1-1-2014

Improving Emergency Department Patient Flow Through Near Real-Time Analytics

Shanshan Qiu
Wayne State University,

IMPROVING EMERGENCY DEPARTMENT PATIENT FLOW THROUGH NEAR REAL-TIME ANALYTICS

by

SHANSHAN QIU

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: INDUSTRIAL ENGINEERING

Approved by:

Advisor Date

DEDICATION

This dissertation is lovingly dedicated to my husband Dawei Wang,

my son Ben Daoyi Wang, and my parents.

ACKNOWLEDGMENTS

First of all, I'd like to acknowledge my tremendous appreciation and deep thanks to my advisors, Dr. Ratna Babu Chinnam, who is also the chair of my PhD dissertation committee, and Dr. Alper E. Murat. They both are excellent mentors, enthusiastic researchers, diligent professors, faithful and sincere friends, and role models for me. Their enthusiasm enlightens my interest in dissertation research work, and there is no doubt in my mind that without their support, encouragement, and help, I could not have completed this dissertation.

Next, I'd like to thank Dr. Simon Ng, the Associate Dean for research of College of Engineering of Wayne State University, for his support to me as my previous supervisor, advice to me as a senior professor and researcher who is very successful in his area, and his company as a good friend.

I would also like to thank Dr. Bassam Batarse, Dr. Xuewen Chen, Dr. Evrim Dalkiran and Dr. Qingyu Yang for serving in my dissertation committee. Dr. Batarse gave me tremendous help in helping under the practical aspects/value of this dissertation. Dr. Xuewen Chen is not only a knowledgeable professor who gave me very good suggestions on my research, but also a very good friend whose humor always made me smile every time we saw each other. Both Dr. Evrim Dalkiran and Dr. Qingyu Yang are young but very knowledgeable. Their enthusiasm for the research inspired me to do high quality research work.

I would like to thank Hakimuddin Neemuchwala and Michael Lederle from VA-CASE for extracting all the data and related information for this research. I would also like to thank the VA-CASE leadership for sponsoring and supporting part of my dissertation research.

In addition, I would like to thank my long-time friend Mike Switney for helping me improve my English and adapt to life in a new country. Besides, I want to thank to my supervisor, Arnold

Babila and my manager Varsha Venkatesh at Ford Motor Credit Company (FMCC) for their generous support for me to maintain the flexible research environment, which allowed me continue working at exciting projects at FMCC while finishing my dissertation.

TABLE OF CONTENTS

Dedication ii

Acknowledgements........ iii

List of Tables ix

List of Figures x

CHAPTER 1: INTRODUCTION........ 1

1.1 Motivation 1

1.2 Research Objectives 6

1.3 Research Scope........ 8

CHAPTER 2: A TRIAGE DECISION SUPPORT FRAMEWORK FOR IMPROVING ED PATIENT FLOW 10

2.1 Introduction 10

2.2 Literature Review 12

2.2.1 General Intervention Methods during Triage for ED Patient Flow Improvement 12

2.2.2 Existing ED Patient Flow Decision Support Frameworks 13

2.2.3 Cost-sensitive Models in Healthcare: Application of Newsvendor Models........ 14

2.3 A Cost Sensitive Ward-bed Reservation Model........ 15

2.3.1 Model Assumptions and Notation........ 16

2.3.2 Cost of Reservation Decisions........ 18

2.3.3 Optimizing Decision: A Newsvendor Model Variant 21

2.4 Model Analysis........ 26

2.4.1 Estimation of Model Parameters 26

2.4.2 Parameter Sensitivity Analysis 32

2.5 Conclusion 36

CHAPTER 3: ADMISSION PREDICTIONS OF EMERGENCY DEPARTMENT PATIENT 38

3.1 Introduction 38

3.1.1 Introduction of Classification Models 39

3.1.2 Feature Selection in Classification 47

3.1.3 Evaluation of Classification 48

3.2 Literature Review 50

3.2.1 Prediction of ED Patient Admission 50

3.2.2 Multi-class Relevance Vector Machine (mRVM) 51

3.3 Integrated Feature Selection mRVM: IF-mRVM 58

3.4 Experiments 63

3.4.1 Performance of IF- mRVM on Benchmark Datasets 63

3.4.2 Performance of IF- mRVM on Emergency Department Data 71

3.5 Conclusion 73

CHAPTER 4: SURVIVAL ANALYSIS MODELS FOR ESTIMATING LENGTH-OF STAY IN EMERGENCY DEPARTMENTS 75

4.1 Introduction 75

4.2 Literature Review 76

4.2.1 Literature: Factors Affecting ED LOS 76

4.2.2 Literature: Statistical Modeling of Skewed & Long-Tailed Data 77

4.2.3 Attempts to Estimate Parts of ED LOS 78

4.3 Survival Analysis Models 78

4.3.1 Semi-parametric Proportional Hazard Model 79

4.3.2 Parametric Hazard Rate Models 80

4.3.3 Neural Network Models for Survival Analysis 80

4.4 Survival Analysis Models for Estimating ED LOS 81

4.4.1 Data 81

4.4.2 Models 82

4.4.2 Experiments & Results 83

4.5 Conclusion 85

CHAPTER 5: CASE STUDY: IMPLEMENTATION OF THE PROPOSED WARD-BED RESERVATION MODELING FRAMEWORK 87

5.1 ED and Ward Data 87

5.2 ED Patient Admission Prediction 90

5.3 Ward-bed Lead-time Estimation 94

5.4 Application of Proposed Cost-sensitive Ward-bed Reservation Model 95

5.4.1 Optimal Reservation Threshold 95

5.4.2 Impact of Proposed Ward-Bed Reservation Model 95

5.5 Conclusion 97

CHAPTER 6: CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH 99

Appendix: Complaint Codes Used by Our Study 102

References 104

Abstract 129

Autobiographical Statement 132

LIST OF TABLES

Table 1 *Confusion Table of a classifier and evaluation measures* 49

Table 2 *Datasets used in the experiments* 63

Table 3 *Results of IF-mRVM* 64

Table 4 *Results of Baseline mRVM* 67

Table 5 *Statistics of patient information in our data* 88

Table 6 *Statistics of ED LOS by Orderable items with highest frequencies* 89

Table 7 *Confusion Table for the RVM classifier on the testing dataset* 92

Table 8 *Results of RVM classifier on the testing dataset when using different minimum probability thresholds for classification* 92

Table 9 *Confusion results of RVM classifier with medicine and surgery ward combined* 93

Table 10 *Results of RVM classifier on the testing dataset when using different minimum probability thresholds for classification* 93

Table 11 *Predicted ward-bed lead-time under different possible and realistic scenarios* 94

Table 12 *Sample of proposed optimal policy results for select medicine ward patients* 96

Table 13 *Savings from making advance reservations according to the proposed policy* 96

LIST OF FIGURES

Figure 1 Research Vision: ED patient flow improvement through proactive planning and decision support models for improved operational intelligence 8

Figure 2. Change of $\boldsymbol{TR}*$ with the ratio $\boldsymbol{CB/CW}$ 24

Figure 3. Cost associated with ward-bed reservation decisions and associated decision tree. 27

Figure 4. Effect of ward-bed lead-time variability on the expected cost. 30

Figure 5. Effect of ward-bed lead-time variability on the expected cost savings opportunity due to perfect ward-bed lead-time information. 31

Figure 6. Effect of Weibull ED LOS scale parameter λ on cost of reservations. 33

Figure 7. $\boldsymbol{CTR}$ and $\boldsymbol{CTNR}$ cost differences under different levels of uncertainty regarding patient's total waiting time from triage to the physician's disposition decision (i.e., $\boldsymbol{t}$) 35

Figure 8. Effect of target-ward utilization on C_TR and C_TNR 36

Figure 9. Effect of ED LOS scale parameter λ and target-ward utilization U on reservation threshold ... 36

Figure 10. Histogram of ED LOS of the patients in our data 83

CHAPTER 1: INTRODUCTION

1.1 Motivation

Emergency department (ED) crowding is an international phenomenon facing emergency physicians, nurses, and their patients, and it has become the subject of significant public and academic attention (Moskop, Sklar et al. 2009, Moskop, Sklar et al. 2009, Boyle, Beniuk et al. 2012, Powell, Khare et al. 2012). ED crowding has been called a national crisis according to a 2006 report of the Institute of Medicine (IOM) Committee(Institute of Medicine 2007) . Many emergency departments (EDs) across the country are crowded. Nearly half the EDs report operating at or above capacity, and 9 out of 10 hospitals report holding or “boarding” admitted patients in the ED while they await inpatient beds (McHugh, Dyke et al. 2011).

ED crowding is known to cause a number of adverse outcomes and are briefly discussed next.

Patient Treatment Delays and Dissatisfaction: Liu et al. (Liu, Hobgood et al. 2003) investigated ED patient flow during periods of acute ED crowding and concluded that patients waited significantly longer for an ED bed when the crowding level is high. Delays have been reported for analgesia, antibiotic therapy, and thrombolysis or percutaneous coronary intervention (Schull, Vermeulen et al. 2004, Hwang, Richardson et al. 2006, Kulstad and Kelley 2009) as well as patients with severe pain (Pines and Hollander 2008). One author has estimated that more people die avoidably as the result of crowding in ED in New Zealand than in road traffic collisions (Johnston 2008).

Patient Mortality: Miro et al. (Miro, Antonio et al. 1999) conducted a study to investigate the health care quality associated with ED overcrowding and noted a statistically significant positive correlation between mortality rates and weekly number of visits (with a correlation coefficient of 0.01). Another study (Sprivulis, Da Silva et al. 2006) associated a combined measure of hospital and ED crowding with an increased risk of mortality at 2, 7, and 30 days after hospital admission, and they concluded that hospital and ED overcrowding is associated with increased mortality. Begley et al. (Begley, Chang et al. 2004) conducted a study in Houston and found higher mortality among trauma patients admitted during ambulance diversion (attributed to ED crowding).

Patients Leaving without Receiving Care: A number of studies have indicated that the rate of patients leaving without being seen closely correlated with waiting times (Kyriacou, Ricketts et al. 1999) and ED occupancy (Polevoi, Quinn et al. 2005) or crowding level (Weiss, Ernst et al. 2005). Specifically, it was reported that in 2007, 1.9 million people—representing 2 percent of all ED visits—left the ED before being seen, typically because of long wait times (Niska, Bhulya et al. 2010). Another study (Rowe, Channan et al. 2006) reported that patients frequently cited long waiting times as a reason for leaving without being seen. Furthermore, some studies investigated the characteristics of the leaving without seen patients, and they reported that a significant part of those patients are those who need urgent medical attention (Baker, Stevens et al. 1991). Patients who left the ED without being seen are also twice as likely to report worsened health problems (Bindman, Grumbach et al. 1991).

Ambulance Diversion: A national survey by Burt et al. (Burt, McCaig et al. 2006) found that approximately 501,000 ambulance diversions occurred in the U.S. during a single year, and approximately 70% of these were from large EDs. A point-prevalence study of ED crowding by Schneider et al. (Schneider, Gallery et al. 2003) found that 11% of U.S. EDs were simultaneously diverting ambulances. A study by Eckstein et al. (Eckstein and Chan 2004) determined the effect of ED crowding on paramedic ambulance availability, and concluded that crowding has resulted in delays for paramedics waiting to transfer patients.

Hospital Financial Losses: A study by Bayley et al. (Bayley, Schwartz et al. 2005) estimated that the hospital studied lost $204 in potential revenue per patient due to extended boarding times. Another study by Krochmal et al. (Krochmal and Riley 1994) estimated that patient boarding in the ED has cost the hospital $6.8 million over three years. A 2006 study conducted by McConnell et al. (McConnell, Richards et al. 2006) at a large academic medical center (AMC) found that each hour on diversion was associated with $1,086 in foregone hospital revenues. Another study conducted at a different AMC (Pines, Batt et al. 2011) showed that a 1-hour reduction in ED boarding time would result in over $9,000 of additional revenue by reducing ambulance diversion and the number of patients leaving the hospital without receiving care.

Harmful to Staff: Studies have found that there are associations between ED congestion and absenteeism, staff sickness, and burnout, which results in experienced staff leaving and more junior staff, or agency staff delivering an increasingly busy and inefficient service (Atzema, Bandiera et al. 2005, Jelinek, Weiland et al. 2010). A survey of Canadian emergency physicians also found that job dissatisfaction was closely related to the perceived scarcity of resources (Rondeau and Francescutti 2005).

In the past two decades, a number of researchers have investigated the causes, consequences and interventions of ED overcrowding (United States General Accounting Office 2003, Hoot and Aronsky 2008, Olshaker 2009, Welch, Asplin et al. 2011). Asplin et al.'s (Asplin, Magid et al. 2003) conceptual model partitions the causes into three interdependent components: *input*, *throughput*, and *output*.

