifiers 28

2. Additional Researched Biomarkers 30

3. Final Features that were used feed into classifiers 32

4. Time Dependent Data Set Sample 34

5. Partitioned Data Sets Detailed Counts (patients with multiple admissions) 35

6. Time Dependent Data Set Sample at Time *t=20* 39

7. Cox Score Calculation Sample 39

8. Confusion Matrix 42

9. Summary of total count of patients, and counts in each class for all models 45

10. Temperature, HR, RR, MAP, and SI Model Coefficients 45

11. Temperature, HR, RR, MAP, and SI Model Confusion Matrix 46

12. Metrics for Temperature, HR, RR, MAP, and SI Model 46

13. Temperature, RR, MAP, Lactate, and WBC Model Coefficients 46

14. Temperature, RR, MAP, Lactate, and WBC Model Confusion Matrix 47

15. Metrics for Temperature, RR, MAP, Lactate, and WBC Model 47

16. Temperature, RR, Creatinine, Lactate, and WBC Model Coefficients 47

17. Temperature, RR, Creatinine, Lactate, and WBC Model Confusion Matrix 47

18. Metrics for Temperature, RR, Creatinine, Lactate, and WBC Model 48

19. Temperature, HR, Creatinine, Lactate, and WBC Model Coefficients 48

20. Temperature, HR, Creatinine, Lactate, and WBC Model Confusion Matrix 48

21. Metrics for Temperature, HR, Creatinine, Lactate, and WBC Model 48

22. HR, RR, MAP, SBP, and DBP Model Coefficients 49

23. HR, RR, MAP, SBP, and DBP Model Confusion Matrix 49

24. Metrics for HR, RR, MAP, SBP, and DBP Model 49

25. Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model Coefficients 50

26. Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model Confusion Matrix 50

27. Metrics for Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model 50

28. DBP and Albumin Model Coefficients 50

29. DBP and Albumin Model Confusion Matrix 51

30. Metrics for DBP and Albumin Model 51

31. MSI Model Coefficients 51

32. MSI Model Confusion Matrix 51

33. Metrics for MSI Model 52

34. ageSI, Age, SBP, and Gender Model Coefficients 52

35. ageSI, Age, SBP, and Gender Model Confusion Matrix 52

36. Metrics for ageSI, Age, SBP, and Gender Model 52

37. Cross Validation Results 55

38. Model Comparisons 59

List of Figures

Figures

1. Patients Selection Criteria 27

2. Cox Model for Temperature, RR, MAP, lactate, and WBC 53

3. Temperature, RR, MAP, lactate, and WBC Model Results 55

Chapter 1

Introduction

Background

Sepsis is an ancient syndrome that has eluded medical practitioners throughout history (Martin, 2012). Hippocrates (460 BC - 370 BC), the Greek physician, talked about rotting flesh and festering wounds as signs of sepsis (Angus & van der Poll, 2013). At a later time, Marcus Terentius Varro, the Roman scholar and writer (116 BC – 27 BC) talked about tiny and invisible airborne creatures that caused dangerous diseases when inhaled (Martin, 2012). Niccolo Machiavelli (1469 – 1527), the Renaissance historian and philosopher, wrote in 1513 about a frenetic fever that was difficult to detect but easy to treat, whereas it would become very difficult to treat but easy to identify at a later stage (Martin, 2012). These syndromes closely matched sepsis (Martin, 2012). With Pasteur and others confirming the germ theory, sepsis was redefined as a systemic infection of the body by pathogenic organisms (germs) that spread in the bloodstream (Angus & van der Poll, 2013). However, despite successfully ridding the body of the invading pathogens, lots of patients did not survive, which led researchers to believe that the body drove the pathogenesis of sepsis not the germs (Angus & van der Poll, 2013). In 1992, the American College of Chest Physicians (ACCP) and the Society of Critical Care Medicine (SCCM) jointly published a consensus definition of sepsis. Sepsis is a systemic inflammatory response of the body due to a microbial infection (King, Bauzá, Mella, &

Remick, 2014; Martin, 2012; Prucha, Bellingan, & Zazula, 2015). This definition remained in effect until 2016, when The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) redefined sepsis as a "life-threatening organ dysfunction caused by a dysregulated host response to infection" (Singer, Deutschman, Seymour, & et al., 2016). In addition, the group of experts of The Third International Consensus found that sepsis and severe sepsis were used interchangeably, thus they eliminated the use of severe sepsis and reclassified the progress of the disease as sepsis that could lead to septic shock (Singer et al., 2016). For the sake of this dissertation, we used both sepsis and severe sepsis diagnosis as they are part of the dataset utilized in this study.

Sepsis is a major worldwide health issue, which leads to death when it progresses to severe sepsis or septic shock (Deepak & Bhat, 2014; Henry, Hager, Pronovost, & Saria, 2015; Marty et al., 2013; Prucha et al., 2015). In the past twenty years, the occurrence of sepsis is increasing not only in developing countries but in Western Europe and the United States as well (Prucha et al., 2015). In the United States, severe sepsis and septic shock will affect 750,000 patients every year resulting in 30% mortality and $15.4 billion in yearly heath care expenditures (Henry et al., 2015; Lausevic & Lausevic, 2012; Lukaszewski et al., 2008; Nguyen et al., 2014).

Medical professionals and researchers have tried early goal-directed therapy to decrease the percentage of deaths in patients suffering from severe sepsis and septic shock (Thiel et al., 2010). They explored timely interventions that involved fluid resuscitation and appropriate antibiotic administration, which proved to optimize the outcomes and reduce mortality (Nguyen et al., 2014; Sawyer et al., 2011).

Despite the progress that has been achieved in the past ten years to detect septic shock early, and despite the advancements in treatment that resulted in reducing mortality, the percentage still remains high (Mohan, Shrestha, Guleria, Pandey, & Wig, 2015; Prucha et al., 2015). The need to implement a system that can identify patients with high risk of septic shock is very crucial (Sawyer et al., 2011). In fact, methods that can identify patients who will experience septic shock in the near future can help improve the outcome (Henry et al., 2015).

Problem Statement

The high mortality rate of sepsis is a major problem that faces the medical and research communities (Deepak & Bhat, 2014; Henry et al., 2015; Lausevic & Lausevic, 2012; Marty et al., 2013; Nguyen et al., 2014; Prucha et al., 2015). Identifying septic shock in a timely manner before it happens is crucial in reducing the mortality rate (Henry et al., 2015; Sawyer et al., 2011).

The septic shock prediction problem was modeled as a binary classification task: patients were classified into two groups – those who were at high risk of suffering a Septic Shock and those who were not – based on information available from clinical observations and laboratory test results. The solution was to train an ensemble classifier on available data and to implement a predictive model for this classification task.

Goal

The study used the extensive data available from the MIMIC-III database to develop a model to predict septic shock. This work explored the ability of the model to increase the accuracy of septic shock prediction before its onset within a certain timeframe. The resulting contribution could help in implementing a forward-looking

computer-assisted decision support in the intensive care unit (ICU), which could allow medical professionals to reduce mortality among sepsis patients.

To support the goal, the dissertation considered the performance of the predictive model at detecting the patients who might have developed septic shock before its onset. First, a cross-validation technique was used to measure accuracy, sensitivity, and specificity (Alberg, Park, Hager, Brock, & Diener-West, 2004; Simon, Subramanian, Li, & Menezes, 2011). An iterative k-fold cross-validation technique, with $k=10$ was used (Beleites, Neugebauer, Bocklitz, Krafft, & Popp, 2013; Refaeilzadeh, Tang, & Liu, 2009). Next, the performance of the model was compared to two different models. The first one was a routine screening protocol for septic shock that uses SIRS criteria, suspicion of infection, and the presence of either hypotension or hyperlactatemia (Henry et al., 2015). The second evaluation was against the TREWScore model – a leading machine learning model developed by Henry et al. (2015).

Research Question

As mentioned, the goal of this study was to develop a model to predict septic shock using ensemble classification. Ensemble classification is known to increase accuracy; thus, the following research question guided the study:

RQ

How can one develop an ensemble model to predict septic shock with acceptable accuracy?

Relevance and Significance

The importance of this research effort is the detection of septic shock before it occurs; therefore, medical professionals can administer the proper on-time treatment to

the patients to reduce the level of mortality (Deepak & Bhat, 2014; Henry et al., 2015; Lausevic & Lausevic, 2012; Marty et al., 2013; Nguyen et al., 2014; Prucha et al., 2015; Sawyer et al., 2011).

Many researchers had contributed to this topic in the past few years. Ho, Lee, and Ghosh (2012) used the MIMIC-II database to construct three different septic shock predictive models with accuracy rate close to 80%. Another significant model is the Quotient Basis Kernel (QBK), which showed a sensitivity of 79.34%, and a specificity of 83.24% (Ribas Ripoll, Vellido, Romero, & Ruiz-Rodríguez, 2014). Henry et al. (2015) used supervised learning methodologies and the MIMIC-II database to construct the targeted real-time early warning score (TREWScore). The model can detect at-risk patients with an accuracy of 0.83 [95% confidence interval (CI), 0.81 to 0.85] at a specificity of 0.67 and a sensitivity of 0.85 within a median of 28.2 [interquartile range (IQR), 10.6 to 94.2] hours before onset (Henry et al., 2015).

The relevance and significance of this research effort is developing an improved method for septic shock prediction using ensemble classification. This new approach to septic shock prediction will increase the prediction accuracy over the previously presented techniques.

Issues

The MIMIC database offers a valuable source of data for clinical and statistical research, however, it used a non-organized and non-standard coding system that led to features' redundancy and ambiguity (Abhyankar, Demner-Fushman, & McDonald, 2012). Besides, the complex nature of clinical data typically suffers from noisy and inconsistent data gathering (Ho, Lee, & Ghosh, 2014; Li, Stuart, & Allison, 2015). For

instance, heart rate was electronically monitored but had to be entered manually into the patient's chart, which led to erroneous or irregular data (Ho et al., 2014). Consequently, a major issue was missing data, which could decrease the dataset size, thus affecting the accuracy of the prediction model (Ho et al., 2014; Li et al., 2015).

Johnson et al. (2016) pointed to the issues that occured during the collection and preprocessing of the clinical data: compartmentalization, corruption, and complexity. Compartmentalization is the distribution of the data across multiple systems, which results in disconnected data that is hard to combine (Johnson et al., 2016). After combining the data, corruption can happen resulting in erronous, missing, or imprecise data (Johnson et al., 2016). Corruption leads to complex data that requires lots of effort to normalize and clean (Johnson et al., 2016). In summary, the available data is noisy and requires significant preprocessing.

Definition of Terms

The following terminologies define measures of predictive accuracy that are used throughout the paper.

True positive (TP)

TP is the prediction or test that correctly identifies the condition when the condition is present (Parikh, Mathai, Parikh, Chandra Sekhar, & Thomas, 2008).

False positive (FP)

FP is the prediction or test that incorrectly identifies the condition when the condition is absent (Parikh et al., 2008).

True negative (TN)

TN is the prediction or test that does not identify the condition when the condition is absent (Parikh et al., 2008).

False negative (FN)

FN is the prediction or test that does not identify the condition when the condition is present (Parikh et al., 2008).

Sensitivity (SN)

SN is the ability of a test to correctly classify a case as positive. It is the probability of testing positive in the presence of a condition (Parikh et al., 2008).

Sensitivity = $\frac{TP}{TP+FN}$

Specificity (SP)

SP is the ability to correctly classify a case as negative. It is the probability of testing negative in the absence of a condition (Parikh et al., 2008).

Specificity = $\frac{TN}{TN+FP}$

Summary

The study aimed at increasing the prediction accuracy of septic shock before its onset. The prediction model was based on a collection of a comprehensive set of features or biomarkers of sepsis and septic shock. The biomarkers were fitted into the Cox proportional hazards model to obtain a score at time *t*. The score was added to the list of biomarkers for the second step. The Random Forest Ensemble was applied to categorize the patients into septic shock class within time *t*, and a No Septic Shock class. The new method, called the Cox Scored Random Forest (CSRF), was based on features that were medically shown to have high impact on the prediction of septic shock.

Chapter 2

Literature Review

Introduction

Researchers realized the importance of predicting septic shock at an early stage after gaining a good understanding of sepsis. The efforts to predict mortality from septic shock started as a manual process to develop a scoring mechanism that uses the available laboratory test results combined with the physicians' clinical observations. With the wide availability of computing equipment, the process of prediction benefited from the usage of automation. Later, researchers started utilizing machine learning techniques to predict the onset of septic shock.

Septic Shock Prediction

Early efforts to predict septic shock used the Limulus amebocyte lysate (LAL) assays for endotoxin (a toxin inside a bacterial cell), but the results were not very successful and accurate (Cohen & McConnell, 1988). A future study refuted these findings as clinical diagnoses had not correlated the presence of endotoxin with multiple organ failure (MOF) patients (Yi et al., 2015). Later, Matsusue, Kashihara, and Koizumi (1988) came up with a scoring system known as the Prognostic Index (PI), which is based

on age, pulse rate, blood urea nitrogen, serum albumin, serum cholesterol and serum potassium. Blomkalns (2006) investigated lactic acid or lactate as a biomarker of septic shock. Lactate does not clear in patients with sepsis, which led the researcher to suggest that increased lactate levels could predict septic shock (Blomkalns, 2006). Chen and Kuo (2007) used heart rate variability (HRV) analysis, which is a technique that observes the variation of beats in the heart rhythm, as an indicator of deterioration for patients with sepsis. The relevance of these efforts is identifying features to use in building prediction models.

Lukaszewski et al. (2008) realized the importance of machine learning techniques to predict the onset of septic shock and created several neural network models. These models used different white blood cells tests (leukocyte IL-1, IL-6, IL-8, IL-10, MCP-1, TNF-@, and FasL) to predict septic shock with 83.09% accuracy (Lukaszewski et al., 2008). The usage of machine learning techniques continued with Wang, Wu, and Wang (2010) making a prediction model based on Support Vector Machine (SVM) to detect severe sepsis (Wang et al., 2010).

Thiel et al. (2010) used the Recursive Partitioning And Regression Tree (RPART) analysis to construct a sepsis prediction model, but the model did not result in high accuracy. Researchers at Barnes-Jewish Hospital in St Louis, Missouri developed a real-time computerized sepsis prediction tool (PT) that utilized partitioning regression tree analysis from data collected from routine laboratory and hemodynamic values (Sawyer et al., 2011). Lausevic and Lausevic (2012) conducted a study to determine septic shock using blood levels of C reactive protein (CRP), immunoreactivity phospholipase A2 group II (PLA2-II), IL-6 and IL-10 concentration, in conjunction with evaluations of

prognostic values of the Simplified Acute Physiology Score (SAPS) II, Injury Severity Score (ISS) score values and multiple organ failure (MOF) signs (Lausevic & Lausevic, 2012).

Ho et al. (2012) constructed three different septic shock predictive models: the first model employed multivariate logistic regression, the second one utilized a linear kernel support vector machine (SVM), and the third used regression trees. The models showed good accuracy rate close to 80% (Ho et al., 2012). The significance of their research is filling missing values using imputation techniques such as the mean feature values and matrix factorization-based approaches (Ho et al., 2012). The imputation process increased accuracy and performance and reduced the use of additional laboratory tests and invasive procedures (Ho et al., 2012). Marty et al. (2013) performed a multivariate logistic regression analysis between the deceased and survivors on lactate clearance and discovered a relation between lactate clearance and concentration and survival status. They concluded that blood lactate concentration and clearance are both an indication of 28-day mortality during severe sepsis or septic shock (Marty et al., 2013).

Researchers at the University of Alabama at Birmingham Hospital developed an automated sepsis detection that would trigger an alert if it met certain criteria based on temperature, respiratory rate, heart rate, and total white blood cell (WBC) count (Nguyen et al., 2014). Deepak and Bhat (2014) presented another effort to predict the outcome of sepsis using C-reactive protein (CRP) and Acute Physiologic and Chronic Health Evaluation (APACHE) II score. Their goal was mainly to contribute a simple, reliable, and inexpensive method utilizing sources that already existed in most medical facilities (Deepak & Bhat, 2014). Ribas Ripoll et al. (2014) presented a sepsis mortality prediction

method using linear algebra, geometry, and statistical inference. They built a kernel for multinomial distributions and named it the Quotient Basis Kernel (QBK), which used the Simplified Acute Physiology Score (SAPS) for ICU patients and the Sequential Organ Failure Assessment (SOFA) to deliver a mortality prediction from sepsis with high accuracy (Ribas Ripoll et al., 2014).

Ho et al. (2014) added a third imputation method to deal with missing data. They incorporated the neighborhood-based imputation that looks for the k-nearest neighbors (KNN) with non-missing data, and takes their mean to fill the missing values (Ho et al., 2014). The significance of the work was allowing models to apply on noisy and incomplete large datasets (Ho et al., 2014). Mohan et al. (2015) analyzed a two-year range of data of patients with sepsis, who were followed from admission until death or discharge from ICU. Their goal was to help formulate better algorithms by offering observation that led to death from septic shock (Mohan et al., 2015).

Henry et al. (2015) used supervised machine learning techniques that consumed different clinical, vital, and laboratory features stored in the MIMIC-II Clinical Database, to develop a model that classifies patients into two groups, one who were at risk of progressing into septic shock and the other who were not at risk (Henry et al., 2015). Based on the model, they built and validated a targeted real-time early warning score (TREWScore) with an accuracy of 83%, (Henry et al., 2015). Mao et al. (2018) used the Gradient tree boosting as an ensemble technique to construct a prediction model utilizing only six vital signs that are routinely checked and measured at medical facilities: systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, peripheral capillary oxygen saturation and temperature. Their model classified patients into Shock and No

Shock with an accuracy of 92%, and four hours before the onset of septic shock it predicted the event with a 96% accuracy (Mao et al., 2018).

Ensemble Classifiers

The study of methods to construct ensemble classifiers is a very active area of research within the field of supervised machine learning (Dietterich, 2000; Ramos-Jimenez, del Campo-Avila, & Morales-Bueno, 2009; Valentini & Masulli, 2002; Zhiwen, Le, Jiming, & Guoqiang, 2015). Single machine learning algorithms or single classifiers search through a space of potential functions or hypotheses to find the best approximation *h* to the unknown function *f* (Dietterich, 2002). The machine learning algorithm determines the best hypothesis by measuring how well a hypothesis *h* matches the function *f* using data points in the training set (Dietterich, 2002). On the other hand, ensemble classifiers construct a set of hypotheses then combine them by taking weighted or unweighted vote (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The result of combining the individual decisions improves the overall performance and delivers a more accurate classification (Dietterich, 2000, 2002; Valentini & Masulli, 2002).

Ensemble classifiers work better because they reduce the inaccuracy of single classifiers (Dietterich, 2000, 2002). Single classifiers suffer from three problems that degrade their performance: statistical, computational, and representational (Dietterich, 2000, 2002). The statistical problem is caused by an insufficient training dataset, which may result in finding multiple optimal hypotheses (Dietterich, 2000, 2002; Valentini & Masulli, 2002). If the algorithm chooses the wrong hypothesis, it will lead to incorrect predictions (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The problem can be resolved by combining the results and getting a better approximation (Dietterich, 2000,

2002; Valentini & Masulli, 2002). The computational problem occurs when the classification algorithm applies local optimization techniques that can get stuck in local minima (optima), hence the algorithm cannot find the best hypothesis (Dietterich, 2000, 2002; Valentini & Masulli, 2002). For example, neural networks employ gradient descent techniques and decision trees apply greedy local optimization approaches in order to minimize error functions over training datasets (Dietterich, 2000, 2002; Valentini & Masulli, 2002). This problem can be reduced or eliminated by applying a weighted combination of the several different local minima (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The representation problem occurs when the space of hypotheses does not contain any good approximation to the unknown function (Dietterich, 2000, 2002; Valentini & Masulli, 2002). In some of these cases, the space can be expanded by combining hypotheses using a weighted sum, which may allow the algorithm to predict a more accurate approximation (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The above-mentioned problems are all resolved or reduced by ensemble classification (Dietterich, 2000, 2002; Valentini & Masulli, 2002), which make ensemble classifiers more accurate, robust, and stable than single classifiers (Zhiwen et al., 2015).

Types of ensemble classifiers

Ensemble classifiers are divided into two groups: non-generative ensembles and generative ensembles (Abad, Zare-Mirakabad, & Rezaeian, 2014; Valentini & Masulli, 2002). Non-generative ensemble methods do not generate new base learners but rather combine a set of well-built base classifiers in a suitable way (Abad et al., 2014; Valentini & Masulli, 2002). Non-generative ensembles use different combining methods, such as employing majority voting to combine the output of a set of base learners, selecting the

best subset of base learners based on their accuracy, or using the Bayes rule to combine the probabilistic output of a set of classifiers (Abad et al., 2014; Valentini & Masulli, 2002). On the other hand, generative ensemble methods generate base classifier by acting on the base learning algorithm or on the structure of the dataset (Abad et al., 2014; Valentini & Masulli, 2002). Generative ensembles work actively to improve diversity and accuracy of the base learners (Abad et al., 2014; Valentini & Masulli, 2002). Examples of generative methods include resampling, feature selection, output coding and mixture of experts, test-and-selection, and randomized methods (Abad et al., 2014; Valentini & Masulli, 2002).

Zhiwen et al. (2015) categorized ensemble classifiers from a different perspective. The first category focuses on how to design and build a new classifier ensemble (Zhiwen et al., 2015). Some examples include: developing graph-based multi-label ensemble classifiers, constructing new classifier ensembles by means of weighted instance selection, and designing a new approach that generates ensembles by clustering data at multiple layers (Zhiwen et al., 2015). The second category concentrates on theoretically exploring and analyzing the properties of a classifier ensemble (Zhiwen et al., 2015). One example is eliminating the redundant classifiers in the ensemble by using an instance-based pruning approach (Zhiwen et al., 2015). Another one is improving the efficiency of the ensemble classifiers using rule migration mechanisms (Zhiwen et al., 2015).

Combining methods

One of the main research areas for ensemble classifiers is the methods to combine the base classifiers to form the ensemble (Verma & Rahman, 2012). The most popular

combining methods are bagging, boosting, and random subspace method (Bagheri & Gao, 2012; Ghavidel, Yazdani, & Analoui, 2013).

The bagging method is a sampling-based approach that uses multiple datasets to generate base classifiers and combine them into the ensemble classifier (Ren & Suganthan, 2012; Valentini & Masulli, 2002; Verma & Rahman, 2012). The training datasets are randomly bootstrapped (drawn with replacement) from the entire training set (Ren & Suganthan, 2012; Valentini & Masulli, 2002; Verma & Rahman, 2012). The aggregation of the base classifiers takes place after performing an average by a majority or weighted vote (Ren & Suganthan, 2012; Valentini & Masulli, 2002; Verma & Rahman, 2012). Bagging works better for small datasets, and improves performance if the induced classifiers are good and not correlated; however, if smaller datasets are used to train individual classifiers, bagging may slightly reduce the performance of some stable algorithms such as the k-nearest neighbor (Bauer & Kohavi, 1999). Besides, sampling for large datasets based on the bootstrap with replicates of the training datasets is not practical (Verma & Rahman, 2012). Bootstrap replicates of large training sets have similar statistical characteristics, since large sets show the real data distribution well (Skurichina, Kuncheva, & Duin, 2002). This will result in constructing similar classifiers and the ensemble will become less diverse and thus less accurate (Skurichina et al., 2002). The randomness introduced by the sampling process in bagging can affect the performance of the ensemble classifier (Verma & Rahman, 2012).

Boosting is an iterative method that generates the base classifiers sequentially (Bauer & Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). For new iterations, the learning algorithm uses a different distribution of the training data (Bauer

& Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). The instance of the training data is assigned a weight in the new iteration based on the performance on the prior iteration (Bauer & Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). Boosting works on the instances of the training data that are hard to classify (Bauer & Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). Such instances have higher weights, which indicate that they are not accurately classified and thus will be included in the next iterations (Bauer & Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). However, boosting does not offer a mechanism to enhance the learning of base classifiers for these instances (Bauer & Kohavi, 1999; Valentini & Masulli, 2002; Verma & Rahman, 2012). The final ensemble classifier is formed by combining the base classifiers using a weighted majority vote (Bauer & Kohavi, 1999; Valentini & Masulli, 2002).

The Random Forest ensemble classifier is based on a collection of tree classifiers (Breiman, 2001; Pal, 2005). Each classifier is generated from a random set of features independently sampled from the input features, and each classifier has a single vote to choose the most popular class to classify the input (Breiman, 2001; Pal, 2005).

Ensemble Classifier Usage in the Medical Field

The use of ensemble classifiers in septic shock prediction has not been established. However, other domains of medical diagnosis benefited from the use of ensemble classifiers to predict progression of diseases and traumatic health situations (Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Srimani & Koti, 2013). Lavanya and Rani (2012) presented an ensemble classifier based on a hybrid of decision trees that relied on the bagging technique to improve the accuracy of breast cancer

prediction. Kelarev, Stranieri, Yearwood, Abawajy, and Jelinek (2012) used ensemble classification, namely the Random Forest, to build a model that outperformed all base classifiers in predicting cardiac autonomic neuropathy (CAN). Williams, Weakley, Cook, and Schmitter-Edgecombe (2013) used single classification techniques, such as naïve Bayes (NB), C4.5 decision tree (DT), back-propagation neural network (NN), and support vector machine (SVM) to detect mild cognitive impairment and dementia, but suggested exploring ensemble classifiers in future studies (Williams et al., 2013). Ali, Majid, and Khan (2014) built multiple ensemble classifiers using various learning algorithms such as Random Forest (RF), SVM, and KNN that performed very well in their experiments (Ali et al., 2014). To predict cancer survivors, Gupta et al. (2014) built three models, where each is an ensemble of 400 SVMs. The study determined that the use of the ensemble classifiers could boost prediction over conventional methods (Gupta et al., 2014). Yao, Guo, and Yang (2015) proposed an ensemble classification tool, which used Random Forests, to predict protein-protein interaction (PPI) networks. Morino et al. (2015) adopted an ensemble classification that generated accurate predictions when tested on a dataset for prostate cancer patients (Morino et al., 2015).

Cox Proportional Hazards Model

The Cox Proportional Hazards (CPH) model has been widely used for survival analysis for censored data (Bonato et al., 2011; Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Tsujitani, Tanaka, & Sakon, 2012). It is one of the most popular models in statistical analysis (Bonato et al., 2011; Wang, Shen, & Thall, 2014). The CPH model is used extensively in clinical and epidemiological studies to mainly estimate the risk ratio (Lin, Chang, & Liao, 2013).

Lin et al. (2013) used small events per predictive variables (EPVs) in Cox regression models to analyze the relationships between protracted low-dose radiation exposure and incidence of leukemia. Wang et al. (2014) proposed a modified Lasso method for the Cox regression model that used adaptive selections of important single covariates. This method had tremendous numerical advantage, especially for survival analysis in biomedical studies, as it helped in identifying key treatment–biomarker interactions to develop individualized treatments (Wang et al., 2014).

Tolosie and Sharma (2014) used the Cox proportional hazards model for multivariate analysis and model building to identify the factors associated with death from tuberculosis. Jackson and Cox (2014) proposed a method to add robustness to the continuous covariate model in the Cox proportional hazards that automatically guards against extreme values and sets asymptotes for the minimum and maximum hazard ratios. The extended model was very useful in clinical studies (Jackson & Cox, 2014). Xu, Sen, and Ying (2014) investigated the consistency of bootstrapping on the Cox proportional hazards model. Honda and Karl Härdle (2014) concentrated on time-varying coefficient Cox regression models to enhance prediction. Wang et al. (2015) proposed an approach called Time Slicing Cox regression (TS-Cox) based on a combination of time-series feature extraction and time-slicing Cox regression method. The new model was applied to predict mortality in ICUs (Wang et al., 2015). Guilloux, Lemler, and Taupin (2016) used high-dimensional covariates with an adaptive estimator of the baseline function in the Cox model, which performed well with simulation data. Wu, Zheng, and Yu (2016) proposed a statistical method based on a semiparametric Logistic-Cox mixture model that worked reasonably for practical sample sizes. Lee, Hudgens, Cai, and Cole (2016)

considered estimating the parameters in the semiparametric marginal structural Cox model to accommodate the effect of prior treatments in biomedical studies. The estimator allowed consistency and asymptotic normality results (Lee et al., 2016).

Random Forest

The Random Forest ensemble classifier has been used on many datasets spanning different environments and industries. The Random Forest ensemble is preferred over other ensembles because it is simple, can be easily parallelized, is relatively robust to outliers and noise, is faster than bagging or boosting, and supplies valuable inside estimates of error, strength, correlation, and variable importance (Breiman, 2001). Besides, Breiman (2001) claims that it is as accurate as Adaboost and occasionally better.

Cutler et al. (2007) used Random Forest on ecology-based datasets and listed several advantages. Compared to other classifiers, Random Forest has the following advantages: classification with very high accuracy; determination of variable importance; flexibility to do classification, survival analysis, regression, and unsupervised learning; capability to model complicated exchanges among features; and the ability to be used as an algorithm to impute missing values (Cutler et al., 2007).

Random Forest is a nonparametric tree-based ensemble classifier that combines the concepts of adaptive nearest neighbors and bagging to effectively infer data (Chen & Ishwaran, 2012). It is a widespread ensemble learning method, which is highly used in data mining and machine learning (Chen & Ishwaran, 2012). The researchers used Random Forest on high-dimensional genomic data analysis, where the results led them to conclude that it predicted outcome accurately (Chen & Ishwaran, 2012).

Lebedev et al. (2014) used Random Forests to predict the onset of Alzheimer's. According to the researchers, Random Forests produced the highest accuracies compared to other algorithms due to its abilities to handle non-linear and high-dimensional data, its robustness to noise, its tuning simplicity, and its effectiveness in parallel processing (Lebedev et al., 2014). In another study, Dauwan et al. (2016) built a Random Forest classifier to enhance the accuracy of differentiating the diagnosis of dementia with Lewy bodies (DLB) from Alzheimer's disease. The Random Forest ensemble is widely and efficiently used in various areas of computational biology (Jia, Liu, Xiao, Liu, & Chou, 2016).

Xia et al. (2015) utilized and enhanced Random Forests to classify hyperspectral images. The ensemble worked efficiently on large data sets with high classification accuracy (Xia et al., 2015). Kulkarni and Lowe (2016) also used Random Forest for analysis of imagery for land cover and achieved excellent accuracy.