Input factors mostly refer to the 'source' of noncritical demand (i.e., patients with non-urgent needs) for ED. Non-urgent visits (situation where low-acuity patients choose the ED for non-urgent medical care) are identified as an important factor causing crowding (Afilalo, Marinovich et al. 2004, Howard, Davis et al. 2005). There is evidence showing that physicians and clinics refer patients to the ED for a variety of reasons, including convenience for after-hours care, reluctance to take on complex cases, liability concerns, and the need for diagnostic testing that cannot be performed in their offices.(Institute of Medicine 2007) *Throughput factors* refer to ED care operations, processes and their efficiency. Inadequate staffing is considered as the main throughput factor that may cause crowding (Lambe, Washington et al. 2003, Schneider, Gallery et al. 2003). *Output factors* mostly refer to availability of timely follow-up appointments, i.e., the efficiency of admitting ED patient to inpatient units. Inpatient boarding and hospital bed shortages have been reported to be the two main output factors (Andrulis, Kellermann et al. 1991, Fromm, Gibbs et al. 1993, Fatovich, Nagree et al. 2005). A good review of studies of the factors that affect ED patient flow by Hoot can be found in (Hoot and Aronsky 2008).

In terms of the interventions to ED, many solutions have been investigated and proposed. Traditional strategies involve increasing capacity of ED, including number of beds (Bazarian, Schneider et al. 1996, Kelen, Scheulen et al. 2001, Ross, Naylor et al. 2001, Moloney, Bennett et al. 2006) or physicians (Shaw and Lavelle 1998, Partovi, Nelson et al. 2001, Bucheli and Martina 2004, Donald, Smith et al. 2005, Russ, Jones et al. 2010, Nestler, Fratzke et al. 2012). Another body of research focuses on applying operations research techniques for ED patient flow improvement. They propose supporting solutions through improved business intelligence. Cochran et al. (Cochran and Roche 2009) derive an open queuing network model of an ED and introduces a new paradigm of ED care that reduces "walk-aways" to increase the capacity of an

ED to treat patients. Discrete event simulation (DES) methods are used to conduct advanced system-level investigation of ED operations by Connelly et al. (Connelly and Bair 2004). Green et al. (Green, Soares et al. 2006) used a Lag SIPP queuing analysis to gain insights in identifying provider staffing patterns to reduce the fraction of patients who leave ED without being seen. Saghafian et al. used a combination of analytic and simulation models to determine whether a 'streaming policy' can improve ED performance, where is it most likely to be effective, and how it should be implemented for maximum performance. For a good systematic review of operations research methods focusing on ED patient flow improvement, see (Saghafian, Austin et al. 2014) .

In the last decade, "lean", which has originated as a production philosophy in automotive manufacturing industry, has been implemented rapidly as a strategy to solve ED crowding and improve patient flow. Lean is a business model and collection of tactical methods that emphasize eliminating non-value added activities (waste) while delivering quality products on time at least cost with greater efficiency (Holweg 2007). Interested readers can find a good review of lean in (Holweg 2007). In the context of patient flow improvement, lean strategies involve process redesign (Spaite, Bartholomeaux et al. 2002, King, Ben-Tovim et al. 2006) physician triage (Han, France et al. 2010), eliminating or combining steps in process (Kolker 2008, Ng, Vail et al. 2010), or lab/imaging orders being administered earlier in the process (Dickson, Singh et al. 2009) and so on. A good review of lean being implemented in ED can be found in a study by Holden (Holden 2011). Another review by Eitel et al. (Eitel, Rudkin et al. 2010) discusses specific methods to improve the ED quality and flow of studies describing the ED as a service business. It was mentioned in (Eitel, Rudkin et al. 2010) that 'the purpose of this review is to serve as a background for emergency physicians and managers interested in applying process reengineering methods to improving ED flow, reducing waiting times, and maximizing patient satisfaction'. Methods discussed include demand management, critical pathways, process-mapping, Emergency Severity Index (ESI) triage, bedside registration, lean and six-sigma management methods, statistical forecasting, queuing systems, discrete event simulation modeling, and balanced scorecards.

Despite many efforts, scientific knowledge remains limited as regards which strategies and pragmatic approaches actually improve patient flow in EDs. It was concluded in (Eitel, Rudkin et al. 2010) that 'it is currently unknown which strategies provide the best solution to fix throughput in the ED'.

In this work, we mostly investigate opportunities for developing effective decision support models that exploit near real-time (NRT) information to enhance the "operational intelligence" within ED, and in turn, patient flow. The distinction between the terms "near real-time" and "real-time" is somewhat relative and here refers to information such as patient vitals, complaints, severity assessment from triage, outcomes of laboratory/imaging work collected while patient is in ED, status of beds in target wards and so on.

1.2 Research Objectives

Research has suggested that if the hospital admissions of ED patients can be predicted early during triage and communicated to different departments of a hospital, then necessary steps can be taken early to reduce transfer delays (Peck, Benneyan et al. 2012). If hospital admission decisions can be predicted in advance (i.e., upon patient triage or soon after), then this information can be passed on to the target inpatient ward departments where staff can begin their preparations early on and thereby reduce patient transfer delays and boarding.

The objective of this research is to develop a comprehensive modeling and analytics framework for improved patient flow in ED by exploiting near real-time information collected and readily available at triage or right after. In particular, the focus is on developing a framework for streamlining ED patient flow through novel cost-sensitive advance ward-bed reservation policies that can be employed during triage or right after. The policies are to be cost-sensitive in that they should optimally account for the different costs that stem from incorrect decisions (e.g., a ward-bed is reserved for a patient who is sent home by the ED physician, potentially blocking the use of the ward-bed by another patient; no reservation is made for a patient that is predicted with high confidence to be admitted right during triage, subjecting the patient to unnecessary waiting past the disposition decision). The advance reservation policies are necessarily driven by predictions of likelihood of ED patient admission as well as predictions of the expected length-of-stay (LOS) by the patient in ED. Two achieve this objective, we have two secondary objectives:

1) To develop models to predict the admission likelihood and target admission wards of ED patients.

2) To develop models to estimate ED LOS of ED patients.

The motivation and the vision behind this research is outlined in Figure 1.

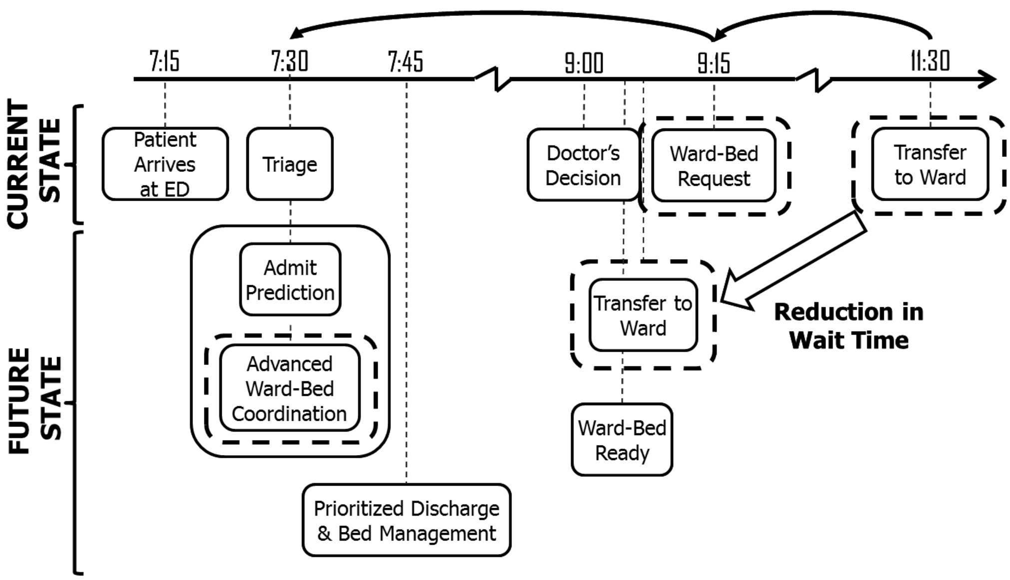


Figure 1 Research Vision: ED patient flow improvement through proactive planning and decision support models for improved operational intelligence

1.3 Research Scope

While there are a lot of solutions for improving ED patient flow as discussed in section 1.1, the focus of this work is to develop a near real-time triage decision support system to reduce ED boarding and improve ED patient flow. We develop a novel variant of a newsvendor modeling framework that integrates patient admission probability prediction within a cost-sensitive bed reservation system to improve the effectiveness of bed coordination efforts and reduce the boarding times of ED patients along with the resulting costs. Specifically, we propose a cost sensitive bed reservation policy that recommends optimal bed reservation times for patients. The policy relies on classifiers that estimate the probability that the ED patient will be admitted using the patient information collected and readily available at triage or right after. The policy is cost sensitive in that it accounts for costs associated with patient admission prediction misclassification as well as costs associated with incorrectly selecting the reservation time.

Our aim is to answer two specific questions to streamline the ED patient flow: (1) What is the optimal admission probability threshold for a patient beyond which a ward-bed reservation should be made? (2) If the decision is to make a ward-bed reservation for a patient, then what is the optimal reservation time slot? Optimization of the reservation time slot is critical since reservations made for a slot earlier than necessary can lead to wasted bed capacity. Similarly, the reservations made for a slot later than necessary leads to increased boarding time for patients. Both alternatives result in a variety of tangible/intangible costs, compromised health outcomes, and patient dissatisfaction. Furthermore, uncertainty in a patient's ED length-of-stay (i.e., time from end of triage until the physician's disposition) and the lead-time to obtain a bed in the target ward, make the reservation decisions challenging. Our cost sensitive ward-bed reservation model effectively accounts for these various costs and uncertainties.

CHAPTER 2: A TRIAGE DECISION SUPPORT FRAMEWORK FOR IMPROVING ED PATIENT FLOW

2.1 Introduction

As patients can arrive at an emergency department (ED) at any time and with any complaint, a key part of the operation of an ED is the prioritization of patients based on their treatment needs, and this process is called 'triage'. Triage is the term derived from a French verb, and it means 'to sort' or 'to choose'. In an ED, triage is normally the first stage the patient passes through. Triage nurses collect basic information such as 'vitals' (arrival mode, height, weight, temperature, blood pressure etc.) and patient complaints and conduct severity assessment of the patient condition. The patients are typically rated/stratified according to the Emergency Severity Index (ESI) from level 1 (most urgent) to level 5 (least resource intensive) according to the level of severity of their situation and this forms a proxy measure of how long an individual patient can safely wait for a medical screening examination and treatment. ESI triage scale is originally developed by ED physicians Richard Wuerz and David Eitel in the U. S. (Gilboy, Travers et al. 1999, Wuerz, Milne et al. 2000)

The Institute of Medicine (IOM) published the landmark report, "The Future of Emergency Care in the United States," and described the worsening crisis of crowding that occurs daily in most EDs (Institute of Medicine 2007) With more patients waiting longer in the waiting room, the accuracy of the triage is critical. Under-categorization (under triage) leaves the patient at risk for deterioration while waiting. Over-categorization (over-triage) uses scarce resources, limiting availability of an open ED bed for another patient who may require immediate care. Rapid, accurate triage of the patient is important for successful ED operations. There is growing interest in the establishment of standards for triage acuity assessment in the United States to support clinical care, ED surveillance, benchmarking, and research activities (National Center for Injury and Control 1997, Gilboy, Travers et al. 1999, Barthell, Coonan et al. 2004, Handler, Adams et al. 2004, Haas, Travers et al. 2008)

Historically, EDs in the United States did not use standardized triage acuity rating systems. Since 2000, there has been a trend toward standardization of triage acuity scales that have five levels (e.g., 1- resuscitation, 2- emergent, 3- urgent, 4- less urgent, 5- non-urgent) (Gilboy 2012). Based on expert consensus of currently available evidence, the American College of Emergency Physicians (ACEP) and the Emergency Nurses Association (ENA) support the adoption of a reliable, valid five-level triage scale such as the ESI (American College of Emergency Physicians 2010). Following the adoption of this position statement, the number of EDs using three-level triage systems has decreased, and the number of EDs using the five-level ESI triage system has increased significantly (McHugh and Tanabe 2011). For more information regarding triage systems and on-going research, see (Gilboy 2012).

Traditionally, the advantages of comprehensive triage are considered as immediate identification of patients with life-threatening or emergent conditions and administration of basic first-aid measures (FitzGerald, Jelinek et al. 2010). However, recently, triage-related interventions to improve patient flow in emergency departments have also been extensively investigated. In this Chapter, we investigate how near real-time (NRT) information that is collected at ED triage can be analyzed to improve the "operational intelligence" and shared with different departments of a hospital to improve the ED patient flow. Specifically, we propose a novel variant of a newsvendor modeling framework that integrates patient admission probability prediction within a cost-sensitive bed reservation system to improve the effectiveness of bed coordination efforts and reduce the boarding times of ED patients along with the resulting costs. The proposed modeling framework is a triage decision support system. Note that although we name the system as a triage decision support system, it allows the decisions to be updates as more information becomes available during the patients' ED visit. For instance, the decision of the optimal reservation time slot for a patient could change if the laboratory/imaging test results come out earlier than expected. In the reminder of this Chapter, we first review the literature that is related to our work, and then we propose our model and investigate the insights of the proposed model. Conclusion and findings are discussed in the last section.

2.2 Literature Review

In this section, we review the studies that related to proposed model. We first discuss general methods that related to intervention on triage to improve ED patient flow; we discuss the previous proposed decision support systems of ED for patient flow improvement.