Insurance big data analysis is another area that Random Forest ensemble outperformed other classification algorithms, such as SVM (Lin, Wu, Lin, Wen, & Li, 2017). Random Forest was better in terms of accuracy and performance within the imbalanced insurance data, and it improved the accuracy of product marketing in comparison to the non-machine learning approaches (Lin et al., 2017).

Random Forest ensemble proved its superiority in classification and prediction of many areas, such as hyperspectral imagery, medical diagnosis, insurance, and Genomics (Chen & Ishwaran, 2012; Dauwan et al., 2016; Lin et al., 2017; Xia, Ghamisi, Yokoya, & Iwasaki, 2018). The interest in the Random Forest ensemble is a result of its following advantages: high performance and rapid prediction; obliviousness to high-dimensional

features; simple parameter tuning; and ability to rank features' importance (Xia et al., 2018).

Septic Shock Biomarkers

In 2001, the National Institutes of Health announced a broad definition of biomarkers as " a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." ("Biomarkers and surrogate endpoints," 2001). Researchers had presented multiple biomarkers for Septic Shock. Rivers et al. (2007) mentioned IL-1ra (150 –30,000 pg/mL), ICAM-1 (2.5–900 ng/ mL), TNF-α (20 –2,000 pg/mL), Caspase-3 (0.1–200 ng/mL), and IL-8 (15–3,000 pg/mL) as biomarkers that change according to Lactate level.

Phua, Koay, and Lee (2008) compared the prognostic utility of biomarkers lactate, procalcitonin (ProCT), and amino-terminal pro-B-type natriuretic peptide (NT-proBNP). The biomarkers were measured together with serum IL-1β, IL-6, IL-10, and TNF-α levels. The researchers concluded that increased lactate levels yielded better prediction than ProCT levels and in turn ProCT were more accurate than NT-proBNP levels. The researchers suggested that serial lactate and ProCT measurements may be used together to enhance the results (Phua et al., 2008). Other studies showed that ProCT is elevated in patients with sepsis, which qualified ProCT as an acceptable biomarker (Azevedo et al., 2012; Becker, Snider, & Nylen, 2010; Kibe, Adams, & Barlow, 2011; McLean, Tang, & Huang, 2015; Riedel, 2012).

Shapiro et al. (2009) defined a panel of biomarkers consisting of neutrophil gelatinase-associated lipocalin, interleukin-1ra, and Protein C. This panel of biomarkers

was a good predictor of severe sepsis, septic shock, and death of patients with suspected sepsis in Emergency Departments (Shapiro et al., 2009).

Lorente et al. (2009) studied the predictive value of Matrix metalloproteinases (MMPs), namely MMP-9 and MMP-10, and tissue inhibitor of matrix metalloproteinases-1 (TIMP-1). The researchers found that patients with sepsis had higher levels of MMP-10 and TIMP-1, higher MMP-10/TIMP-1 ratios, and lower MMP-9/TIMP-1 ratios than did healthy controls. Sepsis patients who did not survive had lower levels of MMP-9, higher levels of TIMP-1, lower MMP-9/TIMP-1 ratio, higher levels of IL-10, and lower TNF-α/IL-10 ratio than did patients who survived (Lorente et al., 2009).

Mikkelsen et al. (2009) found that serum lactate was linked to death independent of clinically apparent organ dysfunction and shock in ED patients with severe sepsis. In another study, Nguyen et al. (2010) found that early lactate clearance decreased the possibility of a septic shock.

Hattori et al. (2009) investigated protein YKL-40 as a potential biomarker of septic shock. The researchers found that the serum levels of YKL-40 were considerably higher and were positively associated with blood levels of IL-6 in patients at risk of getting a septic shock, which suggested that YKL-40 is a biomarker of sepsis (Hattori et al., 2009).

Sturgess et al. (2010) examined diastolic dysfunction, particularly E/é (peak early diastolic transmitral/peak early diastolic mitral annular velocity), as an indicator of septic shock. They concluded that E/é can be used as a predictor of survivability among sepsis patients (Sturgess et al., 2010).

Ricciuto et al. (2011) found that lower angiopoietin-1 plasma levels and higher levels of angiopoietin-2 are associated with death in sepsis patients, which suggested that these two can be used as an indicator of septic shock. The combination of myeloid cells-1 (sTREM-1), ProCT, and polymorphonuclear (PMN) CD64 index was studied as a viable bio score for sepsis (Gibot et al., 2012; Reinhart, Bauer, Riedemann, & Hartog, 2012).

Rivers et al. (2013) suggested the following as biomarkers: interleukin 1β (IL-1 β), IL-1ra, IL-6, IL-8, IL-10, intercellular adhesion molecule (ICAM), tumor necrosis factor-α (TNF-α), caspase 3, D-dimer, high-mobility group protein 1 (HMGB1), vascular endothelial growth factor (VEGF), matrix metalloproteinase (MMP), and myeloperoxidase (MPO).

Berger et al. (2013) used vital signs such as temperature, heart rate (HR), respiratory rate (RR), mean arterial pressure (MAP), and shock index as the features to identify septic shock. Their analysis achieved results similar to SIRS.

Malmir, Bolvardi, and Afzal Aghaee (2014) suggested serum lactate as an indicator of septic shock. The increased level of lactate in patients arriving at the ER was associated with higher death rate.

Gultepe et al. (2014) claimed to achieve an accuracy of 0.99 by utilizing lactate level, temperature, RR, and MAP, and white blood cells (WBC). The researchers used the naïve Bayes algorithm for classification, Gaussian mixture model for clustering, and hidden Markov model for probability distribution.

Carrara, Baselli, and Ferrario (2015) proposed different models that achieved good accuracy levels. The first model was based on RR, temperature, WBC, creatinine,

and lactate. The second one used HR, creatinine, WBC, temperature, and lactate, while the third model utilized SBP, DBP, MAP, HR, RR and cardiac output.

As biomarkers for septic shock, Prucha et al. (2015) suggested C-reactive protein, procalcitonin, cytokines, Lipopolysaccharide binding protein (LBP), and leukocytes. The authors believed that the accuracy of the biomarkers could help in diagnosing the progression of the disease, which would help in the choice of the best treatment.

A group of researchers suggested GCS, HR, RR, SpO_2, temperature, SBP, and DBP as good indicators of septic shock. They applied machine learning techniques to deliver high accuracy results with mostly vital signs (Desautels et al., 2016). Kelly et al. (2016) suggested another combination of biomarkers consisting of α-2 macroglobulin (A2M) and ProCT.

Holder et al. (2016) associated low DBP and serum albumin with the progression to septic shock. Their study showed that an initial level of serum albumin <3.5 g/dL and DBP <52 mmHg has a significant statistical association with progress from sepsis to septic shock.

Sundén-Cullberg et al. (2017) studied the effect of fever in septic patients in the ER who were later admitted to the ICU. Their findings contradicted the common perceptions and current procedures of care of septic patients. They observed that increased body temperature in the ER lowered the mortality rate and shortened the hospital stay for these patients.

Chapter 3

Methodology

Specific Research Method Employed

The goal of this research was to improve prediction of septic shock by using ensemble classifiers. The objective was to predict the onset of a septic shock within 10 to 95 hours before its occurrence. The proposed solution consisted of data collection, feature selection, data cleanup and preparation, training prediction models, validation process, and results based on out of sample examples.

1. Data Collection

The study used data from the MIMIC-III database v1.3, which is a relational database containing data of ICU patients at Beth Israel Deaconess Medical Center (Goldberger et al., 2000; "MIMIC-III Clinical Database," 2015). MIMIC-III is an open access database developed by the MIT Lab for Computational Physiology, containing de-identified health data for more than 40,000 critical care patients, including demographics, vital signs, laboratory tests, medications, and more (Goldberger et al., 2000; "MIMIC-III Clinical Database," 2015). MIMIC-III is an extension of MIMIC-II and augments it with newly collected data between 2008 – 2012 (Goldberger et al., 2000; "MIMIC-III Clinical Database," 2015). The MIMIC-III database v1.3 has records of 46,520 ICU patients with 58,976 admissions (a patient could have multiple admissions), collected at Beth Israel Deaconess Medical Center between 2001 – 2012 (Goldberger et al., 2000; "MIMIC-III

Clinical Database," 2015). The information included laboratory data, therapeutic intervention profiles such as vasoactive medication drip rates and ventilator settings, nursing progress notes, discharge summaries, radiology reports, provider order entry data, International Classification of Diseases, 9th Revision codes, and, for a subset of patients, high resolution vital sign trends and waveforms (Saeed et al., 2011). The privacy of patients was preserved by removing all Protected Health Information (PHI) in order to comply with Health Insurance Portability and Accountability Act standards (Saeed et al., 2011). The database was opened for free access to researchers on February 2010 through the Internet and was accompanied by a detailed manual and data processing tools (Saeed et al., 2011).

The data of the MIMIC-III is temporal. Most fields are time-stamped. Some fields were updated hourly, while others were updated every four hours. Patients were tracked from the time they entered the ICU, this is time where $t=0$, until patients got released from the ICU or passed away. The database had 4,683 patients who were diagnosed with sepsis or severe sepsis (ICD-9 codes: 99591 and 99592), and who were 15 years and older. These patients had 8,696 admissions with 2,585 cases resulting in septic shock (ICD-9 code 785.52).

In this dissertation, we treated patients with multiple admissions as separate cases, that is, we included all the admissions of ICUs patients (Verburg, Holman, Dongelmans, de Jonge, & de Keizer, 2018). Each case contributed to the training of the prediction model. The patients' information and their associated clinical, vital, laboratory test results, and other information were downloaded from the MIT Lab for Computational Physiology as text files, then uploaded to a PostgreSQL database as per instructions and

scripts from the MIT Lab. The required data that included patients' information, admissions, and chart and lab info were extracted from the PostgreSQL database to a Microsoft SQL Server Database for faster processing. The detailed data selection criteria are shown in figure 1.

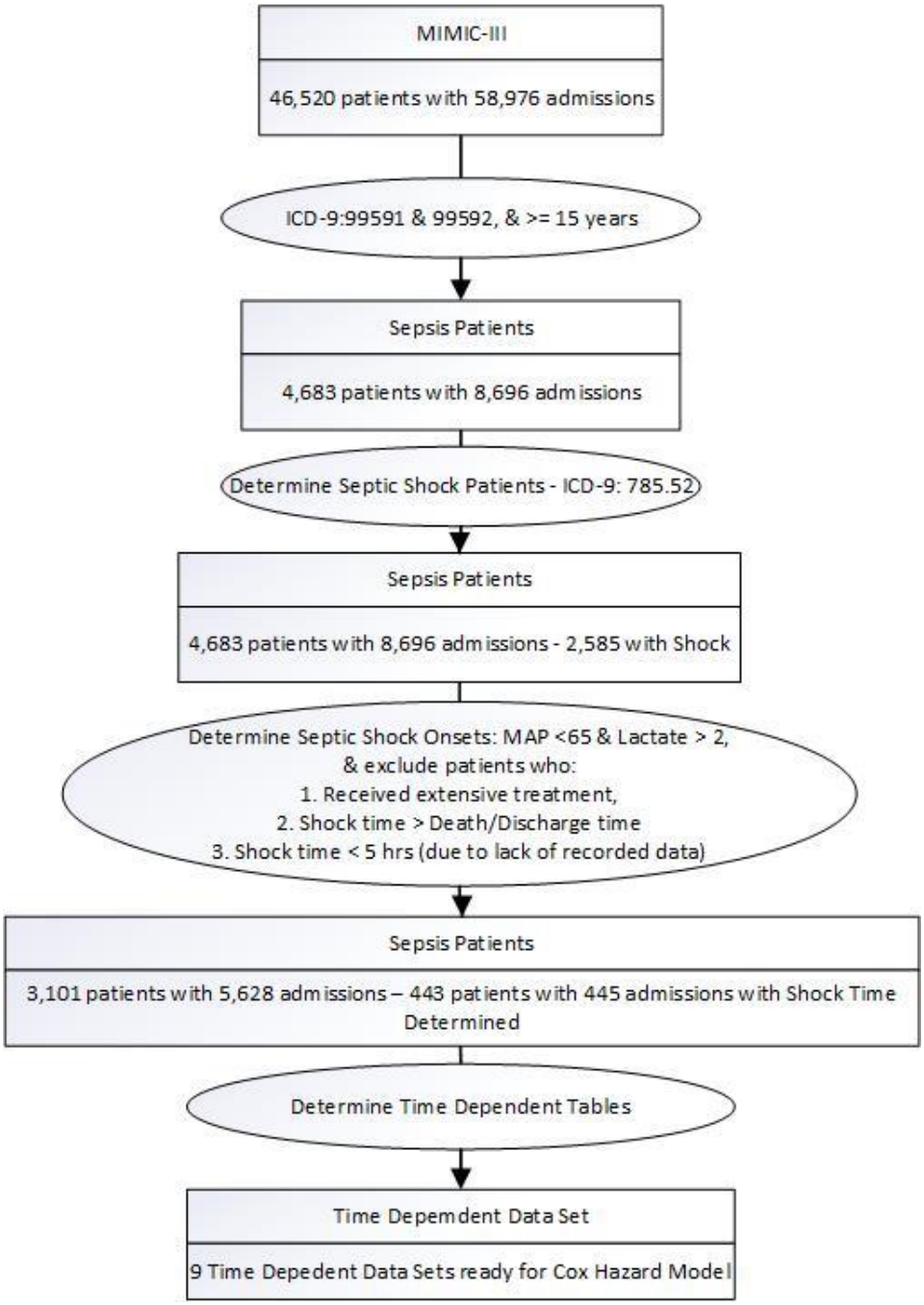

Figure 1 - Patients Selection Criteria

2. Features Selection

Based on the literature review and the established medical standards that define septic shock, we started with a comprehensive set of 46 features, which had good recorded measurements. Only the ones that delivered the best prediction results would be included in the methodology. Table 1 summarizes the list of features:

Table 1

Features that may feed into classifiers

Category	Feature Name	Feature Description	Type	Values/Unit
Clinical	Time since first antibiotics*	Number of Minutes from time antibiotics was first administered in the ICU	Numeric	Minutes
Clinical	6hr Urine Volume*	Total output of urine in the past 6 hours	Numeric	mL
Clinical	Chronic liver disease and cirrhosis	Presence of chronic liver disease and cirrhosis as specified by ICD-9 code 571	Binary	Yes/No
Clinical	Cardiac surgery patient	Patient recovering from a cardiac surgery	Binary	Yes/No
Clinical	Immunocompromised	A patient who received past therapy that suppresses resistance to infection as specified by presence of any ICD-9 in V58.65, V58.0, V58.1, 042, 208.0, 202	Binary	Yes/No
Clinical	SIRS*	Currently showing a minimum of two SIRS criteria	Binary	Yes/No
Clinical	Hematological malignancy	Presence of hematologic malignancy as specified by any ICD-9 code in 200-208	Binary	Yes/No
Clinical	Chronic heart failure	Presence of heart failure as specified by ICD-9 code 428	Binary	Yes/No
Clinical	Chronic organ insufficiency	Such as chronic liver disease, chronic heart failure, chronic respiratory failure, receiving chronic dialysis as specified by one of the ICD-9 codes 571, 585.6, 428.22, 428.32, 428.42, 518.83	Binary	Yes/No
Clinical	Diabetes	Patience is diabetic as specified by ICD-9 code 250	Binary	Yes/No

Clinical	Metastatic carcinoma	As specified by presence of any ICD-9 codes in 140-165, 170-175, 179-199	Binary	Yes/No
Clinical	HIV	Presence of the human immunodeficiency virus (HIV)	Binary	Yes/No
Clinical	Dialysis	The patient is currently undergoing dialysis	Binary	Yes/No
Clinical	Chronic renal insufficiency	The presence of chronic kidney disease caused by damage to the kidneys	Binary	Yes/No
Laboratory	BUN/CR*	The ratio of BUN/creatinine	Numeric	10:1-20:1
Laboratory	Arterial pH	The pH of the blood measured by an arterial line	Numeric	7.35-7.45
Laboratory	PaO_2	Partial pressure of arterial oxygen	Numeric	75-100 mm Hg
Laboratory	BUN	Blood urea nitrogen	Numeric	8-21 mg/dL
Laboratory	Hepatic SOFA*	Hepatic SOFA score calculated based on the bilirubin concentration	Numeric	1-4
Laboratory	WBC	White blood cell count	Numeric	4-10 x 10^9/L
Laboratory	Renal SOFA*	Renal SOFA score calculated on the basis of creatinine concentration	Numeric	1-4
Laboratory	Platelets	The count of Platelet in the bloodstream	Numeric	150-400 x 10^9/L
Laboratory	Glucose	The sugar level in the bloodstream	Numeric	65-110 mg/dL
Laboratory	Chloride	The level of chloride in the blood	Numeric	95-105 mmol/L
Laboratory	Lactate	The presence of lactic acid in the body	Numeric	50-150 U/L
Laboratory	Sodium	The level of sodium in the blood	Numeric	135-145 mmol/L
Laboratory	$PaCO_2$	The level of Partial pressure of arterial carbon dioxide	Numeric	35-45 mm Hg
Laboratory	Creatinine	The level of creatinine (chemical waste product that's produced by your muscle metabolism) in the blood	Numeric	0.8-1.3 mg/dL
Laboratory	Potassium	The level of potassium in the blood	Numeric	3.5-5 mmol/L
Laboratory	Hematocrit	The percentage of the volume of whole blood that is made up of red blood cells	Numeric	40%-52% (men), 36%-47% (women)
Laboratory	Hemoglobin	The level of hemoglobin, which is the protein molecule in red blood cells	Numeric	13-17 g/dL (men), 12-15 g/dL (women)
Laboratory	Aspartate aminotransferase	The level of this enzyme in the body	Numeric	5-30 U/L
Laboratory	C-reactive protein	The level of C-reactive protein (CRP) in the blood	Numeric	< 5 mg/L

Vital	HR	Heart rate	Numeric	60-100 beats/min
Vital	SBP	Systolic blood pressure	Numeric	60-90 mm Hg
Vital	Shock index*	HR/SBP ratio	Numeric	0.5-0.7
Vital	GCS	Glasgow coma score (GCS)	Numeric	3-15
Vital	RR	Respiratory rate	Numeric	Adults: 12-18 breaths per minute
Vital	FiO_2	Fraction of inspired oxygen	Numeric	21%-100%
Vital	Neurologic SOFA*	Neurologic SOFA score calculated on the basis of GCS	Numeric	1-4
Vital	SpO_2	The estimation of the oxygen concentration in the blood	Numeric	96%-100%
Vital	Admission weight	The patient's weight at admission	Numeric	Kg
Vital	Hypotension	The presence of low blood pressure symptoms	Binary	Yes/No
Vital	Current weight	The continuous measurement of the patient's weight	Numeric	Kg
Vital	DBP	Diastolic blood pressure	Numeric	120-139 mm Hg
Vital	Age	Age of patient	Numeric	Years

Note. * Calculated Feature from the electronic health record (EHR)

Additionally, the features listed in Table 2 were extracted from the literature as biomarkers, which can predict septic shock when used individually or as a panel of features. Those features had sparse or no measurements recorded, nevertheless they were listed to raise awareness to start collecting these in future studies.

Table 2

Additional Researched Biomarkers

Biomarker	Category
IL-1ra	Laboratory
ICAM-1	Laboratory
TNF-α	Laboratory
Caspase-3	Laboratory
IL-8	Laboratory
Procalcitonin (ProCT)	Laboratory
Amino-terminal pro-B-type Natriuretic Peptide (NT-proBNP).	Laboratory
IL-1β	Laboratory
IL-6	Laboratory
IL-10	Laboratory
Lipocalin	Laboratory
Protein C	Laboratory

MMP-9	Laboratory
MMP-10	Laboratory
Tissue Inhibitor of Matrix Metalloproteinases-1 (TIMP-1)	Laboratory
TNF-α/IL-10 ratio	Calculated
MMP-9/TIMP-1 ratio	Calculated
YKL-40	Laboratory
Diastolic Dysfunction (E/é)	Clinical
Angiopoietin-1	Laboratory
Angiopoietin-2	Laboratory
sTREM-1, and polymorphonuclear (PMN) CD64	Laboratory

To narrow down the long list, we looked at previous recommendations. Serum lactate was one of the features suggested by many researchers as an indicator of septic shock (Lee & An, 2016; Malmir et al., 2014; Mikkelsen et al., 2009; Phua et al., 2008). Berger et al. (2013) used vital signs such as temperature, heart rate (HR), respiratory rate (RR), mean arterial pressure (MAP), and shock index (SI) as the features to identify septic shock.

Gultepe et al. (2014) utilized temperature, RR, MAP, lactate level, and white blood cells (WBC). As biomarkers for septic shock, Prucha et al. (2015) suggested C-reactive protein, procalcitonin, cytokines, Lipopolysaccharide binding protein (LBP), and WBC. Carrara et al. (2015) proposed 3 different models: first model was based on temperature, RR, creatinine, lactate, and WBC, the second one used temperature, HR, creatinine, lactate, and WBC, and the third model utilized SBP, DBP, MAP, HR, RR and cardiac output.

GCS, HR, RR, SpO_2, temperature, SBP, and DBP were suggested as good indicators of septic shock (Desautels et al., 2016). Holder et al. (2016) associated low DBP and serum albumin with the progression to septic shock. Sundén-Cullberg et al. (2017) suggested temperature as an indicator of septic shock and concluded that fever slows the process. Modified shock index (MSI) has emerged as an early non-invasive

measure, which is calculated by dividing HR by MAP (Jayaprakash, Gajic, Frank, & Smischney, 2018; Torabi, Moeinaddini, Mirafzal, Rastegari, & Sadeghkhani, 2016). Torabi et al. (2016) introduced a new calculated measure age SI (ageSI), defined as age multiplied by SI, and used it with gender and SBP as predictors. Table 3 summarizes the final set of features.

Table 3

Final Features that were used to feed into classifiers

Category	Feature Name	Feature Description	Type
Laboratory	WBC	White blood cell count	Numeric
Laboratory	Lactate	The presence of lactic acid in the body	Numeric
Laboratory	Creatinine	The level of creatinine (chemical waste product that's produced by your muscle metabolism) in the blood	Numeric
Laboratory	C-reactive protein	The level of C-reactive protein (CRP) in the blood	Numeric
Laboratory	Albumin	Albumin test checks liver and kidney function	Numeric
Vital	HR	Heart rate	Numeric
Vital	SBP	Systolic blood pressure	Numeric
Vital	SI	HR/SBP ratio	Numeric
Vital	GCS	Glasgow coma score (GCS)	Numeric
Vital	RR	Respiratory rate	Numeric
Vital	SpO_2	The estimation of the oxygen concentration in the blood	Numeric
Vital	DBP	Diastolic blood pressure	Numeric
Vital	Age	Age of patient	Numeric
Vital	Temperature	Body Temperature	Numeric
Calculated	MAP	Mean Arterial Pressure	Numeric
Calculated	AgeSI	SI enhanced with age	Numeric
Calculated	MSI	Modified Shock Index	Numeric

3. Data Cleanup and Preparation

MIMIC-III is an extension of MIMIC-II, and inherits all it properties (Goldberger et al., 2000; "MIMIC-III Clinical Database," 2015). The databases have missing values,

duplicate values, and wrongly recorded ones. They required massive attention as they would affect the prediction models (Ho et al., 2012; Ho et al., 2014).

The first step was to extract the data for each feature and place it in a temporary table. The measurements of features were time-stamped, where the time of service was based on the admission date. The difference between the time of measurements and the admission date was binned hourly to unify the measurements across all features (Desautels et al., 2016; Mao et al., 2018). Outliers and values with wrong data types were eliminated by nulling them out, thus they will be treated as missing values (Ho et al., 2012; Ho et al., 2014). For any hours with multiple values, the mean values were used, and for missing values a carry forward approach was applied, where the latest bin value was propagated till it reached a bin with a value (Desautels et al., 2016; Mao et al., 2018). In case the first value was missing, the imputation followed a carry backward approach (Desautels et al., 2016; Mao et al., 2018).

In the second step, we determined the onset of the septic shock as it was not identified clearly and had to be calculated. As per Singer et al. (2016), the start of septic shock was the first occurrence of: (1) persistent low blood pressure that required the use of vasopressors (compounds that caused the blood vessels to tighten in order to raise blood pressure) to maintain MAP >= 65mmHg, and (2) serum lactate level >2 mmol/L (18mg/dL) even with adequate volume resuscitation. We excluded patients: (1) who received extensive treatment as they would affect the outcome (Henry et al., 2015), (2) whose Shock time > Death/Discharge time, and (3) whose Shock time < 5 hours due to lack of recorded data. The end result was 3,101 patients with 5,628 admissions, which included 443 patients with 445 admissions having septic shock time determined.

The third step was reconstructing the feature tables and transforming them into time-dependent sets (Therneau, Crowson, & Atkinson, 2018). The time-dependent set has the following columns: (a) ID, which defined each subject uniquely; (b) One or more features – one or more columns where each one represented a feature that fed the Cox model; (c) Full Time, which was the time the event or censor (discharge/death) happened; (d) Start – the start of the time bin; (e) Stop – the end of the time bin; and (f) Event, which is the occurrence or not of the event (Therneau et al., 2018). This data frame allowed running the Cox Hazard Model in order to obtain the hazard coefficients for the risk score calculations (Kim, Park, & Kon, 2013). Table 4 illustrated a sample of a time dependent data set for one patient and one feature, where subject_id and hadm_id (hospital admission id) both represented the unique id, Lactate was an input feature, FullTime was the onset of Septic Shock or death/discharge, tStart and tStop were the beginning and end of the hourly time bin, and Shock was the event. The format of a time-dependent data set mandated that at the end of each time bin the event remained zero till the full time was satisfied, then the event would be recorded as either true or false (Therneau et al., 2018).

Table 4

Time Dependent Data Set Sample

subject_id	hadm_id	Temp	RR	MAP	Lactate	WBC	FullTime	tStart	tStop	Shock
250	124271	36.39	37	94	1.2	17.6	23	0	1	0
250	124271	36.44	40	94	1.2	17.6	23	1	2	0
250	124271	36.44	38	94	1.2	17.6	23	2	3	0
250	124271	36.67	31	94	1.2	17.6	23	3	4	0
250	124271	37.94	38	94	0.8	17.6	23	4	5	0
250	124271	37.94	39	94	0.8	17.6	23	5	6	0
250	124271	37.94	36	94	0.8	17.6	23	6	7	0
250	124271	37.94	18	94	0.8	17.6	23	7	8	0

250	124271	37.94	23	88	0.8	12.7	23	8	9	0
250	124271	37.94	17	96	0.8	12.7	23	9	10	0
250	124271	37.94	23	77	0.8	12.7	23	10	11	0
250	124271	39	24	90	1.1	12.7	23	11	12	0
250	124271	39	26	97	1.1	12.7	23	12	13	0
250	124271	38.67	24	99	1.1	12.7	23	13	14	0
250	124271	38.67	36	74	1.1	12.7	23	14	15	0
250	124271	36.94	29	101	1.1	12.7	23	15	16	0
250	124271	36.94	33	68	1.1	12.7	23	16	17	0
250	124271	35.67	33	69	1.4	27	23	17	18	0
250	124271	35.67	35	67	1.4	27	23	18	19	0
250	124271	36.44	32	75	1.4	27	23	19	20	0
250	124271	36.44	32	70	3.9	27	23	20	21	0
250	124271	36.5	35	70	3.9	27	23	21	22	0
250	124271	36.5	35	62	3.9	21.5	23	22	23	1

The fourth step was to randomly partition the newly formed dataset into an 80% training set and a 20% test set. For validation purposes, the training set was further partitioned into a 10-fold cross validation sets. Table 5 summarizes the actual numbers of the training and test sets as well as the breakdown of each class.

Table 5

Partitioned Data Sets Detailed Counts (patients with multiple admissions)

	Total	0	1
Training Set	4,502	4132	370
Test Set	1126	1051	75

4. Prediction Model

In this dissertation, we developed a prediction model for septic shock based on the features extracted from the MIMIC-III database and were listed in table 3. The prediction model was an extended version of the Random Forest Ensemble called the Cox Enhanced Random Forest (CERF). In this new method, we produced nine preliminary models each consisting of different sets of features from table 3. We generated the Cox hazard

coefficients for each of the nine models. For each model, we calculated Cox risk scores as linear combinations of the features at time *t*, weighted by multivariate Cox proportional hazard coefficients (Kim et al., 2013). We added the score to each model and applied the Random Forest ensemble to determine the final classification at *t* hours before the onset of the shock. We then chose the model that produced the highest accuracy, sensitivity, and specificity as the prediction model for CERF. The detailed steps of the model were as follows:

First step. Based on the literature review, we produced nine preliminary models consisting of different subsets of features. The features had individually or collectively worked as good predictors of septic shock in previous studies. The preliminary nine models are listed below:

1. Temperature, HR, RR, MAP, and SI (Berger et al., 2013)
2. Temperature, RR, MAP, Lactate, and WBC (Gultepe et al., 2014)
3. Temperature, RR, Creatinine, Lactate, and WBC (Carrara et al., 2015).
4. Temperature, HR, Creatinine, Lactate, and WBC (Carrara et al., 2015)
5. HR, RR, MAP, SBP, and DBP (Carrara et al., 2015).
6. Temperature, HR, RR, SBP, DBP, SpO_2, and GCS (Desautels et al., 2016)
7. DBP and Albumin (Holder et al., 2016)
8. MSI (Jayaprakash et al., 2018)
9. ageSI, Age, SBP, and Gender (Torabi et al., 2016)

Second step. Used the Cox proportional hazards model to obtain the coefficients (Li, Zhou, Choubey, & Sievenpiper, 2007), based on the nine sets listed in the first step. Each single run was performed on the whole training set for all patients from time of

admission to the time of the event or the censor time (discharged or died without getting the event).