2.2.1 General Intervention Methods during Triage for ED Patient Flow Improvement

A number of studies have proposed techniques that can be used at triage to improve the ED patient flow. Strategies like *fast track* (patients, following triage or brief evaluation, are divided into different processes) (Fernandes, Christenson et al. 1996, Kilic, Agalar et al. 1998, Bond 2001, Ardagh, Wells et al. 2002, Cooke, Wilson et al. 2002, Rogers, Ross et al. 2004, Patel and Vinson 2005, Darrab, Fan et al. 2006, King, Ben-Tovim et al. 2006, O'Brien, Williams et al. 2006, Rodi, Grau et al. 2006, Sanchez, Smally et al. 2006, Kelly, Bryant et al. 2007, Considine, Kropman et al. 2008, Ieraci, Digiusto et al. 2008, Kwa and Blake 2008), *team triage* (triage handled by a team that includes a physician) (Partovi, Nelson et al. 2001, Richardson, Braitberg et al. 2004, Subash, Dunn et al. 2004, Travers and Lee 2006, Holroyd, Bullard et al. 2007), *point-of-care testing* (performing laboratory analysis in the ED) (Tsai, Nash et al. 1994, Parvin, Lo et al. 1996, Kendall, Reeves et al. 1998, Murray, Leroux et al. 1999, Lee-Lewandrowski, Corboy et al. 2003, Singer, Viccellio et al. 2008), nurse-requested laboratory and imaging tests (e.g., some hospitals have piloted a routine of nurse-requested x-ray) (Thurston and Field 1996, Parris, McCarthy et al. 1997, Lindley-Jones and Finlayson 2000).

A great review of methods of triage-related interventions to improve patient flow in emergency departments was given by Oredsson et al. (Oredsson, Jonsson et al. 2011). They concluded that the effect of fast track on waiting time, length of stay, and patients that left without care was moderately strong. The effect of team triage on left without being seen was relatively strong, but the evidence for all other interventions was limited or insufficient.

2.2.2 Existing ED Patient Flow Decision Support Frameworks

We also identified a few studies that designed decision support tools for ED to improve patient flow. Ahmed et al. (Ahmed and Alkhamis 2009) integrated simulation with optimization to design a decision support tool for the operation of an ED unit at a governmental hospital in Kuwait. They presented a methodology that uses system simulation combined with optimization to determine the optimal number of doctors, lab technicians and nurses required to maximize patient throughput and to reduce patient time in the system subject to budget restrictions. Reeder et al. (Reeder, Burleson et al. 2003) designed real-time emergency analysis of demand indicators to predict ED demand and resource needs for real-time monitoring of ED operations.

Peck et al. (Peck, Benneyan et al. 2012) proposed methodologies to predict ED to inpatient unit admissions at the time of ED triage and recommend starting bed coordination early on while patients are still receiving the ED treatment to reduce the boarding delays. They propose aggregating individual patient admission probabilities into a summative forecast for the near-future inpatient ward-bed demand. They conclude with a recommendation to use this summative forecast as an alternative to traditional ED crowding measures to guide inpatient bed management decisions.

Haupert et al. (Lee and Atallah 2013, Haupert, Lee et al. 2014) developed an ED decision-support system that couples machine learning, simulation, and optimization to help the healthcare administrators to optimize workflow globally, taking into account the uncertainties of incoming injuries and diseases and associated care. The system does not change physical layout, but focuses on process consolidation, operations tracking, and staffing. The system led to ED and trauma efficiency improved throughput by over 16.2% and reduced the number of patients who left without being seen by over 30%.

2.2.3 Cost-sensitive Models in Healthcare: Application of Newsvendor Models

As stated earlier, our proposed framework is based on a modified version of the classic Newsvendor model in inventory management literature. A good discussion of the newsvendor models can be found in (Khouja 1999). In fact, newsvendor model has been used in healthcare widely, in particular for staffing level determination in hospital operating rooms (OR). We refer the interested readers to a good review of studies of newsvendor problems for OR management by Wachtel and Dexter (Wachtel and Dexter 2010). Few studies used variants of newsvendor models for resource planning in hospitals: Olivares et al. (Olivares, Terwiesch et al. 2008) developed an estimation framework that accounts for heterogeneity in the uncertainty faced as well as in the cost parameters, and then applied the proposed model to balance the costs of reserving too much versus too little OR capacity for cardiac surgery cases. He et al. (Biyu He 2012) incorporated workload heterogeneity into the newsvendor problem to determine optimal staffing levels in OR. Green et al. (Green, Savin et al. 2013) used a variant of the newsvendor model to characterize the optimal staffing levels in a hospital under both exogenous and endogenous nurse absenteeism. In this paper, we employ the newsvendor model framework to strike a good balance between making overly early vs. overly late ward-bed reservations for ED patients predicted to be admitted into one of the inpatient wards. The uncertainty stems from the expected length-of-stay (LOS) for the patient in ED. The model seeks to minimize the expected cost associated with keeping the patient waiting after disposition as well as any potential wastage of ward-bed capacity and related costs.

2.3 A Cost Sensitive Ward-bed Reservation Model

For each patient, the inpatient bed reservation model aims to determine whether or not to make the reservation for a bed, the bed reservation time slot, and the target ward while minimizing the expected cost associated with boarding delays and bed capacity wastage. The ideal bed reservation time slot for an ED patient requiring a bed is the time slot that perfectly matches the time of physician's disposition decision. However, uncertainty in the patient's ED LOS prior to physician's disposition decision poses a challenge in achieving this perfect alignment. Using the historical ED LOS data and patient specific information as covariates (e.g., patient's age, complaint, gender etc.), we can estimate the distribution for patient's ED LOS. Based on this distribution, we then identify the optimal bed reservation time slot that minimizes the expected total cost. By comparing this expected cost of making the bed reservation with optimal timing to the expected cost of not making the reservation, we determine whether or not to make the bed reservation and its timing. As discussed in the following sections, since the cost of making a reservation or not depends on the patient's admission probability, the optimal reservation policy (i.e., decision and timing) can be reduced to one of comparing the patient's admission probability to a patient specific admission probability threshold (determined as a function of estimated ED

LOS for the patient as well as the lead-time for obtaining a bed in the target-ward and associated costs).

2.3.1 Model Assumptions and Notation

As is typical with most medical facilities, we assume that hospital's inpatient wards receive patients from ED and other units such as clinics, other wards, and so on. Hence, access to the inpatient unit is competitive, i.e., other patients can take the beds if there is no reservation made in advance for a specific ED patient. On the other hand, if the bed reservation is made for an ED patient, then access to the bed by the other patients is blocked beginning with the reservation time until the patient is transferred to the ward or discharged from the hospital. We further assume that first-in-first-out (FIFO) ED treatment is appropriate for patients of the same acuity level, ESI to be more specific.

In developing the proposed model, we assume that the patient's admission likelihood as well as the target ward can be effectively predicted upon triage or soon after. In addition, we make the assumption that the patient is essentially being considered for admission in a particular ward or being discharged. In other words, any given competition in predicting the admission is between a single target ward and being discharged (model does not entertain reservations in multiple wards for any given patient). This implies that patients being admitted to different wards are clearly separable given the inputs of the admission classifier. The results from our experiments (discussed later) using data from a U.S. Veterans Administration Medical Center (VAMC) strongly support this assumption. For example, our admission classifier reveals that at most 5.4% of the admission predictions at this VAMC are misclassified with respect to the target ward. Most of these misclassifications are even attributable to occasional patient transfers between medicine and surgery wards due to ward-bed shortage. Aggregating the medicine and surgery wards during classification further reduced this misclassification rate to be just 0.2%. It is possible that a patient could be considered for admission to multiple wards, e.g., the classifier might suggest probable admission in more than one ward. In such situations, there are a number of alternative reservation strategies such as making reservations in more than one ward. This extension will be subject of future work.

We also assume that the patient's length-of-stay (LOS) in ED as well as the lead-time to obtain a ward-bed (i.e., the lead-time for obtaining the bed from the time of request) can be estimated. We discuss how we carry out these predictions in our experiments section. Without loss of generality, we restrict the reservation times to discrete time slots (e.g., 15 minute increments) and, for ease of exposition, suppress the indices (e.g., patient, ward, and time slot) from variables and parameters.

We use the following notation in the remainder:

p : Patient admission probability

U : Utilization of inpatient ward-beds

T_R : Bed reservation time slot (decision variable)

t : Patient's length-of-stay in ED (from triage to physician's disposition)

$f(t)$: Probability density function of t

T_l : Expected lead-time to obtain an inpatient ward-bed

C_w : Cost of boarding a patient in ED for unit time

C_B : Cost of wasting bed capacity for unit time

C_{TP} : Expected boarding cost of making a reservation for a patient who ends up being admitted

C_{FP}: Expected boarding cost of making a reservation for a patient who ends up not being admitted

C_{TN} : Expected boarding cost of not making a reservation for a patient who ends up not being admitted

C_{FN} : Expected boarding cost of not making a reservation for a patient who ends up being admitted

C_R : Expected boarding cost of making a reservation for a patient

C_{NR} : Expected boarding cost of making no reservation for a patient

C_{TR} : Expected total cost of making a reservation for a patient

C_{TNR} : Expected total cost of making no reservation for a patient

2.3.2 Cost of Reservation Decisions

In estimating the costs associated with patient waiting, we start the clock at the end of triage instead of starting it at the end of physician's disposition decision. While this does not change the results, it does allow us to present sensitivity analysis results more effectively. Hence, patients will always incur a waiting time cost irrespective of the ward-bed reservation decision and physician's disposition time. Using the notation introduced in Section 2.1, the expected wait time beginning with the triage completion and ending with ED physician's disposition is $\int_0^\infty t\, f(t) d_t$, and the associated expected wait cost is $C_w \int_0^\infty t\, f(t) d_t$. In the remainder, for exposition clarity and analytical tractability, we treat the bed reservation time slot, T_R, to be continuous. Upon determining the optimal reservation time, we select the nearest reservation time slot with the better cost performance to request the bed for.

Cost of Making a Reservation. Given a patient's admission probability p , the probability of "reservation" decision being correct for this patient is p and incorrect is $1-p$. Therefore, the expected boarding cost of making a reservation for this patient is expressed follows:

$$C_R = pC_{TP} + (1-p)C_{FP} \tag{1}$$

Both types of costs C_{TP} and C_{FP} are functions of the bed reservation time slot T_R.[1] When a bed is reserved for a patient at T_R, a bed is guaranteed for the patient effective T_R until either the patient acquires a bed in the ward or discharged without being transferred to the ward. This implies that reservation system will not hold a bed until T_R but will assign a bed at T_R and hold it unoccupied until patient is transferred or discharged. The ideal value of T_R would be to be set it equal to t (i.e., patient's length-of-stay in ED until physician disposition), to achieve perfect coordination between ED and the ward. However, at triage, t is not scalar and can only be estimated as a distribution, $f(t)$. Thus, C_{TP} is either attributable to ED patient waiting for a bed (i.e., $T_R > t$; in patient ward-bed is not ready for the patient at the end of ED treatment) or the cost of blocking another patient from gaining the bed and wasting bed capacity due to a premature reservation time slot (i.e., $T_R < t$; bed is reserved earlier than necessary). Therefore, the bed capacity is wasted and access to the bed by another potential patient who needs it is blocked for duration $(t - T_R)$. The probability that another patient needing a ward-bed will be blocked is assumed to be U, the utilization of the target ward. In the presence of effective bed management information systems, ward waiting list can be queried to improve this estimate. In addition, care should be exercised to ensure that U accounts for seasonality in ward-bed demand (e.g., by time of day, day of the week, month of the year). We denote the time elapsed between T_R and t opportunity time. The expected boarding cost of making a reservation for a patient who ends up being admitted (C_{TP}) can then be expressed as:

[1] For exposition clarity, we first assume that the target ward-bed will be ready for the ED patient in T_R^* time units upon the request from the triage staff. However, it is possible that a bed may not be available at T_R^* but will available later, e.g., $T_l > T_R^*$. We will address this scenario in the end of this section.

$$C_{TP} = \int_0^{T_R} (T_R - t)C_w f(t)d_t + \int_{T_R}^{\infty} (t - T_R)(C_B + C_w)Uf(t)d_t \tag{2}$$

The boarding cost of making a reservation for a patient who ends up not being admitted (C_{FP}) results in the blocking of another patient as well as wasting of the bed capacity from an unnecessary reservation that needs to be canceled. For $T_R < t$, then the opportunity time during which the bed could be used by another patient is $t - T_R$. In case $T_R > t$, then the ED staff can cancel the ward-bed reservation at time t and the opportunity time or cost is zero since the cancellation takes place before the bed is ready. Therefore, C_{FP} takes the following form:

$$C_{FP} = \int_{T_R}^{\infty} (t - T_R)(C_B + C_w)Uf(t)d_t \tag{3}$$

The expected boarding cost of making a reservation for the patient C_R can be calculated by substituting (3) and (2) into (1):

$$C_R = p\left(\int_0^{T_R} (T_R - t)\, C_w f(t)d_t + \int_{T_R}^{\infty} (t - T_R)(C_B + C_w)Uf(t)d_t\right)$$
$$+(1 - p)\left(\int_{T_R}^{\infty} (t - T_R)(C_B + C_w)Uf(t)d_t\right). \tag{4}$$

After some algebraic simplifications, we obtain:

$$C_R = p\int_0^{T_R} (T_R - t)C_w f(t)d_t + \int_{T_R}^{\infty} (t - T_R)(C_B + C_w)Uf(t)d_t \tag{5}$$

Cost of Not Making a Reservation: For a patient with the admission probability $\boldsymbol{p}$, the probability of "no reservation" decision being correct is $\mathbf{1}-\boldsymbol{p}$ and being wrong is $\boldsymbol{p}$. Therefore, the expected boarding cost of this decision is as follows:

$$C_{NR} = (1-p)C_{TN} + pC_{FN} \qquad (6)$$

Without a bed reservation, a patient who ends up being admitted per ED physician's decision would need a bed to be requested after the physicians' disposition decision. Hence, the patient will experience boarding time equivalent to the full ward-bed lead-time, T_l, which is in addition to the time spent in ED until the ED physician's disposition. Therefore, C_{FN} takes the following form:

$$C_{FN} = T_l C_w \qquad (7)$$

There is no cost incurred for not making a reservation for a patient who ends up not being admitted as per ED physician's disposition decision:

$$C_{TN} = 0 \qquad (8)$$

Substituting (7) and (8) into (6), we obtain the following expression for C_{NR}:

$$C_{NR} = pT_l C_w \qquad (9)$$

2.3.3 Optimizing Decision: A Newsvendor Model Variant

Equation (5) is a modified version of the cost function of the classic Newsvendor model in inventory management literature and is a convex function with a unique minimum. We refer the interested readers to (Khouja 1999) for a good discussion of the newsvendor models to identify the optimal time slot (T_R^*) that minimizes (5), we use the first order optimality condition:

$$\frac{d(C_R(T_R))}{dT_R} = pC_w \int_0^{T_R} 1 f(t)\, d_t + (C_B + C_w)U \int_{T_R}^{\infty} (-1) f(t) d_t = 0 \tag{10}$$

that can be simplified to,

$$pC_w F(T_R) - (C_B + C_w)U(1 - F(T_R)) = 0. \tag{11}$$

The T_R^* satisfying the condition in (7) can be expressed as:

$$T_R^* = F^{-1}\left(\frac{(C_B + C_w)U}{pC_w + (C_B + C_w)U}\right). \tag{12}$$

We note that by denoting $\delta = C_B/C_w + 1$ and $\widehat{U} = \delta U$, we can parameterize T_R^* in (8) as:

$$T_R^* = F^{-1}\left(\frac{\widehat{U}}{p + \widehat{U}}\right). \tag{13}$$

Hence, the determination of the optimal reservation slot only requires the knowledge of relative costs of bed wastage and patient waiting in addition to the utilization and admission probability. The critical fractile, $\widehat{U}/(p + \widehat{U})$, is a function of relative cost adjusted utilization ($\widehat{U}$) and admission probability p. The relative cost adjusted utilization ($\widehat{U}$) represents the cost of early reservation (i.e., cost of under-forecasting in the classical newsvendor model) which jointly captures waiting cost of another patient needing a bed and cost of bed wastage. The admission probability (p) represents the cost of late reservation (i.e., cost of over-forecasting in the classical newsvendor model) which captures the waiting cost of the patient with the reservation. As the cost of patient waiting increases, $C_B/C_W \to 0$, $\delta \to 1$, the cost of early reservation approaches to the utilization of the ward capacity, i.e., $\widehat{U} \to U$, and the optimal reservation time slot is solely determined through the target ward utilization and patient's admission probability. The effect of increasing the cost of bed wastage relative to the cost of waiting is to increase the weight of utilization from the baseline level of U. One interesting realization of the critical fractile is when the patient's admission likelihood equals to the relative cost adjusted utilization ($\widehat{U} = p$) and the reservation time is set at mean ED LOS, i.e., $T_R^* = \int_0^\infty t\,f(t)d_t$. This realization occurs when $C_B/C_W = C_{B/W}^* = (p - U)/U$. When $C_B/C_W > C_{B/W}^*$ $\left(< C_{B/W}^*\right)$, it is better to reserve the bed later (earlier) than the mean ED LOS. Clearly $C_B/C_W \geq 0$ is always true, hence when the admission probability is less than the ward utilization, we have $C_B/C_W \geq 0 > C_{B/W}^*$ and it is always better to reserve the bed later than the mean ED LOS regardless of theC_B and C_W.

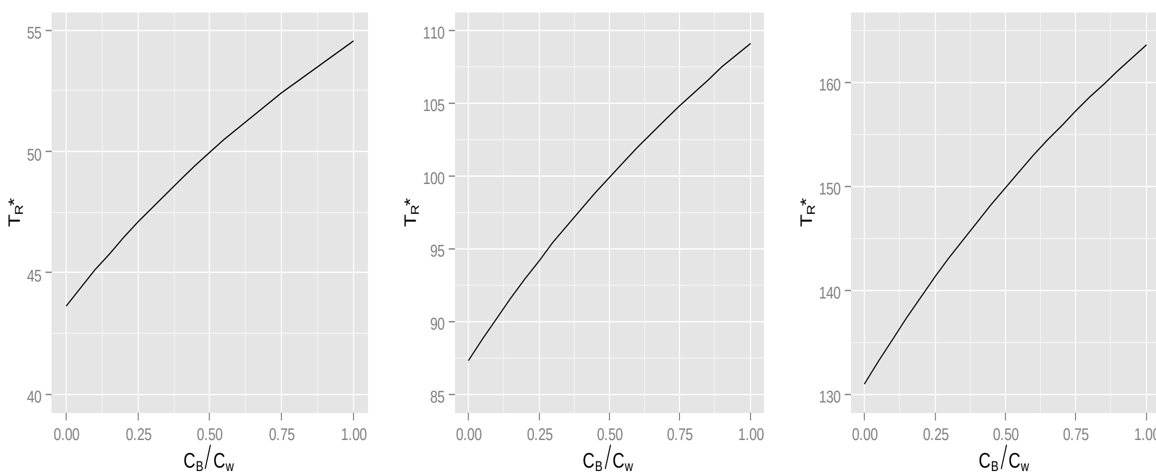
The cost parameters C_B and C_W affect the decision of making a reservation as well as the timing of the reservation. While the former is a dichotomous decision, the latter is continuous. It can be shown that, for the Weibull ED LOS distribution, T_R^* and its sensitivity to C_B/C_W can be expressed as,

$$T_R^* = \lambda(\ln(\alpha))^{\frac{1}{k}}. \tag{14}$$

where λ and k are the scale and shape parameters of the Weibull distribution, respectively.

$$\frac{dT_R^*}{d\left(\frac{C_B}{C_W}\right)} = \frac{\lambda}{k\mu\alpha}(\ln(\alpha))^{\frac{1-k}{k}}. \tag{15}$$

where $\mu = U/p$ and $\alpha = (1 + \delta\mu)$. We investigated the sensitivity of T_R^* with respect to C_B/C_W at different scale (λ) and shape (k) parameters for the Weibull distribution. As an example, Fig. 1 illustrates how T_R^* varies with C_B/C_W when target ward utilization U = 0.8 and patient probability of admission $p = 0.7$. The values of k in (a), (b), (c) are all set to 2 (found very reasonable for data from our VAMC), and the λ values are also reasonably set to 50, 100, 150 minutes, respectively. We only considered the C_B/C_W range of [0,1], which is the case for most practical settings. Increasing C_B/C_W from 0 all the way to 1, the value of T_R^* increase is about 25% for all instances with an approximate increase of 2.5% per C_B/C_W increment of 0.1. In comparison, for a given C_W, each 0.1 increment ofC_B/C_W, starting from 0 through 1, corresponds to about infinite (C_B/C_W=0 to 0.1) to 11% (C_B/C_W=0.9 to 1.0) increase in C_B at each step. Hence, the change in T_R^* is significantly dampened with respect to the changes in cost parameters, especially in the lower range of the C_B/C_W . In most hospitals, C_W is usually much higher than C_B, e.g., $C_B/C_W \ll 1$, which makes the T_R^* even less sensitive to changes in C_B/C_W.



(a) λ=50 minutes (b) λ=100 minutes (c) λ=150 minutes

Figure 2. Change of $\boldsymbol{T_R^*}$ with the ratio $\boldsymbol{C_B/C_W}$

Setting: ED LOS follows a Weibull distribution with k = 2; U = 0.8 and p=0.7.

Even though the triage staff would request the target ward-bed to be ready for the ED patient in T_R^* time units from the time of triage, the availability of the requested bed at T_R^* is not guaranteed. In other words, T_R^* is the time slot for which a bed is requested but it may not be feasible to make a bed available at that time if T_R^* is earlier than the expected inpatient ward-bed lead-time, T_l.

Therefore, we compare T_R^* to T_l, i.e., if $T_R^* > T_l$, then T_R^* is feasible and optimal. Alternatively, if $T_R^* < T_l$, then T_l is the earliest time slot that the bed is available for the patient. In this case, it is straightforward to show that T_l is the optimal reservation time slot since C_R is a convex function and is monotone and increasing in T_R for $\forall\, T_R > T_R^*$.

As noted earlier, in estimating the costs, we assume that the patient waiting time starts with the completion of the triage instead of starting it at the end of physician's disposition decision. Accounting for the cost of patient waiting for ED physician's disposition, the total expected cost of making an advanced reservation for the patient is expressed as follows:

$$C_{TR} = \begin{cases} C_w \int_0^\infty t\, f(t) d_t + p \int_0^{max(T_R^*, T_l)} (max(T_R^*, T_l) - t) C_w f(t) d_t) U f(t) d_t \\ \qquad\qquad + \int_{max(T_R^*, T_l)}^\infty (t - max(T_R^*, T_l))(C_B + C_w) U f(t) d_t \end{cases} \tag{16}$$

Accounting also for the cost of patient waiting for ED physician's disposition, which is $C_w \int_0^\infty t\, f(t) d_t$, the total expected cost of making no advanced reservation for the patient is as follows:

$$C_{TNR} = C_W \int_0^\infty t\, f(t) d_t + p T_l C_w \tag{17}$$

Overall, in order to make the ward-bed reservation decision for a patient, we compare the estimated costs C_{TNR} and C_{TR} and select the decision that leads to a lower expected cost. Fig. 3 illustrates these different costs in the form of a decision tree, where $T_B = \max\{T_R^*, T_L\}$ is the time when the bed will be ready for the patient upon reservation. Lastly, it can be shown that, as in the case of T_R^* decision, the cost parameters C_B and C_W affect the decision of making a reservation through the parameter $\delta = C_B/C_w + 1$.

2.4 Model Analysis

2.4.1 Estimation of Model Parameters

For a patient with admission probability p, we need to calculate the relative costs reported in (15), (16) and (17) to find the optimal ward-bed reservation decisions. In particular, we need estimates for C_B , C_W, $f(t)$, and T_l. In this section, we discuss how we estimate all these parameters. The values of $\boldsymbol{C_B}$ and $\boldsymbol{C_w}$ can be challenging to estimate and vary generally across different hospitals depending on the organization type, facility size, location, management, etc. For instance, there are significant differences between for-profit facilities and non-profit facilities such as the VA medical centers. A number of earlier studies have estimated these values. For example, Falvo et al. ***(Falvo, Grove et al. 2007)*** reported that without the 10,397 hours of inpatient boarding time to obstruct the normal cycle of bed turnover at a particular ED, the unit could have generated additional median net revenue of $3,960,264. While this estimate, as an opportunity cost, does not account for the costs associated with negative health outcomes from keeping patients waiting and/or the costs associated with patient dissatisfaction, it leads to a lower bound of $\boldsymbol{C_W = \$380.90}$ per hour in the U.S. The actual cost of patient waiting might be much higher when we consider other practical aspects. On the other extreme, e.g., units such as the VA medical centers are now required to transfer the ED patient to a different (VA or private) hospital if there is no bed available in ICU within 2 hours, step down within 2 hours, and medicine ward in 6 hours, and have to pay for their transfer as well as the often higher external treatment cost (fee-based). As for the cost of maintaining a hospital bed, Health Management Associates recently estimated it to be around $400 for a day for the State of Texas state hospitals ***(Health Management Associates 2011)***, suggesting $\boldsymbol{C_B = \$16.67}$ per hour. In the absence of more accurate cost estimates, we employ these optimistic (low) reported estimates for our numerical experiments.

As discussed in Section 3, the decision to make the reservation and the reservation timing decision both depend only on the relative importance of the cost parameters, i.e., C_B/C_w. As illustrated through an example, the sensitivity of T_R^* is significantly dampened for the low C_B/C_w values found in practice, e.g., using the estimates in (Falvo, Grove et al. 2007) and (Health Management Associates 2011), $C_B/C_w = \$16.67/\$380.90 = 0.044$. Since the decision to make a reservation is dichotomous, given the correct decision of making the reservation, the sensitivity of the expected cost is limited by the sensitivity of T_R^* to the parameters C_B and C_w.