The Cox proportional hazards model is a statistical technique for survival analysis of data (Walters, 2009). Survival models predict hazard at time t as a function of the input variables. In addition, the model allows separating the effects of treatment from other triggering features (Walters, 2009). The Cox proportional hazards (CPH) model is used extensively for survival analysis for censored data (Bonato et al., 2011; Hothorn et al., 2006; Tsujitani et al., 2012). It is one of the most widespread models in statistical analysis (Bonato et al., 2011; Wang et al., 2014). The CPH model is used broadly in clinical studies for risk ratio estimation (Lin et al., 2013). This method had helped researchers achieve good results in medical predictions and risk estimations (Guilloux et al., 2016; Jackson & Cox, 2014; Lee et al., 2016; Tolosie & Sharma, 2014; Wang et al., 2014; Wang et al., 2015; Wu et al., 2016).

The model is specified as follows:

$$h(t) = h_0(t) \times e^{\sum_{i=1}^{n} \beta_i \times X_i}$$

The quantity $h_0(t)$ is the baseline or underlying hazard function and corresponds to the probability of triggering the event, the septic shock, when all the explanatory features are zero (Walters, 2009). X_i represents the i^{th} predictor in the features' set. The regression coefficients β_i give the proportional change in the hazard, related to changes in the explanatory features. β is assessed with the maximum likelihood estimate (MLE) method, which is the value that makes the feature the most probable. Using the Survival Library in R, the *coxph* function was used to determine the Cox model including the coefficients (Therneau, 2018).

Third step. In this step, we obtained the Cox Risk Score at time *t*. The score is derived from the Cox Proportional Hazard Ratio shown below:

$$HR = \frac{h(t)}{h_0(t)} = e^{\sum_{i=1}^{n} \beta_i \times X_i}$$

Fox and Weisberg (2011); Kim et al. (2013); Staley et al. (2017) took the natural logarithm (ln) of each side of the Cox proportional hazards regression model, to relate the log of the relative hazard to a linear function of the predictors, thus producing the new score that looked as follows:

$$Score\ (t) = ln\left(\frac{h(t)}{h_0(t)}\right) = ln\left(e^{\sum_{i=1}^{n} \beta_i \times X_i}\right)$$

$$\Rightarrow Score(t) = \sum_{i=1}^{n} \beta_i \times X_{it}$$

where n = number of features, β_i= the coefficient of the i^{th} feature, and X_{it} = Value of the i^{th} feature at time t (Kim et al., 2013).

To get the risk score for each of the nine preliminary models, we filtered the training set to the values of the features at time *t*. In our case, we chose *t* to be equal to 20 hours before septic shock based on the models selected for comparison. Henry et al. (2015) predicted shock with a median of 28.2 [interquartile range (IQR), 10.6 to 94.2] hours before onset, and Mao et al. (2018) at four hours before the onset of septic shock; hence to improve accuracy, sensitivity, and specificity over Henry et al. (2015) and achieve metrics close to Mao et al. (2018), we chose 20 hours that was above the average of both predictions and the rounded average of the lower range and the median of Henry et al. (2015), that is, (28.2 + 10.6)/2 = 19.4, which was rounded up to 20. Since Time *t* was determined to be 20 hours before onset, to get the record that has the values at time *t*,

we subtracted the stop time (tStop) of the hourly bin from the full time with the result equaling 20 (FullTime – tStop = *t*). The values of the features from that record were used to calculate the score using the formula above. The score was added to the list as an enhancing feature. From Table 4, the records for the patient were reduced to one record as displayed in Table 6.

Table 6

Time Dependent Data Set Sample at Time t=20

subject_id	hadm_id	Temp	RR	MAP	Lactate	WBC	FullTime	tStart	tStop	Shock
250	124271	36.44	38	94	1.2	17.6	23	2	3	1

The event class at the full time was used at Time *t*. In this example, Shock was equal to one at tStop = FullTime = 23, therefore, upon reduction the Shock was set to one. Generally, the hourly bin at Time *t* varied for each patient as shown in Table 7.

Table 7

Cox Score Calculation Sample

subject_id	hadm_id	Temp	RR	MAP	Lactate	WBC	Score	FullTime	tStart	tStop	Shock
21	111970	37.28	14	67	2.7	38.6	0.345756	40	19	20	1
124	134369	35.89	17	76	1	8	-1.03133	379	358	359	0
157	110545	36.67	16	86	1.4	5.8	-1.27883	222	201	202	0
191	142081	37.56	16	126	1.3	13.7	-2.14925	266	245	246	0
211	101148	37	19	82	1.1	7.6	-1.14437	300	279	280	0
250	124271	36.44	38	94	1.2	17.6	-0.93867	23	2	3	1
275	129886	37.44	16	93	6.2	8.5	-0.39273	35	14	15	1
305	122211	37.39	19	67	1	7	-0.77668	638	617	618	0
323	143334	36.56	25	68	1.63	8.3	-0.56906	144	123	124	0
357	145674	37.11	12	67	2.4	11.4	-0.46387	107	86	87	1
530	149648	37.94	30	65	1.5	13.6	-0.29705	479	458	459	0
618	181546	37.28	19	78	1.7	11.9	-0.79677	304	283	284	0
638	149359	35.83	23	74	1.4	8.7	-0.79945	95	74	75	1
690	135389	36.06	19	73	0.9	11.5	-0.85171	547	526	527	0
801	187764	36.94	21	135	1	8	-2.54384	1074	1053	1054	0
894	157870	38.67	30	76	2.2	16.5	-0.3617	194	173	174	1

905	150569	37.44	32	116	2.6	9.1	-1.52877	264	243	244	1
914	124723	37.11	25	69	0.4	4	-0.95873	363	342	343	0
1006	147743	36.22	33	77	1.1	10.8	-0.75257	563	542	543	0
1006	189081	36.67	28	68	0.7	6	-0.78251	679	658	659	0
1006	199286	38.22	28	77	2.2	21.1	-0.29705	235	214	215	1
1141	153413	36.61	25	75	2.8	15.9	-0.31201	46	25	26	1
1331	114467	35.89	36	65	1.7	24.5	0.090139	26	5	6	1
1332	161256	36.83	28	109	1	8.9	-1.73641	313	292	293	0
1332	165244	35.89	19	104	1.3	9.8	-1.64303	404	383	384	0
1386	150628	34.83	13	52	1.6	5.9	-0.38449	894	873	874	0

Fourth step. In this step, we trained the Random Forest Ensemble classifier on the training sets of each Cox enhanced data sets. We used the Random Forest Library in R for the purpose of training model (Liaw, 2018). From the literature review, ensemble classifiers have not been used extensively to predict septic shock. However, other areas in the medical domain have benefited from the use of ensemble classifiers to predict progression of diseases and traumatic health situations (Kourou et al., 2015; Srimani & Koti, 2013). Lavanya and Rani (2012) used an ensemble classifier to improve the accuracy of breast cancer prediction. Kelarev et al. (2012) utilized ensemble classification to outperform all base classifiers in predicting cardiac autonomic neuropathy (CAN). Williams et al. (2013) suggested exploring ensemble classifiers to improve predictions. Ali et al. (2014) built multiple ensemble classifiers that performed very well in their experiments. To predict cancer survivors, Gupta et al. (2014) determined that the use of the ensemble classifiers can boost prediction over conventional methods. Yao et al. (2015) proposed an ensemble classification tool to predict protein-protein interaction (PPI) networks. Morino et al. (2015) generated accurate predictions for prostate cancer patients.

The proposed Cox Enhanced ensemble method, CERF, classified patients into two classes: A Septic Shock class - patients who were predicted to go into septic shock

20 hours before onset, and a No Septic Shock class - patients who most likely did not progress into septic shock.

Fifth step. For performance measurements, validation purposes, and parameters tuning, the *k-fold* cross validation technique was used with the number of folds *k = 10* (Beleites et al., 2013). Performance and parameters of the model could be affected by systematic deviations (bias) and random uncertainty (variance), therefore, the cross-validation process provided a mechanism to reduce both the bias and variance (Beleites et al., 2013), and to avoid over-fitting the training data (Refaeilzadeh et al., 2009). The method included:

1) Arrange the training set in random order.
2) Divide the training set into *k* folds or subsets (each fold size = *n/k*; *n*=number of records in the training set).
3) For i = 1 to k
    a) Train each individual model of the ensemble on all subsets except fold *i*.
    b) Test the ensemble classifier using fold *i*. Each individual Cox model is tested.
    c) Compute Accuracy (i), Sensitivity (i), Specificity (i), and Error Rate (i).
4) Compute Accuracy, Sensitivity, Specificity, and Error Rate. These are the averages for all iterations.

Sixth step. In this step, we tested the model on the test data set at 20 hours before the event full time and recorded the results. The prediction function in the Random Forest Library calculated these metrics: Accuracy, Sensitivity, Specificity, and Error Rate. The confusion matrix was also created based on the predictions.

Summary

The prediction model classified 4,683 patients diagnosed with sepsis or severe sepsis. These are patients with ICD-9 codes 995.91 and 995.92. The cases that developed into septic shock (ICD-9 code 785.52) were 2585 cases, out of which 1624 septic shock patients died. The mortality rate was high at 62.82%.

The effort was to reduce the rate of mortality by using an ensemble classification model to predict the patients who would progress into septic shock before its onset. The CERF prediction model considered 17 features that were divided into nine different combinations. The performance of the nine models was measured using the following:

- The Confusion Matrix: it is shown in Table 4, and it reports how the model classifies the various fault groups in comparison to the actual classification, and it consists of TP, FP, TN, and FN (Bowes, Hall, & Gray, 2012).

  Table 8

  Confusion Matrix

	Predicted False	Predicted True
Actual False	TN	FP
Actual True	FN	TP

- Sensitivity (SN): it is the measure of correctly classified positive cases (Parikh et al., 2008; Steyerberg, Calster, & Pencina, 2011; Steyerberg et al., 2010).
- Specificity (SP): it is the measure of correctly classified negative cases (Parikh et al., 2008; Steyerberg et al., 2011; Steyerberg et al., 2010).
- Accuracy: it is the Correct Classification Rate (CCR) (Bowes et al., 2012).

  Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

- Error Rate = $\frac{FP+FN}{TP+TN+FP+FN}$

The performance measures of the prediction model in this dissertation were compared to two prominent models. The first one was a routine screening protocol for septic shock that used SIRS criteria, suspicion of infection, and the presence of either hypotension or hyperlactatemia. This model achieved a specificity of 0.64 (FPR, 0.36) and a sensitivity of 0.74 (Henry et al., 2015). The second comparison was against the TREWScore model – a leading machine-learning model, which achieved a much higher sensitivity of 0.85 and specificity of 0.67 (Henry et al., 2015).

Chapter 4

Results

Overview

The goal of this dissertation was to develop a prediction model to classify and predict septic shock 20 hours before its onset using CERF, a Cox Enhanced Random Forest ensemble method. This chapter presents the results of the models, the validation of the final model, and the comparison against the routine screening protocol for septic shock and against the TREWScore model (Henry et al., 2015).

Model Results

We ran nine prediction models using a two-step method. The first step was to obtain the Cox Model coefficients for each separately, calculate the risk scores using the equation provided in the methodology section at time *t=20*, and add the score of each model to the features of that model. The second step was to apply the Random Forest ensemble on the enhanced feature sets of each model separately. The dataset for each model was reduced when the features were combined. One of the drawbacks of the MIMNIC III database was that features were not recorded for all patients all the time. Patients had records of one feature but lacked other features. This caused the size of the datasets to shrink. In Table 9, we summarized the total numbers of patients with multiple admissions for each model and supplied a breakdown of the numbers in each class for all models.

Table 9

Summary of total count of patients, and counts in each class for all models

Model		Total	0	1
Temperature, HR, RR, MAP, and SI	Training	723	546	177
	Test	177	139	38
Temperature, RR, MAP, Lactate, and WBC	Training	578	404	174
	Test	142	106	36
Temperature, RR, Creatinine, Lactate, and WBC	Training	1412	1238	174
	Test	337	301	36
Temperature, HR, Creatinine, Lactate, and WBC	Training	1412	1238	174
	Test	337	301	36
HR, RR, MAP, SBP, and DBP	Training	759	576	183
	Test	188	148	40
Temperature, HR, RR, SBP, DBP, SpO_2, and GCS	Training	722	545	177
	Test	177	139	38
DBP and Albumin	Training	450	279	171
	Test	117	81	36
MSI	Training	759	576	183
	Test	188	148	40
ageSI, Age, SBP, and Gender	Training	769	586	183
	Test	191	151	40

1. Temperature, HR, RR, MAP, and SI Model

First step - used the Cox Model and got the coefficients shown in Table 10:

Table 10

Temperature, HR, RR, MAP, and SI Model Coefficients

	Temperature	HR	RR	MAP	SI
Coefficients	-0.046626	0.010944	0.018253	-0.020364	0.306089

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 11 and the other metrics shown in Table 12:

Table 11

Temperature, HR, RR, MAP, and SI Model Confusion Matrix

	0	1
0	131	14
1	8	24

Table 12

Metrics for Temperature, HR, RR, MAP, and SI Model

95% CI	Accuracy	Sensitivity	Specificity
(0.8179, 0.9204)	0.8757	0.6316	0.9424

The model showed good results as shown in Table 12. The accuracy and specificity were very good but the model did not deliver high on sensitivity. This meant that the model did not pick up enough true positives. Even with the presence of MAP, one of the features that determined septic shock, the positive detection rate did not improve.

2. Temperature, RR, MAP, Lactate, and WBC Model

First step - used the Cox Model and got the coefficients shown in Table 13:

Table 13

Temperature, RR, MAP, Lactate, and WBC Model Coefficients

	Temperature	RR	MAP	Lactate	WBC
Coefficients	0.035297	0.001920	-0.016974	0.142865	0.021088

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 14 and the other metrics shown in Table 15:

Table 14

Temperature, RR, MAP, Lactate, and WBC Model Confusion Matrix

	0	1
0	103	4
1	3	32

Table 15

Metrics for Temperature, RR, MAP, Lactate, and WBC Model

95% CI	Accuracy	Sensitivity	Specificity
(0.9011, 0.98)	0.9507	0.8889	0.9717

Based on the results shown in Table 15, this was the highest performing model, with excellent accuracy and specificity, but very good sensitivity. This model did well detecting the positive values. The presence of both MAP and Lactate, the two deterministic features of septic shock onset, had a very high impact on the higher detection.

3. Temperature, RR, Creatinine, Lactate, and WBC Model

First step - used the Cox Model and got the coefficients shown in Table 16:

Table 16

Temperature, RR, Creatinine, Lactate, and WBC Model Coefficients

	Temperature	RR	Creatinine	Lactate	WBC
Coefficients	0.063327	0.001915	0.018966	0.171590	0.034384

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 17 and the other metrics shown in Table 18:

Table 17

Temperature, RR, Creatinine, Lactate, and WBC Model Confusion Matrix

	0	1
0	297	21
1	4	15

Table 18

Metrics for Temperature, RR, Creatinine, Lactate, and WBC Model

95% CI	Accuracy	Sensitivity	Specificity
(0.8924, 0.9514)	0.9258	0.41667	0.98671

The model showed good results as shown in Table 18. The accuracy and specificity were very good but sensitivity was very low. This model did not pick up enough true positives. The presence of Lactate, which is one of the determining features of septic shock timing, did not improve the positive detection rate. It seemed the other features had a clear negative impact on sensitivity, which was very obvious in the results.