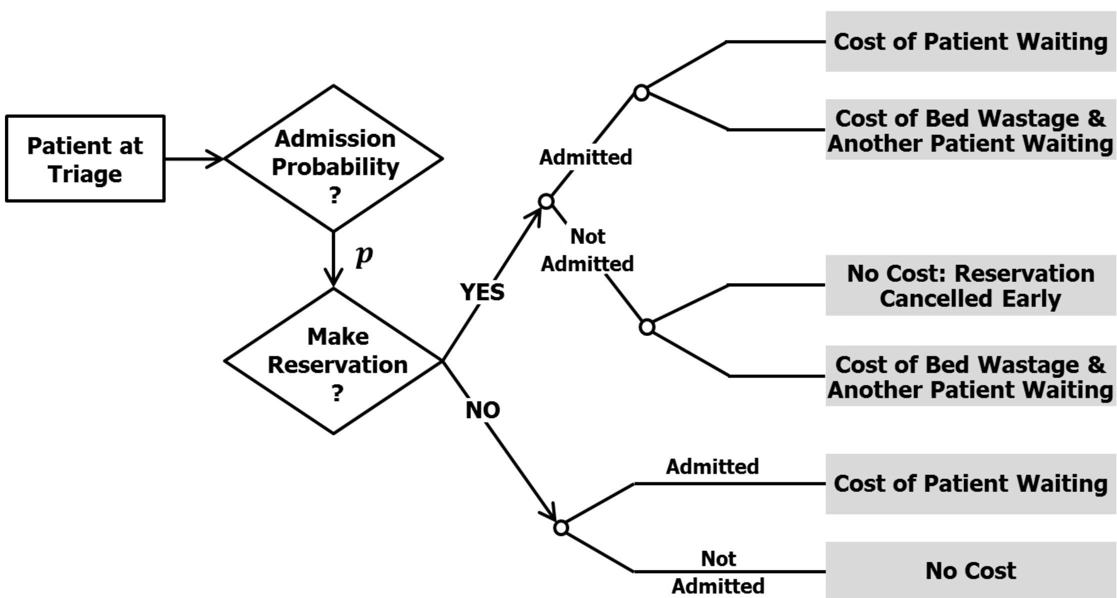


Figure 3. Cost associated with ward-bed reservation decisions and associated decision tree.

To estimate the ED LOS distribution, $f(t)$, one can infer it from the historical patient waiting time data recorded in the ED information systems. Weibull distribution, a versatile and commonly used distribution in the literature for modeling the total ED length of stay, proved to be very effective in fitting our historical data from our VAMC. Our experiments indicate that the calculated costs regarding advance reservations are very sensitive to the scale parameter and not to the shape parameter of the fitted Weibull distribution.

In VA and many other healthcare systems, there exist IT bed management systems (BMS) for real-time tracking of patient movements, status of wards, and bed availability. These systems allow administrative and clinical staff to record manage and report on the planning, patient-movement, patient occupancy, and other activities related to management of beds. At VA, BMS offers the following features needed to estimate the ward-bed lead-time whenever a request for a bed is made: real-time display of patient and bed occupancy status for all beds in the facility; support for emergency management; and reports to facilitate discharge appointments. It also provides information regarding the number of empty beds, number of beds available for female patients, number of beds out of service, and lead-time to clean vacated beds. In addition, it provides reports of the waiting lists for wards, discharges planned for the next day, and any scheduled admissions.

Given the growing availability and adoption of BMS systems in most hospitals and their effective real-time information features, we assume that whenever a bed reservation is being attempted by the ED staff for a target ward, a reliable point estimate for ward-bed lead-time T_l can be obtained rather than relying on the historical data for estimating a lead-time distribution. Extension of the proposed bed reservation model to explicitly account for any uncertainty in the target ward-bed lead-time, T_l, will be the subject of future work. However, in the remainder of this section, we investigate how the variability of ward-bed lead-time T_l affects the expected cost of reservation decisions as well as saving opportunities associated with access to perfect lead-time information. The experiments were conducted with the following ward parameters:

utilization U = 0.85; capacity of m = 30 beds. For the baseline scenario, the expected T_l is estimated to be 53 minutes according to the Kingman's expression (18) for $G/G/m$ queue representation of the ward (using t_e=4 days and ${C_a}^2 + {C_e}^2$ = 0.25) (Kingman 1962):

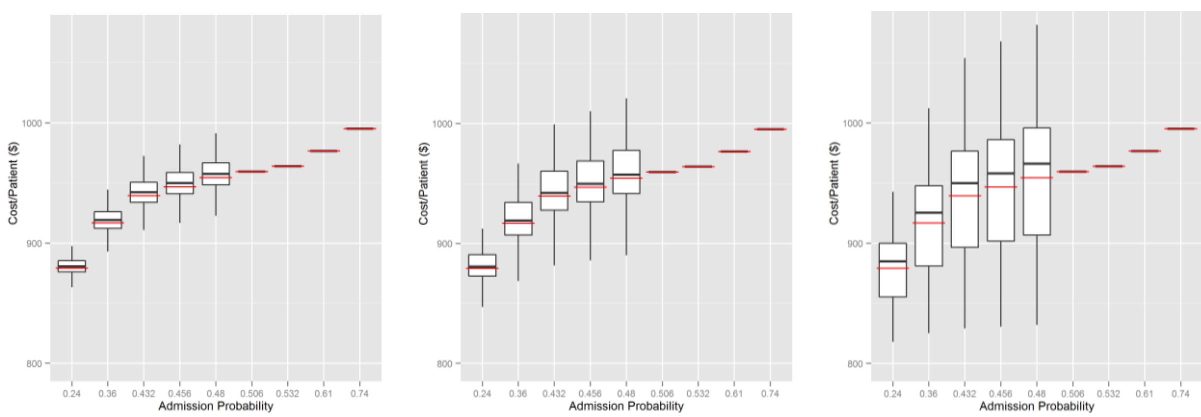
$$E(T_l) \approx \left(\frac{{C_a}^2 + {C_e}^2}{2}\right)\left(\frac{U^{\sqrt{2(m+1)}-1}}{m(1-U)}\right)t_e \quad (18)$$

where C_a is the coefficient of variation of inter-arrival times of patients to the ward, C_e is coefficient of variation of patient's LOS in ward-beds, U is the utilization of the ward, m is the number of beds in the ward (assuming that the staff levels are more than adequate and the bed availability mostly determines the access to inpatient wards access), and t_e is the mean LOS of a patient in the ward. We note that the choice of T_l =53 minutes or its generation process using the Kingman's expression is not restrictive as our goal is to investigate the effect of different levels of variability in T_l on the robustness of the proposed bed reservation model.

In order to assess the sensitivity of the expected costs and potential savings with respect to the uncertainty in T_l, we conducted a series of experiments. In this experimentation, we generated a set of 100 normally distributed lead-times around the mean μ = 53 minutes with varying levels of variance σ^2 ranging from low- to high-degree of variability, i.e., c_v=σ/μ = 0.1, 0.2, and 0.4. In all these experiments, the scale and shape parameters for the ED LOS Weibull distribution are assumed to be $\lambda = 150$ minutes and $k = 2$, respectively.

The sensitivity analysis tracks the robustness of the proposed model in two aspects. Firstly, we investigate the deviation in expected cost when the realized T_l is different from the expected T_l used in determining the reservation decision and timing. In this approach, we maintain the original decisions (whether to make a reservation and its timing) as before. Secondly, we also investigate the potential savings that can be realized if we had perfect information regarding the true ward-bed lead-time T_l. In this approach, we re-optimize both the decision to make a reservation as well as its timing given the perfect knowledge of T_l.

Figure 4 reports the effect of ward-bed lead-time variability on the expected cost from the first approach for different patient admission probabilities. These admission probabilities, on the x-axis, are carefully distributed around the breakeven probability of 0.48, which is the admission probability where the cost of making a reservation or not is same (given the selected ward parameters and ED LOS distribution). The box-plots in (a), (b) and (c) show the quartiles of the cost distributions associated with c_v of 0.1, 0.2, and 0.4, respectively. In each box-plot, the expected cost associated with the expected T_l (i.e., 53 minutes) is highlighted through the red line.

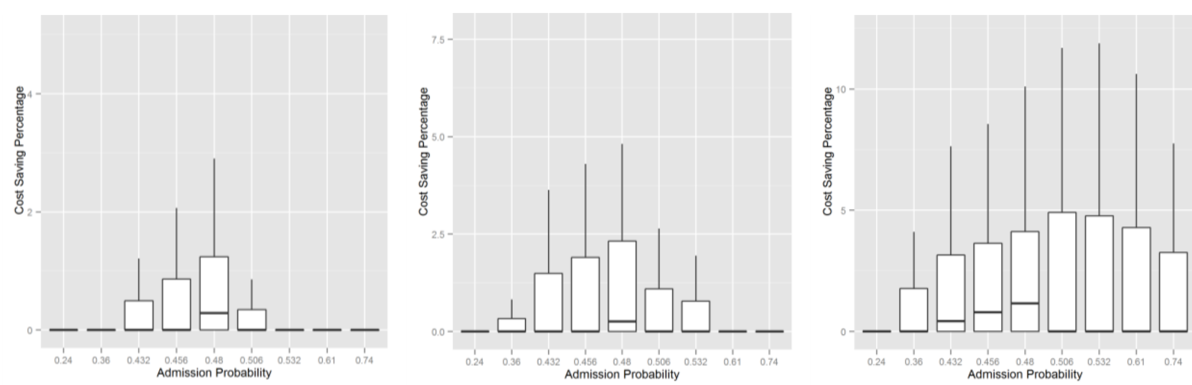


(a) c_v= 0.1 (b) c_v= 0.2 (c) c_v= 0.4

Figure 4. Effect of ward-bed lead-time variability on the expected cost.

In all three levels of the variability, when the admission probability is greater than the breakeven level of 0.48, the expected costs are same for all T_l instances. This is because, given the selected baseline scenario parameters, the optimal reservation time T_R^* is much higher than any of the T_l instances generated. Accordingly, the bed request time, determined by $\max\{T_R^*, T_l\}$, always coincides with T_R^*, making the expected cost independent of T_l. We also note that the median expected cost increases with the lead-time variability. In the extreme case, the predicted expected cost underestimates the median expected cost by at most 1.7% when c_v=0.4. Lastly, we observe that the variability in the expected cost increases with increasing c_v and admission probability.

In the second approach, we calculate the potential savings associated with making the optimal reservation decisions based on perfect information for ward-bed lead-time T_l. We calculate the savings rate as the percent reduction in the expected cost when the reservation decisions are optimized under perfect lead-time information. Figure 4 shows the savings realized under different levels of lead-time variability. The most significant observation is that the median savings are bounded with at most 1.3% across all variability levels. The savings rate increases with the increasing variability in the lead-time. The variance in savings attains its peak around the breakeven admission probability for the predicted lead-time.



(a) c_v= 0.1 (b) c_v= 0.2 (c) c_v= 0.4

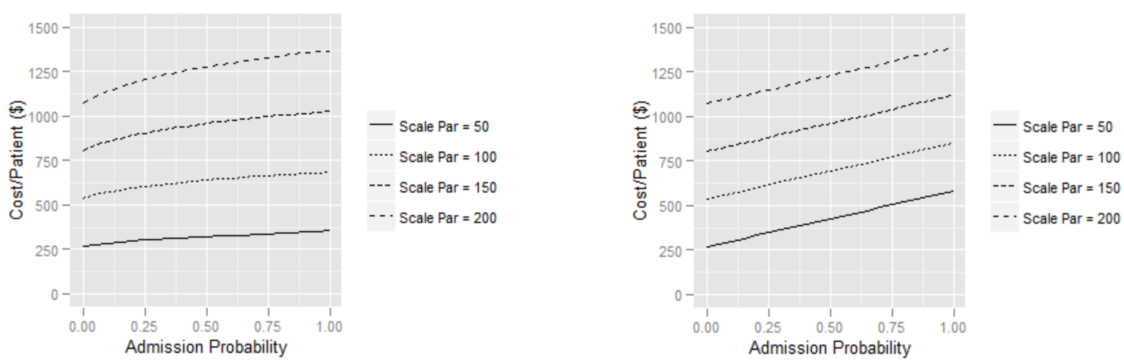
Figure 5. Effect of ward-bed lead-time variability on the expected cost savings opportunity due to perfect ward-bed lead-time information.

2.4.2 Parameter Sensitivity Analysis

In what follows, we report on the results from conducting a number of sensitivity analysis experiments to study the impact of critical model and facility parameter settings on optimal ward-bed reservation decisions.

We first investigate the joint impact of patient admission probability (p) and level of uncertainty regarding patient's LOS from triage to physician's disposition (i.e., the scale parameter of Weibull distribution for t) on the relative costs of making or not making ward-bed reservations, C_{TR} and C_{TNR}, respectively. In conducting these experiments, we use the following estimates for the ward parameters: utilization U = 0.85, capacity of m = 30 beds, average length of patient stay t_e=4 days, and exhibits some variability in arrival and service processes ${C_a}^2 + {C_e}^2$ = 0.25. The shape parameter for the ED LOS Weibull distribution is set to be $k = 2$. As noted earlier, C_W and C_B are assumed to be \$380.90 and \$16.67 per hour, respectively.

Figure 6 illustrates the impact of admission probability and ED LOS distribution's scale parameter on the expected cost of ward-bed reservation decisions. The first observation is that as the scale parameter for t increases, the expected cost per patient increases. This is expected because an increase in scale parameter implies higher variability in the time from triage to physician's disposition, preventing us from effectively selecting the ward-bed reservation slot time (i.e., T_R^*). The second observation is that both C_{TR} and C_{TNR} increase with patient admission probability (p). This is also expected because as p increases, patient needs admission with higher probability and will experience some level of additional waiting due to our inability to facilitate a perfect transition from ED to the target ward. When p = 0, both costs are exactly equal and simply capture the cost associated with patient waiting time from triage to physician's disposition of no admission. At the other extreme, when p = 1, the cost differences are purely attributable to the consequence of a proactive ward-bed reservation vs. a reactive request such that the patient is being admitted with certainty, and hence, the proactive reservation does not suffer from the risk of a false positive reservation. Even when p = 1, the optimal decision will still depend on the levels of the model parameters. When 0 < p < 1, if we choose to make a ward-bed reservation, we have to also account for the costs associated with both false positive and false negative reservations. While the expected cost per patient increases linearly with p when we make no ward-bed reservations, the rate of cost increase decreases with patient admission probability if we choose to make the ward-bed reservation. Overall, optimal rational decision for any given patient admission probability is to select the action that leads to a lower expected cost.



(a) Expected cost of making reservation (b) Expected cost of making no reservation

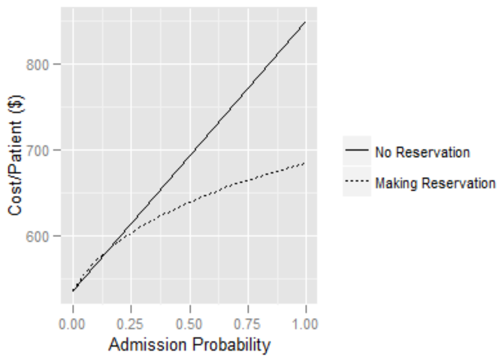
Figure 6. Effect of Weibull ED LOS scale parameter λ on cost of reservations.

(Ward Parameters: U = 0.85; m = 30 beds; t_e=4 days; ${C_a}^2 + {C_e}^2$ = 0.25)

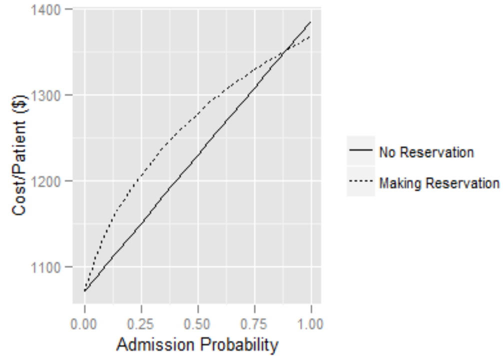
To better illustrate these cost sensitivities, Figure 7 superimposes the C_{TR} and C_{TNR} plots for two relatively extreme settings of the Weibull scale parameter for t. When the scale parameter is low and admission probability is high (e.g., Fig. 7 (a)), there are significant cost savings from making an advanced ward-bed reservation. On the other hand, if there is significant uncertainty in the ED LOS prior to physician's disposition, not only can the savings decrease but can also lead to higher costs under advanced ward-bed reservations. However, irrespective of the model parameter settings, the plots allow us to identify the optimal action that can reduce the expected cost per patient, given the probability of admission. Given the monotone structure of the costs (C_{TR} and C_{TNR}) as a function of patient specific admission probability p, the optimal decision can be reduced to one based on an admission probability reservation threshold (p^*). In Figure 7, when the Weibull distribution scale parameter is 100 minutes, $p^* \approx 0.15$, and it increases to $p^* \approx 0.87$ when the scale parameter goes up to 200 minutes. The threshold is patient specific for estimated ED LOS varies for individual patients based on their triage information and other covariates.

While employing the same ward parameters, Figure 8 illustrates the impact of target-ward utilization on the C_{TR} and C_{TNR}. Here the ED LOS distribution scale parameter is set to be 150 minutes. As expected, with the increasing target-ward utilization, the relative benefit of the advanced reservation increases.

Finally, Figure 9 illustrates the impact of ED LOS distribution scale parameter and target-ward utilization on optimal reservation thresholds (p^*). The two plots show the results for wards with 30 and 40 beds and at different levels of utilization. Obviously, as the ward utilization increases, from 80% to 92.5%, the reservation threshold decreases, making advanced ward-bed reservations more attractive. In addition, as the capacity of the target ward increases, the reservation threshold increases because of risk pooling benefits (Kulkarni, Magazine et al. 2004).



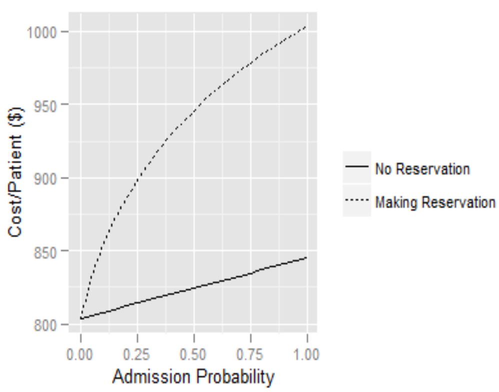
(a) ED LOS Scale parameter: λ=100 minutes



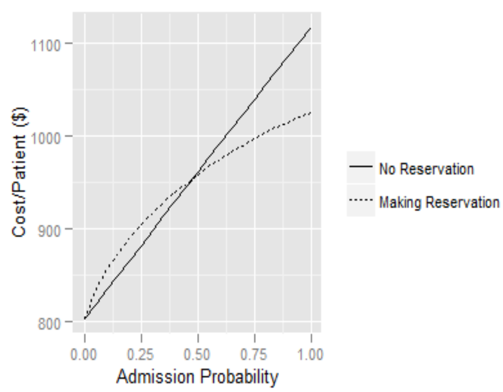
(b) ED LOS Scale parameter: λ=200 minutes

Figure 7. $\boldsymbol{C_{TR}}$ and $\boldsymbol{C_{TNR}}$ cost differences under different levels of uncertainty regarding patient's total waiting time from triage to the physician's disposition decision (i.e., $\boldsymbol{t}$)

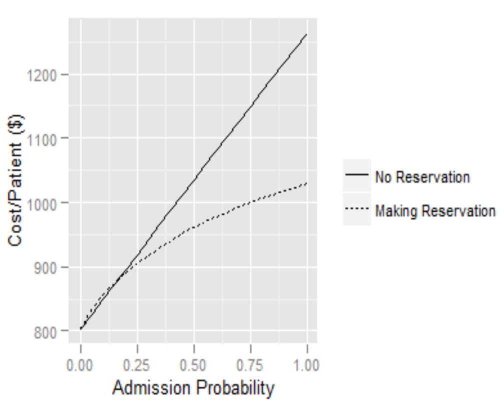
(Ward Parameters: U = 0.85; m = 30 beds; t_e=4 days; ${C_a}^2 + {C_e}^2$ = 0.25)



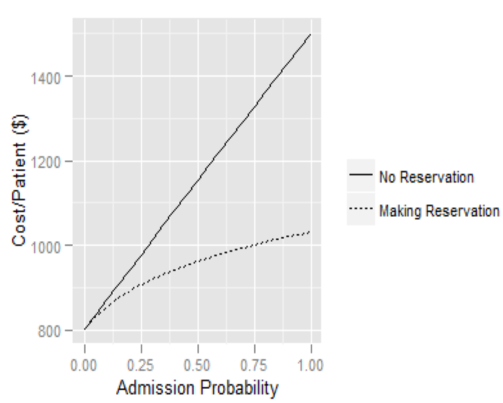
(a) Utilization U = 0.8



(b) Utilization U = 0.85



(c) Utilization U = 0.875

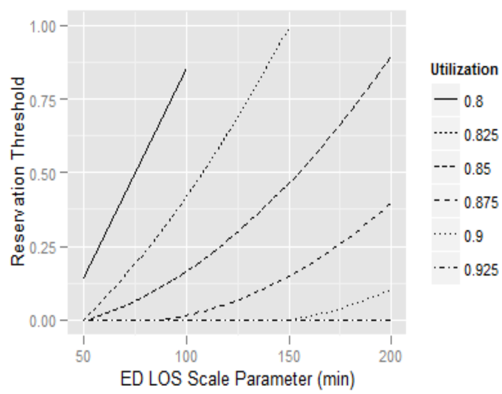


(d) Utilization U = 0.9

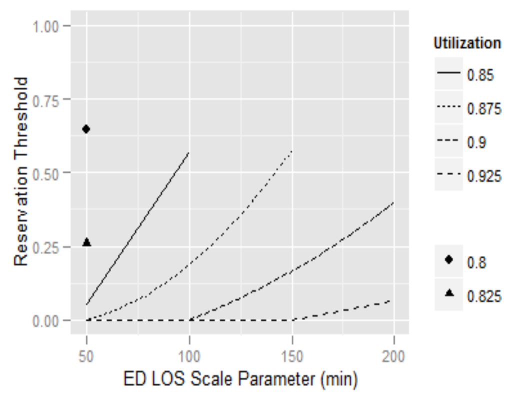
Figure 8. Effect of target-ward utilization on C_TR and C_TNR

(Ward Parameters: m = 30 beds; t_e=4 days; ${C_a}^2 + {C_e}^2$ = 0.25;

ED LOS scale parameter λ= 150 minutes)



(a) Ward beds: 30



(b) Ward beds: 40

Figure 9. Effect of ED LOS scale parameter λ and target-ward utilization U

on reservation threshold

(Ward Parameters: t_e=4 days; ${C_a}^2 + {C_e}^2$ = 0.25)

2.5 Conclusion

Although a number of methods of improving ED patient flow have been proposed in literature, there is no evidence that which one is the best way. This Chapter develops a framework for streamlining ED patient flow. Being different from many proposed methods that involve redesigning the ED system or increasing the capacity, our proposed framework focuses on using near real-time (NRT) information to enhance the "operational intelligence" and in turn promote effective communication between ED and the downstream department to coordinate the patient flow and reduce the ED patient boarding.

The framework is able to take advantage of the existing database sources that many hospitals already have to exploit the information that can be used to improve the efficiency in ED. Machine learning, statistical and operation research methods are all used to develop this framework; specifically, a cost sensitive ward-bed reservation policy based on the admission likelihood prediction of the ED patients is developed. The policy identifies an admission probability as the threshold for making the reservation decision. It also recommends an optimal bed reservation time slot based on a modified news-vender model to minimize the cost of patient waiting and bed wastage. In our framework, each patient’s admission probability is evaluated against an optimal admission probability threshold for the patient. This threshold is determined by trading off the cost associated with making an optimal reservation (e.g., an optimized reservation time slot) with not making a reservation at all. In addition, our proposed approach explicitly accounts for the ED length-of-stay (LOS), which is uncertain, and bed acquisition lead-time reported by target ward bed management system (BMS).

The proposed models are tested using extensive historical data from several mid-west U.S. Department of Veterans Affairs Medical Centers (VAMCs), and this work has been accepted for publication in the journal Health Care Management Science (Qiu, Chinnam et al. 2014).

CHAPTER 3: ADMISSION PREDICTIONS OF EMERGENCY DEPARTMENT PATIENT

3.1 Introduction

As stated in Chapter 2, the proposed cost sensitive ward-bed reservation framework relies on the ability to predict the probability that any ED patient will be admitted using patient information collected and readily available at triage or right after. As discussed in Chapter 2, the triage staff collect information such as patient's complaint, arrival mode, height, weight, age, gender, and vitals (e.g., temperature, blood pressure) to carry out severity assessment of the patient's condition. Using this and other health history information of the ED patient to predict the likelihood of admission leads to a "classification" problem, which is a mainstream research topic in statistics and machine learning.

Given that, patients come to the ED with a wide range of complaints (running into the thousands) and health history backgrounds, the development of an effective patient admission predication classifier can be challenging. While many hospitals in the United States are now maintaining historical electronic medical records (EMR) for the ED patients, pure "data-driven" approaches for building the classifiers from historical records alone might not be effective. Even when the hospital has historical records for years, because of the diversity and range of patient's complaints and backgrounds, it is possible that not enough sample records are available for all complaint categories and disease cases. In addition, other features such as facility characteristics like policies, diversity in staff training/experience, and capacity constraints can introduce additional variability/uncertainty into the admission decision process, and in turn, the burden on the classifier.

All these complexities call for an ED patient admission prediction classifier that is scalable (e.g., can handle a large number of input complaint features/variables and a large number of patient records), able to conduct feature selection to intelligently drop any redundant/irrelevant features from the input, and most importantly, make accurate predictions. In this section, we first review some popular classification methods and then propose an extension to a highly effective probabilistic classifier known as the Relevance Vector Machine (RVM), to achieve integrated feature selection during model training. We conclude the Chapter with results from a variety of experiments on synthetic and real-world datasets.

3.1.1 Introduction of Classification Models

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known (Tan, Steinbach et al. 2005). The goal is that previously unseen records should be assigned a "class" as accurately as possible. Usually, the given dataset is divided into training and testing datasets, with training set used to build the model and test set used to validate it. In another technique, known as cross-validation, the training set is divided into mutually exclusive and equal-sized subsets and for each subset the classifier is trained on the union of all the other subsets. The average of the error rate of each subset is therefore an estimate of the error rate of the classifier.