4. Temperature, HR, Creatinine, Lactate, and WBC Model

First step - used the Cox Model and got the coefficients shown in Table 19:

Table 19

Temperature, HR, Creatinine, Lactate, and WBC Model Coefficients

	Temperature	HR	Creatinine	Lactate	WBC
Coefficients	-0.002077	0.013166	0.030142	0.168251	0.032635

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 20 and the other metrics shown in Table 21:

Table 20

Temperature, HR, Creatinine, Lactate, and WBC Model Confusion Matrix

	0	1
0	294	18
1	7	18

Table 21

Metrics for Temperature, HR, Creatinine, Lactate, and WBC Model

95% CI	Accuracy	Sensitivity	Specificity
(0.8924, 0.9514)	0.9258	0.5000	0.97674

The model's accuracy and specificity were very high, but the sensitivity was low as shown in Table 21. Many positive values were wrongly classified as negative, which explained the high specificity. Similar to the previous model, the presence of Lactate did not improve the positive detection rate.

5. HR, RR, MAP, SBP, and DBP Model

First step - used the Cox Model and got the coefficients shown in Table 22:

Table 22

HR, RR, MAP, SBP, and DBP Model Coefficients

	HR	RR	MAP	SBP	DBP
Coefficients	0.0130021	0.0164549	-0.0417143	0.0003254	0.0249370

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 23 and the other metrics shown in Table 24:

Table 23

HR, RR, MAP, SBP, and DBP Model Confusion Matrix

	0	1
0	138	20
1	10	20

Table 24

Metrics for HR, RR, MAP, SBP, and DBP Model

95% CI	Accuracy	Sensitivity	Specificity
(0.7801, 0.8897)	0.8404	0.5000	0.9324

The model delivered less than the previous one as illustrated Table 24. The lower sensitivity showed the model's inability to detect positive values at a higher rate. Like one of the previous models, the presence of MAP did not improve the detection rate.

6. Temperature, HR, RR, SBP, DBP, SpO_2, and GCS Model

First step - used the Cox Model and got the coefficients shown in Table 25:

Table 25

Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model Coefficients

	Temp	HR	RR	SBP	DBP	SpO_2	GCS
Coefficients	-0.026826	0.012868	0.029240	-0.009402	-0.008089	-0.046293	-0.170831

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 26 and the other metrics shown in Table 27:

Table 26

Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model Confusion Matrix

	0	1
0	134	13
1	5	25

Table 27

Metrics for Temperature, HR, RR, SBP, DBP, SpO2, and GCS Model

95% CI	Accuracy	Sensitivity	Specificity
(0.844, 0.9386)	0.8983	0.6579	0.9640

The model results displayed in Table 27 showed very good accuracy and specificity, but average sensitivity. The combination of these features did not deliver as discussed in prior research efforts.

7. DBP and Albumin Model

First step - used the Cox Model and got the coefficients shown in Table 28:

Table 28

DBP and Albumin Model Coefficients

	DBP	Albumin
Coefficients	-0.011217	-0.554037

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 29 and the other metrics shown in Table 30:

Table 29

DBP and Albumin Model Confusion Matrix

	0	1
0	64	15
1	17	21

Table 30

Metrics for DBP and Albumin Model

95% CI	Accuracy	Sensitivity	Specificity
(0.6364, 0.8048)	0.7265	0.5833	0.7901

Table 30 showed very low results, which did not put this model at a useful level. The two features did not work well together and therefore the model was deemed useless and unproductive.

8. MSI Model

First step - used the Cox Model and got the coefficients shown in Table 31:

Table 31

MSI Model Coefficients

	MSI
Coefficients	1.2034

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 32 and the other metrics shown in Table 33:

Table 32

MSI Model Confusion Matrix

	0	1
0	141	20
1	7	20

Table 33

Metrics for MSI Model

95% CI	Accuracy	Sensitivity	Specificity
(0.798, 0.9032)	0.8564	0.5000	0.9527

Table 33 showed good accuracy and excellent specificity, but bad sensitivity. The number of true positives was very low, which put the model at a useless level.

9. ageSI, Age, SBP, and Gender Model

First step - used the Cox Model and got the coefficients shown in Table 34:

Table 34

ageSI, Age, SBP, and Gender Model Coefficients

	ageSI	Age	SBP	Gender	DBP
Coefficients	0.015063	-0.008210	-0.006527	0.145446	0.0249370

Second step – applied the Random Forest Ensemble and got the confusion matrix shown in Table 35 and the other metrics shown in Table 36:

Table 35

ageSI, Age, SBP, and Gender Model Confusion Matrix

	0	1
0	143	21
1	8	19

Table 36

Metrics for ageSI, Age, SBP, and Gender Model

95% CI	Accuracy	Sensitivity	Specificity
(0.7893, 0.8959)	0.8482	0.4750	0.94702

The new calculated feature, ageSI, did not add value to the model, as Table 36 illustrated. The sensitivity was very low. As a result, the model was ruled out.

Selected Model

The second model delivered the best results among the tested models. The presence of both Lactate and MAP helped the model achieve better overall percentages. As a matter of fact, the Cox Hazard Model showed that Lactate, MAP, and WBC were the most significant features that affected the time the event happened, as Figure 2 illustrated.

```
Call:
coxph(formula = Surv(tStart, tStop, Shock) ~ Temperature + RR +
    MAP + Lactate + WBC, data = Temp_RR_MAP_Lactate_WBC_Classifier)

  n= 198693, number of events= 356

                coef exp(coef)  se(coef)      z Pr(>|z|)
Temperature  0.035297  1.035927  0.054042  0.653    0.514
RR           0.001920  1.001922  0.008419  0.228    0.820
MAP         -0.016974  0.983169  0.004092 -4.148 3.36e-05 ***
Lactate      0.142865  1.153574  0.012528 11.403  < 2e-16 ***
WBC          0.021088  1.021312  0.004624  4.561 5.10e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2 - Cox Model for Temperature, RR, MAP, lactate, and WBC

The Cox risk score was calculated for all patients twenty hours before the start of the shock, where we fed the measurements of the features at time *t=20* and the coefficients produced by the Cox model into the Cox Risk Score equation defined in the Method. This score was added to the features as a new and additional calculated feature. The Random Forest ensemble classifier was used to get the final classification, thus coming up with a new method called CERF – the Cox Enhanced Random Forest Prediction Model. The model demonstrated impressive results. The Accuracy was 0.9507 with 95% CI: (0.9011, 0.98), Sensitivity was 0.8889, Specificity was 0.9717, and Error Rate was 6.23%. Figure 3 shows the full measures that were obtained from fitting the Random Forest ensemble on the enhanced feature list.

```
Confusion Matrix and Statistics Reference
Prediction        0     1
         0        103   4
         1        3     32