There are great many techniques developed for classification. In this section, we introduce some the most popular classification methods that have been proposed in the literature. Generally, all techniques fall into one of the following three categories: artificial intelligence based methods, artificial neuron network (ANN) techniques and statistics methods. An excellent review of classification is given in a study by Kotsiantis et al. (Kotsiantis 2007).

Artificial Intelligence based Classification: In artificial intelligence classification area, two main methods are decision trees and rule-based classifiers. Decision tree is a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features (Tan, Steinbach et al. 2005). Decision tree classification works by recursively selecting the best attribute to split the data and expanding the leaf nodes of the tree until the stopping criterion is met. The choice of best split test condition is determined by comparing the impurity of child nodes and also depends on which impurity measurement is used. Rule-based classifiers on the other hand make use of set of IF-THEN rules to classify the instances. The purpose is to construct the smallest rule-set that is consistent with the training data. Usually the procedure is as follows: a separate-and-conquer algorithm (covering algorithms) search for a rule that explains a part of its training instances and separates these instances. The remaining instances are then recursively conquered by learning more rules. The procedure continues until no instances remain. Furnkranz et al. (Furnkranz 1999) provided an excellent overview of existing work in rule-based methods. Note that the decision tree induction can be considered as learning a set of rules simultaneously.

The best characteristic of a rule-based classifier is its comprehensibility. People can easily understand why a decision tree classifies an instance as belonging to a specific class; it is also easy to incorporate the domain knowledge in the classifier. The disadvantages of decision trees include the following (Breiman 1984): (1) Decision-tree causes over fitting easily because learners can create an over-complex trees that do not generalize the data well. (2) The problem of learning an optimal decision tree is NP-complete, so practical decision-tree learning algorithms are based on heuristic algorithms, which cannot guarantee the return of a globally optimal decision tree. (3) There are concepts that are hard to learn because decision trees do not express them easily.

Artificial Neural Networks based Classification: This family of classification methods is based on the notion of a perceptron (Rosenblatt 1962). It includes among others the (single-layered) Perceptron (Littlestone and Warmuth 1994), the Multi-Layered Perceptron (MLP) (Rumelhart, Hinton et al. 1986, Zhang 2000), and the Radial Basis Function (RBF) networks (Robert and Lakhmi 2001).

Single-layered perceptron is the simplest form of an artificial neural network (ANN). It consists of a single neuron with adjustable synaptic weights and bias; during training, the perceptron algorithm converges and positions the decision surface in the form of hyperplane between two classes by adjusting synaptic weights. A single layered perceptron makes its predictions based on a linear predictor function combining a set of weights with the feature vector, and it can only conduct classification with two classes and linearly separable sets of instances.

Multi-layered perceptrons have been developed to tackle non-separable and non-linearly separable learning problems (Rumelhart, Hinton et al. 1986). Multi-layered perceptrons consist of an input layer, flexible number of hidden layers and an output layer; each hidden layer can have a flexible number of neurons (the size of the input and output layers is explicitly dictated by the dimensionality of the dataset). The input layer 'accepts' the vector of predictor variable values and standardizes these values, and then distributes the values to each of the neurons in the hidden layer with a bias. In the hidden layer, the value from each input neuron is multiplied by a weight, and the resulting weighted values are added together producing a combined value, which is fed into a transfer function. The transfer function then outputs a value, which is then distributed to the output layer. The output layer then creates the output of the whole network.

The network training can involve a variety of techniques, often employing some form of a nonlinear optimization technique that involves adjusting the weights on the network arcs to reduce the errors/loss. A study by Zhang et al. (Zhang 2000) provides the details and a good overview of ANNs.

Radial Basis Function (RBF) networks have been also widely applied in many science and engineering fields (Robert and Lakhmi 2001). An RBF network is also an ANN that uses radial basis functions as activation functions (Broomhead and Radar 1988). They are typically three-layer feed-forward networks, and the nodes in the hidden layer implement a set of radial basis functions (e.g. Gaussian functions). The output nodes implement linear summation functions as in an MLP.

The advantage of the ANN based classifier is that it is non-parametric; therefore, this family of methods can handle the classification problems when domain knowledge is lacking and the relationship between input and output is complicated. The networks do not make any assumption regarding the underlying probability density functions or other probabilistic information about the pattern classes under consideration in comparison to other probability based models (Mu-Chun, Woung-Fei et al. 1996). Also, they are adaptive and simple to implement. The disadvantage of this family of classifiers is that there a large number of iterations required for the learning process, so they are usually computationally expensive (Jackson, Beale et al. 1990). In addition, they can also suffer from convergence to local optimal solutions.

Statistical Learning Algorithms: Statistical approaches are characterized by having an explicit underlying probability model, which provides a probability that an instance belongs to any particular class, rather than simply carrying out a classification (Kotsiantis 2007). This family includes among others the Linear Discriminant Analysis (LDA) and the related Fisher's linear discriminant methods (Friedman 1989), Discriminant Correspondence Analysis (Mika, Ratsch et al. 1999), Maximum entropy (Csiszár 1996), the extremely popular Naïve Bayes classifiers (Nilsson 1965), Bayesian networks (Jensen 1996), and Instance-based learning (Mitchell 1997).

LDA and the related Fisher's linear discriminant methods are simple methods used in statistics to find the linear combination of features which best separate two or more classes of objects (Friedman 1989). Instance-based learning (IBL) is a family of learning algorithms that, instead of performing explicit generalization, compares new problem instances with instances seen in training, which have been stored in memory (Russell and Norvig). IBL algorithms assume that similar instances have similar classification, so it classifies instances based on the classification of their most similar neighbors. The most well-known IBL method is the k-NN classification algorithm which looks at the k nearest neighbors of a new instance to decide which class the new instance should belong to. The main advantage of k-NN methods is their simplicity and lack of parametric assumptions. In the presence of a large enough training set, these methods perform surprisingly well, especially when each class is characterized by multiple combinations of predictor values. There are two difficulties with the practical exploitation of the power of the k-NN approach (Shmueli 2010). First, although no time is required to estimate parameters from the training data, the time to find the nearest neighbors in a large training set can be prohibitive. For large training sets, it requires large memory and is slow when making a prediction. Second, prediction accuracy can quickly degrade when number of attributes grows.

Bayesian networks are the most well-known representative of statistical learning algorithms. A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. It consists of a directed acyclic graph of ‘nodes’ and ‘links’ that explains a system. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. The values of the nodes are defined in terms of different, mutually exclusive, ‘states’ (Marcot, Steventon et al. 2006). The relationships between nodes are described by conditional probability distributions that capture the dependences between variables. Bayesian networks have several advantages: (1) they can readily handle incomplete datasets and (2) they can learn the causal relationships between the variables. Construction of prior knowledge is relatively straightforward by constructing “causal” edges between any two factors that are believed to be correlated. Combining domain knowledge and data, they provide an efficient method for preventing the over fitting of data. There are also some limitations to Bayesian networks models. First, while Bayesian models are a useful way to model expert knowledge, it may be difficult to get experts to agree on the structure of the model and the nodes that are important to be included. Furthermore, experts may be challenged to express their knowledge in the form of probability distributions (Pollino, Woodberry et al. 2007, Uusitalo 2007).

Naive Bayesian classifier (NB) is also a Bayesian network model. A naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions, i.e., it assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. The major advantage of the NB classifier is its short computational time for training. In addition, since the model has the form of a product, it can be converted into a sum through the use of logarithms – with significant consequent computational advantages. If a feature is numerical, the usual procedure is to discretize it during data pre-processing (Yang and Webb 2003). Although with strong independence assumption, it often outperforms more sophisticated classifiers. When the number of predictors is very large, it still can achieve high performance even if the assumption of (conditionally) independent predictors is far from true. The disadvantages of NB classifier are as follows: (1) It requires a very large number of records to obtain good results. (2) If a predictor category is not present in the training data, naive Bayes assumes that a new record with that category of the predictor has zero probability. This can be a problem if this rare predictor value is important (Shmueli 2010). The presence of a large training set helps mitigate this effect.

Support Vector Machines and Relevance Vector Machines: Support Vector Machines (SVM) constitute a relatively new class of supervised machine learning techniques and are known to be very effective for classification (Vapnik 1995). It is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. For linear separable classification problem, SVM employs quadratic programming (QP) to find the best classifier boundary hyperplane for global optima. The best classifier boundary hyperplane is called 'the maximum margin hyperplane', which is, among all hyperplanes that separate the classes, the one that gives the greatest separation between the classes. Therefore, the optimal separating hyperplane maximizes the margin of the training data. Support vectors are the observation points which lie exactly on the margin. Regardless of the number of dimensions or size of dataset, the number of support vectors could be as little as two. Finding support vectors and the maximum margin hyperplane belongs to a standard class of optimization problems known as quadratic programming optimization problem (Nocedal and Wright 2006). For classification problems that are not linearly separable, SVM uses a kernel function to implicitly map the model inputs into high-dimensional feature space, where the separation is easier. The kernel trick, which allows the SVM to carry out all the computations in the relatively low-dimensional input space while generating a hyperplane in the high-dimensional feature space, is considered the key feature of SVM. The main advantages of SVM are as follows: (1) It often results in very accurate classifiers. (2) It causes less over fitting and is robust to noise. SVM also has few disadvantages: (1) SVM is intrinsically a binary classifier. To carry out multi-class classification, pair-wise classifications, which we will discuss later, have to be used (e.g., one class against all others, for all classes). (2) It is computationally expensive, and be less practical for the large-scale tasks. (3) It is not easy to incorporate domain knowledge into the classifier. A good review of SVMs can be found in (Burges 1998) and a more recent book (Cristianini and Shawe-Taylor 2000) .

Relevance Vector Machines (RVM) are usually considered an extension of SVM because they too involve the kernel trick of SVM. However, the idea of RVM is different from SVM. Since RVM method is heavily employed in this dissertation, we will review it in detail in a separate section.

Depending on how many classes are involved, a classification problem can be a binary classification (if there are only two classes) problem or a multi-class classification problem (if there are more than two classes). Some classification algorithms naturally permit the use of more than two classes, others are by nature binary algorithms; these can, however, be turned into multinomial classifiers by a variety of strategies (Wu, Lin et al. 2004). Among these strategies one-against-all (Rifkin and Klautau 2004) and one-against-one techniques (Hastie and Tibshirani 1998) are the two most popular such decomposition methods. For a C-class ($C > 2$) problem, one-against-all approach constructs C classifiers, each of which separates one class from all the rest. The ith classifier ($i = 1, \ldots C$) is trained with all the training examples of the ith class with positive labels and all the others with negative labels, and the class which classifies the test datum with the greatest margin is chosen as the final class. One-against-one classification constructs one binary classifier for each pair of the classes to separate members of one class from members of the other; thus $(|C|(|C|-1)/2$ classifiers are built, and a testing sample is assigned to the class that is selected by the most classifiers. One-against-one is usually more suitable since the one-against-all method causes more severe imbalanced problem and it involves high training complexity because the number of training samples is large (Hsu and Lin 2002). However, one-against-one classifiers involve far more computational burden given the need for building a number of sub-classifiers.

3.1.2 Feature Selection in Classification

Feature subset selection is the process of identifying and removing as many irrelevant and redundant features as possible (e.g., in building a classifier) (Yu and Liu 2004). It aims to reduce the dimensionality of the data and enables data mining algorithms to operate faster and more effectively. Feature selection not only makes training and applying a classifier more efficient by decreasing the size of the effective vocabulary but also often increases classification accuracy by eliminating noise features.

Since the possible feature choices and the complexity increases exponentially with the number of features, searching for an optimal feature subset from a high dimensional feature space is an NP-hard problem. Therefore, exhaustive search is impractical for high dimensional data. In this section, we review two most popular feature selection methods: ‘filters’ and ‘wrappers’.

Filtering: The basic idea of filtering is to assign a heuristic score to each feature to filter out the “obviously” useless ones. In the filtering approach, the evaluation of features is independent of the classification algorithm. Some of the popular scoring methods include the Markov blanket and Mutual Information, and a survey of these methods can be found in (Yang and Pedersen 1997). The advantage of filtering is that it is very fast and simple to apply; the disadvantage is that it does not take into account interactions between features (e.g., seemingly useless features can be useful when grouped with others).

Wrapper Methods: Wrapper methods evaluate features using criteria related to the specific classification algorithm. The objective function is a pattern classifier, which evaluates feature subsets by their predictive accuracy on test data by statistical resampling or cross-validation. In wrapper methods, the learning algorithm is a black box, and it is just used to compute objective function, then the search is conducted. A wide range of search strategies can be used, including best-first, branch-and-bound, simulated annealing, genetic algorithms (Kohavi and John 1997). The search procedures are often used based on heuristics but they cannot guarantee the selection of the optimal subset. Wrapper methods generally achieve better accuracy than simple filtering methods since they explicitly consider the interaction between the classifier and the dataset, albeit at a far greater computational burden.