               Accuracy: 0.9507
                 95% CI: (0.9011, 0.98)
            Sensitivity: 0.8889
            Specificity: 0.9717
         Pos Pred Value: 0.9143
         Neg Pred Value: 0.9626
             Error Rate: 6.23%
```

Figure 3 - Temperature, RR, MAP, lactate, and WBC Model Results

Model Validation

Random Forest had self-test and built-in validation (Strobl, Malley, & Tutz, 2009); nevertheless, for extra validation, we used the k-fold cross validation technique with the number of folds *k = 10* (Beleites et al., 2013). In addition, cross validation helped in performance measurements and parameters tuning (Beleites et al., 2013). It also provided a mechanism to reduce both the bias and variance (Beleites et al., 2013), and to avoid over-fitting the training data (Refaeilzadeh et al., 2009).

Before we tested the model on the test data set, the training set was arranged in random order, then divided into 10 folds or subsets. The model was trained on nine folds and tested on one, with the process repeated 10 times.

Table 37 demonstrated the 10 folds results and their average. The validation confirmed the results that the model reached. The accuracy did not go below 0.9, and at different runs, the model was able to reach a 100% accuracy, sensitivity, and specificity. This demonstrates the power and validity of CERF – the new Cox Enhanced Random Forest Ensemble.

Table 37

Cross Validation Results

	Accuracy	95% CI	Sensitivity	Specificity	Error Rate
Fold 1	0.9216	(0.8112, 0.9782)	0.8261	1	5.50%
Fold 2	0.8971	(0.7993, 0.9576)	0.6842	0.9796	5.49%
Fold 3	0.902	(0.7859, 0.9674)	0.8	0.9444	6.26%
Fold 4	0.9496	(0.8585, 0.9894)	0.8235	1	6.55%
Fold 5	1	(0.9351, 1)	1	1	7.46%
Fold 6	0.9583	(0.8575, 0.9949)	0.8889	1	6.98%
Fold 7	0.9434	(0.8434, 0.9882)	0.9091	0.9524	6.48%
Fold 8	0.9538	(0.871, 0.9904)	0.9091	0.9767	7.21%
Fold 9	0.9444	(0.8461, 0.9884)	1	0.925	6.11%
Fold 10	0.9189	(0.8318, 0.9697)	0.9	0.9259	5.95%
Average	0.93891		0.8741	0.9704	6.40%

Model Comparison

CERF – the Cox Enhanced Random Forest Prediction Model, has delivered remarkable results with Accuracy at 0.9507 (95% CI: [0.9011, 0.98]), Sensitivity at 0.8889, Specificity at 0.9717, and Error Rate at 6.23%. See Figure 3 for the full measures.

The performance measures of CERF are compared to two prominent models. The first model is the routine screening protocol for septic shock that used SIRS criteria, suspicion of infection, and the presence of either hypotension or hyperlactatemia. The model achieved a specificity of 0.64 (FPR, 0.36) and a sensitivity of 0.74 (Henry et al., 2015). The second comparison was against the TREWScore model – a leading machine-learning model, with an accuracy of 0.83 [95% confidence interval (CI), 0.81 to 0.85] at a specificity of 0.67 and a sensitivity of 0.85 within a median of 28.2 [interquartile range (IQR), 10.6 to 94.2] hours before onset (Henry et al., 2015). CERF has an Accuracy of

0.9507 (95% CI: [0.9011, 0.98]), Sensitivity of 0.8889, and Specificity of 0.9717 with 20 hours before the onset of the shock. CERF solidly exceeded both models.

Additionally, we will extend the comparison to another recent model. The model, called InSight, classified patients into Shock and No Shock with an accuracy of 0.96 (95% CI: [0.94, 0.98]), Sensitivity of 0.80, and Specificity of 0.95 four hours before the onset of septic shock (Mao et al., 2018). CERF achieved a very close accuracy (less than 1% difference) but with an extra sixteen hours of lead time and better sensitivity.

Summary

This chapter presented nine prediction models using a two-step method. The first step was to obtain the Cox Model coefficients and add it to the feature set, and the second step was to use the Random Forest ensemble on the enhanced set. The most prominent model is picked to introduce CERF – the Cox Enhanced Random Forest Prediction Model. The model is validated using a k-fold cross validation technique with $k = 10$. The validation strengthened the superior result achieved by the model.

The model was then compared to three different models with one of them very recent and CERF showed superiority over the compared models. CERF predicted the onset of septic shock 20 hours before it happened with an Accuracy of 0.9507 (95% CI: [0.9011, 0.98]), Sensitivity of 0.8889, and Specificity of 0.9717, and at one instance delivering 100% in all measures during validation.

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Overview

This dissertation presented CERF - an enhanced method to classify and predict septic shock more accurately than previous methods presented by other researchers. The method combined the strengths of the Cox Hazard Model with the effectiveness of the Random Forest ensemble. This chapter draws the conclusions of this dissertation and the implications towards the current standing of septic shock prediction in particular, and towards medical prediction in general. It then discusses recommendations for future work and ends with a summary of the chapter.

Conclusions

The focus of this study was to answer the following question:

RQ

How can one develop an ensemble model to predict septic shock with acceptable accuracy?

A single classifier searches for the best approximation or hypothesis to the unknown function (Dietterich, 2002). The algorithm measures how well a hypothesis matches the function to determine the best one using data points in the training set (Dietterich, 2002). In contrast, an ensemble classifier constructs a set of hypotheses then combine them using a combining method (Dietterich, 2000, 2002; Valentini & Masulli,

2002). The result improves the overall performance and delivers a more accurate classification (Dietterich, 2000, 2002; Valentini & Masulli, 2002).

Ensemble classifiers work better because they reduce the effect of the three problems that affect single classifiers' performance (Dietterich, 2000, 2002). The first problem is statistical and caused by an insufficient training dataset, which may result in finding multiple optimal hypotheses (Dietterich, 2000, 2002; Valentini & Masulli, 2002). If the algorithm chooses the wrong hypothesis, it will lead to incorrect predictions (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The problem can be resolved by combining the results and getting a better approximation (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The second one is the computational problem, which occurs when the classification algorithm applies local optimization techniques that can get stuck in local minima (optima), therefore, the algorithm cannot find the best hypothesis (Dietterich, 2000, 2002; Valentini & Masulli, 2002). If a weighted combination of the several different local minima is applied, the problem can be reduced or eliminated (Dietterich, 2000, 2002; Valentini & Masulli, 2002). The third problem is representational that occurs when the space of hypotheses does not contain any good approximation to the unknown function (Dietterich, 2000, 2002; Valentini & Masulli, 2002). In this case, an ensemble classifier can help expand the space and allow a more accurate approximation (Dietterich, 2000, 2002; Valentini & Masulli, 2002).

The use of ensemble classification is not very well established in septic shock prediction. However, it has benefited other medical domains (Kourou et al., 2015; Srimani & Koti, 2013). Lavanya and Rani (2012) presented an ensemble classifier based on a hybrid of decision trees with the bagging technique to improve the accuracy of

breast cancer prediction. Kelarev et al. (2012) used the Random Forest ensemble to build a model that outperformed all base classifiers in predicting cardiac autonomic neuropathy (CAN). Ali et al. (2014) built multiple ensemble classifiers using Random Forest (RF), SVM, and KNN that performed very well in their experiments. To predict cancer survivors, Gupta et al. (2014) built three ensemble classifiers, which boosted prediction over conventional methods. Yao et al. (2015) proposed Random Forests to enhance the prediction of protein-protein interaction (PPI) networks. Morino et al. (2015) generated accurate predictions with ensemble classification when tested on a dataset for prostate cancer patients.

CERF delivered a superior result compared to existing models. The method combined the power of the Cox Hazard model, which calculates the hazard that features can have on the status of the outcome. Thus, calculating a score that can help the Random Forest Ensemble classify more accurately. Table 38 summaries the comparison of the models and reveals the superiority of CERF.

Table 38

Model Comparisons

Model	Accuracy	Sensitivity	Specificity	Hours Before Onset
Routine Screening Protocol	--	0.74	0.64	--
TREWScore	0.83	0.85	0.67	Median:28.2 [interquartile range (IQR), 10.6 to 94.2]
InSight	0.96	0.8	0.95	4
CERF	0.9507	0.8889	0.9717	20

Implications

The process of improving prediction relies heavily on data preparation, the choice of algorithms, and the enhancement to the existing algorithm. Predicting the outcome of

disease before it happens, plays a very important role in deciding the treatment that could save the patients' lives. As the problem statement elaborated alongside the extensive literature review, identifying septic shock in a timely manner before it happens is crucial in reducing the mortality rate. The methodology in this dissertation has delivered a tool called CERF that improved the predictability of septic shock with higher accuracy, sensitivity, and specificity. CERF can utilize a limited number of inputs from any EMR and deliver a prediction.

Besides, another novel feature of CERF is its portability. The method can be packaged to work with any EMR system to deliver continuous predictions as vitals, laboratory tests, and clinical observations are being recorded.

In summary, the effort presented in this dissertation advanced the current state of the septic shock prediction problem. The tool is able to generate better predications, thus allowing better knowledge of patients' statuses, and helping medical professionals decide on treatments.

Recommendations

Based on the results of this work, there are many recommendations that could improve the effort of this dissertation. As discussed earlier, this study aimed at improving the prediction of septic shock. One recommendation is varying the input as it can change the output. Using a different dataset that has been processed and cleaned differently can have a two-fold impact: validate the current results and improve prediction through fine tuning the input of the features.

A second recommendation is to utilize the other prediction models and combine the result through a voting mechanism or any ensemble combining techniques. A third

recommendation is test the ability of CERF to be used in a different medical prediction problem.

The fourth and final recommendation is to add an unsupervised machine learning technique to continuously enhance the tool based on the previous prediction accuracies. This is an ambitious recommendation, which if implemented successfully, can lead to continuous improvements of medical predictions. The results of such implementation can have a very good impact on saving patients' lives.

Summary

This dissertation improved the prediction of septic shock by using machine learning techniques. The study used data from the MIMIC-III database v1.3, which has records of 46,520 ICU patients with 58,976 admissions collected at Beth Israel Deaconess Medical Center between 2001 – 2012 (Goldberger et al., 2000; "MIMIC-III Clinical Database," 2015). From the available vital, laboratory, and clinical measurements, and based on prior research, we used temperature, RR, MAP, Lactate, and WBC as the input or features for the method (Gultepe et al., 2014). The data was then cleaned up, where outliers and values with wrong data types were eliminated by nulling them out and treating them as missing values (Ho et al., 2012; Ho et al., 2014). For any measurement with multiple values, the mean value was used, and for missing values a carry forward or backward approach was applied depending on the location of the missing item (Desautels et al., 2016; Mao et al., 2018). The data for each feature was then reconstructed and transformed into time-dependent sets (Therneau et al., 2018). The data was then divided into a training set and a test set.

The two-step prediction model CERF - the Cox Enhanced Random Forest Prediction Model, was introduced. The first step was to obtain the Cox model coefficients, calculate a score at a time corresponding to 20 hours before the onset of septic shock, and add the new feature to the feature set. The model was validated using a k-fold cross validation technique with $k = 10$. The validation strengthened the superior result achieved by the model. The second step was to use the Random Forest ensemble on the enhanced set. The model was then compared to three different models: The Routine Screening Protocol, TREWScore, and InSight. CERF predicted the onset of septic shock 20 hours before it happened with an Accuracy of 0.9507 (95% CI: [0.9011, 0.98]), Sensitivity of 0.8889, and Specificity of 0.9717, beating all three models.

In conclusion, CERF delivered results superior to the previous prominent models. This research effort advanced the current status of septic shock prediction by improving the prediction accuracy, thus adding a contribution to the general body of knowledge.

References

Abad, S. K. J., Zare-Mirakabad, M. R., & Rezaeian, M. (2014). *An approach for classifying large dataset using ensemble classifiers.* Paper presented at the Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on.

Abhyankar, S., Demner-Fushman, D., & McDonald, C. J. (2012). Standardizing clinical laboratory data for secondary use. *Journal of biomedical informatics, 45*(4), 642-650. doi:10.1016/j.jbi.2012.04.012

Alberg, A. J., Park, J. W., Hager, B. W., Brock, M. V., & Diener-West, M. (2004). The Use of "Overall Accuracy" to Evaluate the Validity of Screening or Diagnostic Tests. *Journal of General Internal Medicine, 19*(5 Pt 1), 460-465. doi:10.1111/j.1525-1497.2004.30091.x

Ali, S., Majid, A., & Khan, A. (2014). IDM-PhyChm-Ens: intelligent decision-making ensemble methodology for classification of human breast cancer using physicochemical properties of amino acids. *Amino acids, 46*(4), 977-993. doi:10.1007/s00726-013-1659-x

Angus, D. C., & van der Poll, T. (2013). Severe Sepsis and Septic Shock. *New England Journal of Medicine, 369*(9), 840-851. doi:doi:10.1056/NEJMra1208623

Azevedo, J. R. A. d., Torres, O. J. M., Czeczko, N. G., Tuon, F. F., Nassif, P. A. N., & Souza, G. D. d. (2012). Procalcitonin as a prognostic biomarker of severe sepsis and septic shock. *Revista do Colégio Brasileiro de Cirurgiões, 39*(6), 456-461. doi:10.1590/S0100-69912012000600003

Bagheri, M. A., & Gao, Q. (2012). *An efficient ensemble classification method based on novel classifier selection technique*. Paper presented at the Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, Craiova, Romania.

Bauer, E., & Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning, 36*(1-2), 105-139. doi:10.1023/a:1007515423169

Becker, K. L., Snider, R., & Nylen, E. S. (2010). Procalcitonin in sepsis and systemic inflammation: a harmful biomarker and a therapeutic target. *British Journal of Pharmacology, 159*(2), 253-264. doi:10.1111/j.1476-5381.2009.00433.x

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta, 760*, 25-33. doi:10.1016/j.aca.2012.11.007

Berger, T., Green, J., Horeczko, T., Hagar, Y., Garg, N., Suarez, A., . . . Shapiro, N. (2013). Shock Index and Early Recognition of Sepsis in the Emergency Department: Pilot Study. *Western Journal of Emergency Medicine, 14*(2), 168-174. doi:10.5811/westjem.2012.8.11546

Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. (2001). *Clinical Pharmacology & Therapeutics, 69*(3), 89-95. doi:10.1067/mcp.2001.113989

Blomkalns, A. L. (2006). Lactate-a marker for sepsis and trauma. *Emergency Medicine Cardiac Research and Education Group, 2*, 43-49.

Bonato, V., Baladandayuthapani, V., Broom, B. M., Sulman, E. P., Aldape, K. D., & Do, K.-A. (2011). Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics, 27*(3), 359-367. doi:10.1093/bioinformatics/btq660

Bowes, D., Hall, T., & Gray, D. (2012). *Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix*. Paper presented at the Proceedings of the 8th International Conference on Predictive Models in Software Engineering, Lund, Sweden.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

Carrara, M., Baselli, G., & Ferrario, M. (2015). Mortality Prediction Model of Septic Shock Patients Based on Routinely Recorded Data. *Computational and Mathematical Methods in Medicine, 2015*, 761435. doi:10.1155/2015/761435

Chen, W. L., & Kuo, C. D. (2007). Characteristics of heart rate variability can predict impending septic shock in emergency department patients with sepsis. *Academic emergency medicine, 14*(5), 392-397. doi:10.1197/j.aem.2006.12.015

Chen, X., & Ishwaran, H. (2012). Random Forests for genomic data analysis. *Genomics, 99*(6), 323-329. doi:https://doi.org/10.1016/j.ygeno.2012.04.003

Cohen, J., & McConnell, J. S. (1988). Limulus assay in prediction of septic shock. *The Lancet, 331*(8595), 1165. doi:10.1016/S0140-6736(88)91977-0

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for classification in ecology. *Ecology, 88*(11), 2783-2792.

Dauwan, M., van der Zande, J. J., van Dellen, E., Sommer, I. E. C., Scheltens, P., Lemstra, A. W., & Stam, C. J. (2016). Random Forest to differentiate dementia with Lewy bodies from Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis,*

Assessment & Disease Monitoring, 4, 99-106. doi:https://doi.org/10.1016/j.dadm.2016.07.003

Deepak, C., & Bhat, S. (2014). Prediction of outcome in patients with sepsis using C-reactive protein & APACHE II scoring system. *IOSR Journal of Dental and Medical Sciences, 13*(3), 17-20.

Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., . . . Das, R. (2016). Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics, 4*(3), e28. doi:10.2196/medinform.5909

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning *Multiple Classifier Systems* (Vol. 1857, pp. 1-15): Springer Berlin Heidelberg.

Dietterich, T. G. (2002). Ensemble learning. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2 ed., pp. 405-408). Cambridge, MA: The MIT Press.

Fox, J., & Weisberg, S. (2011). Cox Proportional-Hazards Regression for Survival Data in R *An R Companion to Applied Regression*. Los Angeles: Sage Publishing.

Ghavidel, J., Yazdani, S., & Analoui, M. (2013, 28-30 May 2013). *A new ensemble classifier creation method by creating new training set for each base classifier.* Paper presented at the Information and Knowledge Technology (IKT), 2013 5th Conference on.

Gibot, S., Béné, M. C., Noel, R., Massin, F., Guy, J., Cravoisy, A., . . . Charles, P.-E. (2012). Combination Biomarkers to Diagnose Sepsis in the Critically Ill Patient. *American Journal of Respiratory and Critical Care Medicine, 186*(1), 65-71. doi:10.1164/rccm.201201-0037OC

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., . . . Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215-e220. doi:10.1161/01.cir.101.23.e215

Guilloux, A., Lemler, S., & Taupin, M.-L. (2016). Adaptive estimation of the baseline hazard function in the Cox model by model selection, with high-dimensional covariates. *Journal of Statistical Planning and Inference, 171*, 38-62. doi:http://dx.doi.org/10.1016/j.jspi.2015.11.005

Gultepe, E., Green, J. P., Nguyen, H., Adams, J., Albertson, T., & Tagkopoulos, I. (2014). From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *Journal of the American*

Medical Informatics Association : JAMIA, 21(2), 315-325. doi:10.1136/amiajnl-2013-001815

Gupta, S., Tran, T., Luo, W., Phung, D., Kennedy, R. L., Broad, A., . . . Khasraw, M. (2014). Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. *BMJ open, 4*(3), e004007. doi:10.1136/bmjopen-2013-004007

Hattori, N., Oda, S., Sadahiro, T., Nakamura, M., Abe, R., Shinozaki, K., . . . Hirasawa, H. (2009). YKL-40 IDENTIFIED BY PROTEOMIC ANALYSIS AS A BIOMARKER OF SEPSIS. *Shock, 32*(4), 393-400. doi:10.1097/SHK.0b013e31819e2c0c

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. doi:10.1126/scitranslmed.aab3719

Ho, J. C., Lee, C. H., & Ghosh, J. (2012). *Imputation-enhanced prediction of septic shock in ICU patients.* Paper presented at the Proceedings of the ACM SIGKDD Workshop on Health Informatics, Beijing, China.

Ho, J. C., Lee, C. H., & Ghosh, J. (2014). Septic Shock Prediction for Patients with Missing Data. *ACM Transactions on Management Information Systems, 5*(1), 1-15. doi:10.1145/2591676

Holder, A. L., Gupta, N., Lulaj, E., Furgiuele, M., Hidalgo, I., Jones, M. P., . . . Birnbaum, A. (2016). Predictors of early progression to severe sepsis or shock among emergency department patients with nonsevere sepsis. *International Journal of Emergency Medicine, 9*(1), 10. doi:10.1186/s12245-016-0106-7

Honda, T., & Karl Härdle, W. (2014). Variable selection in Cox regression models with varying coefficients. *Journal of Statistical Planning and Inference, 148*, 67-81. doi:http://dx.doi.org/10.1016/j.jspi.2013.12.002

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics, 7*(3), 355-373. doi:10.1093/biostatistics/kxj011

Jackson, R. J., & Cox, T. F. (2014). A Robust Parameterization for Unbounded Covariates Within the Cox Proportional Hazards Model. *International Journal of Statistics in Medical Research, 3*(4), 331.

Jayaprakash, N., Gajic, O., Frank, R. D., & Smischney, N. (2018). Elevated modified shock index in early sepsis is associated with myocardial dysfunction and mortality. *Journal of Critical Care, 43*, 30-35. doi:10.1016/j.jcrc.2017.08.019

Jia, J., Liu, Z., Xiao, X., Liu, B., & Chou, K.-C. (2016). pSuc-Lys: predict lysine succinylation sites in proteins with PseAAC and ensemble Random Forest approach. *Journal of theoretical biology, 394*, 223-230.

Johnson, A. E. W., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D., & Clifford, G. D. (2016). Machine Learning and Decision Support in Critical Care. *Proceedings of the IEEE, 104*(2), 444-466. doi:10.1109/JPROC.2015.2501978

Kelarev, A. V., Stranieri, A., Yearwood, J. L., Abawajy, J., & Jelinek, H. F. (2012). *Improving classifications for cardiac autonomic neuropathy using multi-level ensemble classifiers and feature selection based on Random Forest*. Paper presented at the Proceedings of the Tenth Australasian Data Mining Conference - Volume 134, Sydney, Australia.

Kelly, B. J., Lautenbach, E., Nachamkin, I., Coffin, S. E., Gerber, J. S., Fuchs, B. D., . . . Han, J. H. (2016). Combined biomarkers discriminate a low likelihood of bacterial infection among surgical intensive care unit patients with suspected sepsis. *Diagnostic Microbiology and Infectious Disease, 85*(1), 109-115. doi:http://dx.doi.org/10.1016/j.diagmicrobio.2016.01.003

Kibe, S., Adams, K., & Barlow, G. (2011). Diagnostic and prognostic biomarkers of sepsis in critical care. *Journal of Antimicrobial Chemotherapy, 66*(suppl 2), ii33-ii40. doi:10.1093/jac/dkq523

Kim, S., Park, T., & Kon, M. A. (2013). *Computational methods for cancer survival classification using intermediate information.* Paper presented at the IWBBIO.

King, E. G., Bauzá, G. J., Mella, J. R., & Remick, D. G. (2014). Pathophysiologic mechanisms in septic shock. *Laboratory investigation, 94*(1), 4-12.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, 13*, 8-17.

Kulkarni, A., & Lowe, B. (2016). Random Forest Algorithm for Land Cover Classification. *International Journal on Recent and Innovation Trends in Computing and Communication, 4*(3), 58-63.

Lausevic, Z., & Lausevic, M. (2012). *Early Detection of Sepsis, MOF and Outcome Prediction in Severely Traumatized Patients* (L. Azevedo Ed. 1 ed.). Rijeka, Croatia: InTech.

Lavanya, D., & Rani, K. U. (2012). Ensemble decision tree classifier for breast cancer data. *International Journal of Information Technology Convergence and Services (IJITCS), 2*(1), 17-24. doi:10.5121/ijitcs.2012.2103

Lebedev, A. V., Westman, E., Van Westen, G. J. P., Kramberger, M. G., Lundervold, A., Aarsland, D., . . . Simmons, A. (2014). Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical, 6*, 115-125. doi:https://doi.org/10.1016/j.nicl.2014.08.023

Lee, H., Hudgens, M. G., Cai, J., & Cole, S. R. (2016). Marginal Structural Cox Models with Case-Cohort Sampling. *Stat Sin, 26*(2), 509-526. doi:10.5705/ss.2014.015

Lee, S. M., & An, W. S. (2016). New clinical criteria for septic shock: serum lactate level as new emerging vital sign. *Journal of Thoracic Disease, 8*(7), 1388-1390. doi:10.21037/jtd.2016.05.55

Li, P., Stuart, E. A., & Allison, D. B. (2015). Multiple imputation: A flexible tool for handling missing data. *JAMA, 314*(18), 1966-1967. doi:10.1001/jama.2015.15281

Li, Z., Zhou, S., Choubey, S., & Sievenpiper, C. (2007). Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE transactions, 39*(3), 303-315. doi:10.1080/07408170600847168

Liaw, A. (2018). Package'randomForest'. Breiman and Cutlers Random Forests for classification and regression. Version 4.6-12.

Lin, I. F., Chang, W. P., & Liao, Y.-N. (2013). Shrinkage methods enhanced the accuracy of parameter estimation using Cox models with small number of events. *Journal of clinical epidemiology, 66*(7), 743-751. doi:http://dx.doi.org/10.1016/j.jclinepi.2013.02.002

Lin, W., Wu, Z., Lin, L., Wen, A., & Li, J. (2017). An Ensemble Random Forest Algorithm for Insurance Big Data Analysis. *IEEE Access, 5*, 16568-16575.

Lorente, L., Martín, M. M., Labarta, L., Díaz, C., Solé-Violán, J., Blanquer, J., . . . Páramo, J. A. (2009). Matrix metalloproteinase-9, -10, and tissue inhibitor of matrix metalloproteinases-1 blood levels as biomarkers of severity and mortality in sepsis. *Critical Care, 13*(5), 1-9. doi:10.1186/cc8115

Lukaszewski, R. A., Yates, A. M., Jackson, M. C., Swingler, K., Scherer, J. M., Simpson, A., . . . Brooks, T. J. (2008). Presymptomatic prediction of sepsis in intensive care unit patients. *Clinical and Vaccine Immunology, 15*(7), 1089-1094. doi:10.1128/CVI.00486-07

Malmir, J., Bolvardi, E., & Afzal Aghaee, M. (2014). Serum lactate is a useful predictor of death in severe sepsis and septic shock. *Reviews in Clinical Medicine, 1*(3), 97-104.

Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., . . . Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital

sign data in the emergency department, general ward and ICU. *BMJ open, 8*(1). doi:10.1136/bmjopen-2017-017833

Martin, G. S. (2012). Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. *Expert review of anti-infective therapy, 10*(6), 701-706. doi:10.1586/eri.12.50

Marty, P., Roquilly, A., Vallée, F., Luzi, A., Ferré, F., Fourcade, O., . . . Minville, V. (2013). Lactate clearance for death prediction in severe sepsis or septic shock patients during the first 24 hours in Intensive Care Unit: an observational study. *Annals of intensive care, 3*(1), 3. doi:10.1186/2110-5820-3-3

Matsusue, S., Kashihara, S., & Koizumi, S. (1988). Prediction of mortality from septic shock in gastrointestinal surgery by probit analysis. *The Japanese journal of surgery, 18*(1), 18-22. doi:10.1007/BF02470841

McLean, A. S., Tang, B., & Huang, S. J. (2015). Investigating sepsis with biomarkers. *BMJ, 350*. doi:10.1136/bmj.h254

Mikkelsen, M. E., Miltiades, A. N., Gaieski, D. F., Goyal, M., Fuchs, B. D., Shah, C. V., . . . Christie, J. D. (2009). Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. *Critical care medicine, 37*(5), 1670-1677.

MIMIC-III Clinical Database. (2015). *MIT Lab for Computational Physiology.*

Mohan, A., Shrestha, P., Guleria, R., Pandey, R. M., & Wig, N. (2015). Development of a mortality prediction formula due to sepsis/severe sepsis in a medical intensive care unit. *Lung India: official organ of Indian Chest Society, 32*(4), 313-319. doi:10.4103/0970-2113.159533

Morino, K., Hirata, Y., Tomioka, R., Kashima, H., Yamanishi, K., Hayashi, N., . . . Aihara, K. (2015). Predicting disease progression from short biomarker series using expert advice algorithm. *Scientific Reports, 5*(8953). doi:10.1038/srep08953

Nguyen, H. B., Loomba, M., Yang, J. J., Jacobsen, G., Shah, K., Otero, R. M., . . . Rivers, E. P. (2010). Early lactate clearance is associated with biomarkers of inflammation, coagulation, apoptosis, organ dysfunction and mortality in severe sepsis and septic shock. *Journal of Inflammation, 7*(1), 1-11. doi:10.1186/1476-9255-7-6

Nguyen, S. Q., Mwakalindile, E., Booth, J. S., Hogan, V., Morgan, J., Prickett, C. T., . . . Wang, H. E. (2014). Automated electronic medical record sepsis detection in the emergency department. *PeerJ, 2*, e343. doi:10.7717/peerj.343

Pal, M. (2005). Random Forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217-222. doi:10.1080/01431160412331269698

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology, 56*(1), 45-50.

Phua, J., Koay, E., & Lee, K. (2008). Lactate, procalcitonin, and amino-terminal pro-B-type natriuretic peptide versus cytokine measurements and clinical severity scores for prognostication in septic shock. *Shock, 29*. doi:10.1097/SHK.0b013e318150716b

Prucha, M., Bellingan, G., & Zazula, R. (2015). Sepsis biomarkers. *Clinica Chimica Acta, 440*, 97-103. doi:10.1016/j.cca.2014.11.012

Ramos-Jimenez, G., del Campo-Avila, J., & Morales-Bueno, R. (2009, Nov. 30 2009-Dec. 2 2009). *Hybridizing Ensemble Classifiers with Individual Classifiers.* Paper presented at the Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation *Encyclopedia of database systems* (pp. 532-538): Springer.

Reinhart, K., Bauer, M., Riedemann, N. C., & Hartog, C. S. (2012). New approaches to sepsis: molecular diagnostics and biomarkers. *Clinical microbiology reviews, 25*(4), 609-634. doi:10.1128/CMR.00016-12

Ren, Y., & Suganthan, P. N. (2012, 9-12 July 2012). *Empirical comparison of bagging-based ensemble classifiers.* Paper presented at the Information Fusion (FUSION), 2012 15th International Conference on.

Ribas Ripoll, V. J., Vellido, A., Romero, E., & Ruiz-Rodríguez, J. C. (2014). Sepsis mortality prediction with the Quotient Basis Kernel. *Artificial Intelligence in Medicine, 61*(1), 45-52. doi:10.1016/j.artmed.2014.03.004

Ricciuto, D. R., dos Santos, C. C., Hawkes, M., Toltl, L. J., Conroy, A. L., Rajwans, N., . . . Liles, W. C. (2011). Angiopoietin-1 and angiopoietin-2 as clinically informative prognostic biomarkers of morbidity and mortality in severe sepsis*. *Critical care medicine, 39*(4), 702-710. doi:10.1097/CCM.0b013e318206d285

Riedel, S. (2012). Procalcitonin and the role of biomarkers in the diagnosis and management of sepsis. *Diagnostic Microbiology and Infectious Disease, 73*(3), 221-227. doi:http://dx.doi.org/10.1016/j.diagmicrobio.2012.05.002

Rivers, E. P., Jaehne, A. K., Nguyen, H. B., Papamatheakis, D. G., Singer, D., Yang, J. J., . . . Klausner, H. (2013). Early Biomarker Activity in Severe Sepsis and Septic

Shock and a Contemporary Review of Immunotherapy Trials: Not a Time to Give Up, But to Give It Earlier. *Shock, 39*(2), 127-137. doi:10.1097/SHK.0b013e31827dafa7

Rivers, E. P., Kruse, J. A., Jacobsen, G., Shah, K., Loomba, M., Otero, R., & Childs, E. W. (2007). The influence of early hemodynamic optimization on biomarker patterns of severe sepsis and septic shock. *Critical care medicine, 35*(9), 2016-2024.

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L.-W., Moody, G., . . . 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), 952-960. doi:10.1097/CCM.0b013e31820a92c6

Sawyer, A. M., Deal, E. N., Labelle, A. J., Witt, C., Thiel, S. W., Heard, K., . . . Kollef, M. H. (2011). Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. *Critical care medicine, 39*(3), 469-473. doi:10.1097/CCM.0b013e318205df85

Shapiro, N. I., Trzeciak, S., Hollander, J. E., Birkhahn, R., Otero, R., Osborn, T. M., . . . Milzman, D. (2009). A prospective, multicenter derivation of a biomarker panel to assess risk of organ dysfunction, shock, and death in emergency department patients with suspected sepsis. *Critical care medicine, 37*(1), 96-104.

Simon, R. M., Subramanian, J., Li, M.-C., & Menezes, S. (2011). Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. *Briefings in Bioinformatics, 12*(3), 203-214. doi:10.1093/bib/bbr001

Singer, M., Deutschman, C. S., Seymour, C., & et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA, 315*(8), 801-810. doi:10.1001/jama.2016.0287

Skurichina, M., Kuncheva, L., & Duin, R. W. (2002). Bagging and Boosting for the Nearest Mean Classifier: Effects of Sample Size on Diversity and Accuracy. In F. Roli & J. Kittler (Eds.), *Multiple Classifier Systems* (Vol. 2364, pp. 62-71): Springer Berlin Heidelberg.

Srimani, P., & Koti, M. S. (2013). Medical diagnosis using ensemble classifiers-a novel machine-learning approach. *Journal of Advanced Computing, 1*, 9-27. doi:10.7726/jac.2013.1002

Staley, J. R., Jones, E., Kaptoge, S., Butterworth, A. S., Sweeting, M. J., Wood, A. M., & Howson, J. M. M. (2017). A comparison of Cox and logistic regression for use in genome-wide association studies of cohort and case-cohort design. *European Journal Of Human Genetics, 25*, 854. doi:10.1038/ejhg.2017.78

https://www.nature.com/articles/ejhg201778#supplementary-information

Steyerberg, E. W., Calster, B. V., & Pencina, M. J. (2011). Performance Measures for Prediction Models and Markers: Evaluation of Predictions and Classifications. *Revista Española de Cardiología (English Edition), 64*(09), 788-794.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.), 21*(1), 128-138. doi:10.1097/EDE.0b013e3181c30fb2

Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychological methods, 14*(4), 323-348. doi:10.1037/a0016973

Sturgess, D. J., Marwick, T. H., Joyce, C., Jenkins, C., Jones, M., Masci, P., . . . Venkatesh, B. (2010). Prediction of hospital outcome in septic shock: a prospective comparison of tissue Doppler and cardiac biomarkers. *Critical Care, 14*(2), 1-11. doi:10.1186/cc8931

Sundén-Cullberg, J., Rylance, R., Svefors, J., Norrby-Teglund, A., Björk, J., & Inghammar, M. (2017). Fever in the Emergency Department Predicts Survival of Patients With Severe Sepsis and Septic Shock Admitted to the ICU*. *Critical care medicine, 45*(4), 591-599. doi:10.1097/ccm.0000000000002249

Therneau, T. (2018). A package for survival analysis in R. R package version 2.42-3. *Retrived from http://CRAN. R-project. org/package= survival*.

Therneau, T., Crowson, C., & Atkinson, E. (2018). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. Retrieved from https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

Thiel, S. W., Rosini, J. M., Shannon, W., Doherty, J. A., Micek, S. T., & Kollef, M. H. (2010). Early prediction of septic shock in hospitalized patients. *Journal of Hospital Medicine, 5*(1), 19-25. doi:10.1002/jhm.530

Tolosie, K., & Sharma, M. K. (2014). Application of Cox Proportional Hazards Model in Case of Tuberculosis Patients in Selected Addis Ababa Health Centres, Ethiopia. *Tuberculosis Research and Treatment, 2014*, 11. doi:10.1155/2014/536976

Torabi, M., Moeinaddini, S., Mirafzal, A., Rastegari, A., & Sadeghkhani, N. (2016). Shock index, modified shock index, and age shock index for prediction of mortality in Emergency Severity Index level 3. *The American Journal of Emergency Medicine, 34*(11), 2079-2083. doi:10.1016/j.ajem.2016.07.017

Tsujitani, M., Tanaka, Y., & Sakon, M. (2012). Survival Data Analysis with Time-Dependent Covariates Using Generalized Additive Models. *Computational and Mathematical Methods in Medicine, 2012*, 986176. doi:10.1155/2012/986176

Valentini, G., & Masulli, F. (2002). Ensembles of Learning Machines. In M. Marinaro & R. Tagliaferri (Eds.), *Neural Nets* (Vol. 2486, pp. 3-20): Springer Berlin Heidelberg.

Verburg, I. W. M., Holman, R., Dongelmans, D., de Jonge, E., & de Keizer, N. F. (2018). Is patient length of stay associated with intensive care unit characteristics? *Journal of Critical Care, 43*, 114-121. doi:https://doi.org/10.1016/j.jcrc.2017.08.014

Verma, B., & Rahman, A. (2012). Cluster-Oriented Ensemble Classifier: Impact of Multicluster Characterization on Ensemble Classifier Learning. *Knowledge and Data Engineering, IEEE Transactions on, 24*(4), 605-618. doi:10.1109/TKDE.2011.28

Walters, S. J. (2009). What is a Cox Model? *University of Oxford Medical Sciences Division.* Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/cox_model.pdf

Wang, L., Shen, J., & Thall, P. F. (2014). A modified adaptive Lasso for identifying interactions in the Cox model with the heredity constraint. *Statistics & Probability Letters, 93*, 126-133. doi:10.1016/j.spl.2014.06.024

Wang, S., Wu, F., & Wang, B.-H. (2010). Prediction of Severe Sepsis Using SVM Model. In H. R. Arabnia (Ed.), *Advances in Computational Biology* (Vol. 680, pp. 75-81): Springer New York.

Wang, Y., Chen, W., Heard, K., Kollef, M. H., Bailey, T. C., Cui, Z., . . . Chen, Y. (2015). Mortality Prediction in ICUs Using A Novel Time-Slicing Cox Regression Method. *AMIA Annual Symposium Proceedings, 2015*, 1289-1295.

Williams, J. A., Weakley, A., Cook, D. J., & Schmitter-Edgecombe, M. (2013). *Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia.* Paper presented at the Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence.

Wu, R.-f., Zheng, M., & Yu, W. (2016). Subgroup Analysis with Time-to-Event Data Under a Logistic-Cox Mixture Model. *Scandinavian Journal of Statistics*, n/a-n/a. doi:10.1111/sjos.12213

Xia, J., Ghamisi, P., Yokoya, N., & Iwasaki, A. (2018). Random Forest Ensembles and Extended Multiextinction Profiles for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing, 56*(1), 202-216.

Xia, J., Liao, W., Chanussot, J., Du, P., Song, G., & Philips, W. (2015). Improving Random Forest with ensemble of features and semisupervised feature extraction. *IEEE Geoscience and Remote Sensing Letters, 12*(7), 1471-1475.

Xu, G., Sen, B., & Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electronic journal of statistics, 8*(1), 1345-1379. doi:10.1214/14-EJS927

Yao, J., Guo, H., & Yang, X. (2015). PPCM: Combing Multiple Classifiers to Improve Protein-Protein Interaction Prediction. *International Journal of Genomics*.

Yi, J., Slaughter, A., Kotter, C. V., Moore, E. E., Hauser, C. J., Itagaki, K., . . . Peltz, E. (2015). A "clean case" of systemic injury: mesenteric lymph after hemorrhagic shock elicits a sterile inflammatory response. *Shock, 44*(4), 336-340. doi:10.1097/shk.0000000000000431

Zhiwen, Y., Le, L., Jiming, L., & Guoqiang, H. (2015). Hybrid Adaptive Classifier Ensemble. *Cybernetics, IEEE Transactions on, 45*(2), 177-190. doi:10.1109/TCYB.2014.2322195