Note that there is another family of methods for reducing dimensionality of the data. This technique is called feature construction/transformation (e.g., PCA or LDA), and these methods construct new features from the basic feature set (Markovitch and Rosenstein 2002) and the newly generated features may lead to the creation of more concise and accurate classifiers. However, this is out of the scope of this research.

An excellent review about the methods of feature selection was given by Guyon et al. (Guyon and Elisseeff 2003).

3.1.3 Evaluation of Classification

A classification model is evaluated by applying it to test data with known target values and comparing the predicted values with the known values. There are many evaluation criteria that can be used to evaluate the performance of a classifier. The basic and most straight forward is a confusion matrix/Table, which displays the number of correct and incorrect predictions made by the model compared with the actual classifications in the test data. Based on the confusion matrix, criteria such as precision, sensitivity (or true positive rate, *TPR*), specificity, accuracy, and *F*-Score can be calculated. A typical confusion Table and associate terminology and performance measures are shown in Table 1.

Table 1 *Confusion Table of a classifier and evaluation measures*

		Predicted Outcome		
		Positive	**Negative**	**Measures:**
Actual Outcome	**Positive**	*TP*: True positives	*FN*: False negatives (Type II Errors)	Sensitivity or True Positive Rate (TPR): $TP/(TP+FN)$
	Negative	*FP*: False positives (Type I Errors)	*TN*: True negatives	Specificity or True Negative Rate: $TN/(FP+TN)$
	Measures:	Precision: $TP/(TP+FP)$ False Positive Rate (FPR): $FPR = 1 - Precision$	Negative Predictive Value: $TN/(TN+FN)$	Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$

In Information Retrieval domain, sensitivity is known as hit rate or recall. F-score can be used as a single measure of performance: $F = 2 \times \frac{Precision \times Recall}{Precision + Recall}$.

Receiver Operating Characteristics (ROC), a coordinate system used for visualizing classifier performance, is another widely used evaluation method. For a ROC, *TPR* is plotted on the Y-axis and *FPR* is plotted on the X-axis, and the area under the ROC curve (AUC) measures the discriminating ability of a binary classification model.

In some classification problems, some types of misclassifications may be considered worse than others. For example, in the cancer prediction problem, predicting a cancer patient to have no cancer is much worse than predicting a non-cancer patient to have cancer. Situations such as this can be addressed by cost-sensitive learning, i.e., taking the cost of every type of error into account, and then the objective becomes one of minimizing the total cost of misclassification.

Since the reservation system proposed in Chapter 2 requires the prediction of probability of admission for the ED patient to different inpatient wards, we need the classifier to be able to conduct multi-class classification and output probabilistic results. For this purpose, we employ Relevance Vector Machines (RVM) for carrying out the classification task. In addition, to further improve the performance of the classifier, we extend the standard multi-class RVM to integrally consider feature selection during learning. By doing so, we also eliminate the need for employing filter or wrapper techniques for building a highly effective and compact classifier.

3.2 Literature Review

In this section, we review the studies that conduct ED patient admission prediction. We also review the work that is related to the RVM since it is closely related to our proposed ED patient admission model.

3.2.1 Prediction of ED Patient Admission

Recently published literature offers a number of classification models to predict hospital admissions of ED patients: Leegon et al. (Leegon, Jones et al. 2005, Leegon and Aronsky 2006) propose a Bayesian network approach, Li et al. (Li, Guo et al. 2009, Li, Guo et al. 2012) incorporate semantic information of medical terms, and Sun et al. (Sun, Heng et al. 2011) apply a logistic regression model. Peck et al. (Peck, Benneyan et al. 2012) discussed the concept of a pull system in the hospital and used discrete event simulation to study how it may benefit ED patient flow. They show that early prediction of ED patient admissions does indeed have the ability to improve the patient flow and reduce the effects of non-value added delays. Stover-Baker et al. (Stover-Baker, Stahlman et al. 2012) conducted a prospective study to determine if an ED nurse can determine at triage if a patient will be admitted to an inpatient unit and concluded that triage nurses demonstrated a high sensitivity and specificity in admission prediction and could potentially save many hours in requesting an inpatient bed. Xie et al. (Xie 2013) validated that proportional hazard (PH) and logistic regression models can be used to provide reasonably accurate prediction of hospital admission for ED patients, with the PH model offering more accurate predictions. Cameron et al. (Cameron, Rodgers et al. 2014) created a 6-variable core (triage category, age, National Early Warning Score (NEWS), arrival by ambulance, referral source, and recent admission) core to estimate the probability of admission at triage. Although the accuracies of the proposed ED patient admission prediction models are promising, most of these models attempt "binary" predictions by classifying each patient as going to be admitted by the hospital or not upon patient's triage. None of these previous works conducted multi-class classification to identify the target inpatient ward. In addition, there are no probabilistic results in terms of the target ward admission.

3.2.2 Multi-class Relevance Vector Machine (mRVM)

Original RVM Model: In supervised learning, a set of input vectors $\{\boldsymbol{x_n}\}_{n=1}^{N}$ along with corresponding targets $\{t_n\}_{n=1}^{N}$ are given to us (the targets $\{t_n\}_{n=1}^{N}$ are real values in regression and class labels in classification), and we wish to learn the dependency of the target in the input and then predict target t for previously unseen $\boldsymbol{x}$.

Predictions are usually based on assuming a parametric function $y(\boldsymbol{x})$ that is defined over the input space, and then a learning process being conducted to infer the parameters of this function. A popular and widely used function for $y(x)$ is the linear model:

$$y(\boldsymbol{x};\boldsymbol{w}) = \sum_{i=1}^{N} w_i K\,(\boldsymbol{x},\boldsymbol{x_i}) + w_0 \tag{19}$$

where $K(\boldsymbol{x},\boldsymbol{x_i})$ is a kernel function effectively defining one basis function for each data point in the training set, $\boldsymbol{w} = \left(w_1, w_{2,} \dots, w_N\right)^T$ are the weights reflecting the importance of the training data points.

RVM model is originally proposed by Tipping(Tipping (2001)). This method employs a Bayesian framework to create probabilistic results for class membership and to make the model sparse. In (Tipping (2001)), a standard probabilistic formulation is followed and the targets are assumed to be samples from the model with additive noise:

$$t_n = y(\boldsymbol{x_n};\boldsymbol{w}) + \varepsilon_n\,, \tag{20}$$

where the function $y(\boldsymbol{x})$ is defined as in (19), and ε_n are independent samples from some noise process, further assumed to follow a Gaussian distribution with zero mean and variance σ^2:

$$p\big(\varepsilon_n|\sigma^2)\big) = \boldsymbol{N}(\varepsilon_n|0,\sigma^2) \tag{21}$$

Therefore, $p\big(t_n|\sigma^2)\big) = \boldsymbol{N}(t_n|y(\boldsymbol{x_n}),\sigma^2)$**,** and the complete dataset can be written as:

$$p(\boldsymbol{t}|\boldsymbol{w},\sigma^2\,) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left|\left|\boldsymbol{t}-\boldsymbol{Kw}\right|\right|^2\right\}, \tag{22}$$

where $\boldsymbol{t} = (t_1, \dots t_N)^T$, $\boldsymbol{w} = (w_0, \dots w_N)^T$ and $\boldsymbol{K}$ is the $N \times (N+1)$ matrix with $\boldsymbol{K} = [\boldsymbol{K}(\boldsymbol{x}_1), \boldsymbol{K}(\boldsymbol{x}_2), \dots, \boldsymbol{K}(\boldsymbol{x}_N)]^T$**,** where $\boldsymbol{K}(\boldsymbol{x}_n) = [1, K(\boldsymbol{x}_n,\boldsymbol{x}_1), K(\boldsymbol{x}_n,\boldsymbol{x}_2), \dots, K(\boldsymbol{x}_n,\boldsymbol{x}_N)]^T$.

Maximum likelihood estimation of parameters $\boldsymbol{w}$ and σ^2 in (22) would lead to severe over-fitting because there are as many parameters in the model as training examples. Therefore, additional constraints have to be imposed on the parameters, and RVM adopts the standard Bayesian methodology of defining a prior probability distribution over the parameters $\boldsymbol{w}$:

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{i=0}^{N} N(w_i \big| 0, \alpha_i^{-1}), \tag{23}$$

where $\boldsymbol{\alpha}$ a vector of $N + 1$ hyperparameters, and there is an individual hyperparameter 'controlling' every weight.

The introduction of an individual hyperparameter for every weight is considered as the key feature of the RVM formulation (Majumder 2005, Tipping (2001)) because it makes the goal of sparsity to be possible to achieve; by defining a Gamma distribution hyperprior over $\boldsymbol{\alpha}$, i.e.,

$$p(\boldsymbol{\alpha}) = \prod_{i=0}^{N} Gamma(\alpha_i|a, b) \tag{24}$$

and enforcing the parameters a and b to small values, the RVM makes the hyperprior non-informative (flat). Therefore, many of the α_i are driven to very large values, and thus the standard deviation of the $\boldsymbol{w}$ Gaussian distribution would be very small and the posterior probability of the associated weights would be concentrated at zero, implying that the corresponding model weights w_i can be effectively pruned out. Those training vectors that are associated with non-zero weights are called 'relevance vectors'. They capture the data's underlying distribution and represent 'prototypical' examples of respective classes.

Based on defined prior, the posterior of the weight is obtained using the Bayes rule:

$$p(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{-(N+1)}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}|\boldsymbol{w}-\boldsymbol{\mu}|^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})\right\}, \tag{25}$$

where the posterior covariance and the mean are as following two equations, respectively:

$$\boldsymbol{\Sigma} = (\boldsymbol{K}^T \boldsymbol{B} \boldsymbol{K} + \boldsymbol{A})^{-1} \tag{26}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \boldsymbol{K}^T \boldsymbol{B} \boldsymbol{t} \tag{27}$$

with $\boldsymbol{A} = diag(\alpha_{0,} \alpha_1, \dots, \alpha_N)$ and $\boldsymbol{B} = \sigma^{-2} \boldsymbol{I_N}$.

By integrating out parameters $\boldsymbol{w}$, the marginal likelihood can be obtained:

$$p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{-(N+1)}{2}} |\mathbf{B}^{-1} + \boldsymbol{K}\boldsymbol{A}^{-1}\boldsymbol{K}^T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{t}^T (\mathbf{B}^{-1} + \boldsymbol{K}\boldsymbol{A}^{-1}\boldsymbol{K}^T)^{-1}\boldsymbol{t}\right\}, \tag{28}$$

and then we maximize (28) with respect to $\boldsymbol{\alpha}$ and σ^2.

The values of $\boldsymbol{\alpha}$ and σ^2 which maximize the marginal likelihood cannot be obtained in closed form, so they have to be iteratively re-estimated until some suitable convergence criteria, i.e., many of the α_i tend to infinity, have been satisfied. In (Tipping (2001)), for $\boldsymbol{\alpha}$, differentiation of (28), equating to zero and rearranging, following the approach of MacKay (MacKay 1992), gives:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2}, \tag{29}$$

where μ_i is the i-th posterior mean weight from (27) and γ_i is defined by:

$$\gamma_i = 1 - \alpha_i \Sigma_{\mathrm{ii}}, \tag{30}$$

with Σ_{ii} the i-th diagonal element of the posterior weight covariance from (26) computed from current $\boldsymbol{\alpha}$ and σ^2 values.

For the noise variance σ^2, differentiation leads to the re-estimate:

$$(\sigma^2)^{new} = \frac{||t - \emptyset\mu||^2}{N - \Sigma_{\mathrm{i}} \gamma_i}, \tag{31}$$

where the 'N' in the denominator refers to the number of data examples.

The prediction is then made based on the posterior distribution over the weights, conditioned on the maximizing value α_{MP} and σ_{MP}^2, i.e. , for a new datum $\boldsymbol{x}_*$, the predictive distribution is calculated as following using (25):

$$p(\boldsymbol{t}_*|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = \int p(t_*|\boldsymbol{w}, \sigma_{MP}^2) p(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d_{\boldsymbol{w}} \,. \tag{32}$$

Since both terms in the integrand are Gaussian, this is readily computed, giving:

$$p(\boldsymbol{t}_*|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = N(t_*|y_*, \sigma_*^2),$$

with

$$y_* = (\boldsymbol{\mu}^T \boldsymbol{\phi}(\boldsymbol{x}_*), \tag{33}$$

$$\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\phi}(\boldsymbol{x}_*)^T \Sigma \boldsymbol{\phi}(\boldsymbol{x}_*) \,. \tag{34}$$

Extension of RVM to mRVM: The work in (Tipping (2001)) focuses on regression problem and binary classification problem, and multiclass classification problem was briefly mentioned without details. Extension of RVM to multiclass classification problem was investigated recently by Psorakis et al. (Psorakis 2010).

The notation of mRVM is slightly different from RVM: for a C–class classification problem $(C \geq 2)$ with N training data points, $t_n \in \{1, .. C\}$ $(n = 1, ... N)$ represent the class labels of the training data points; $\mathbf{W} \in \mathcal{R}^{N \times C}$ is the weight matrix with the c-th column $\boldsymbol{w}_c$ representing the weights of the training data points to the class c $(c = 1, ... C)$; $\mathbf{K} \in \mathcal{R}^{N \times N}$ is the kernel matrix with the n-th row $\mathbf{k}_n = K(\boldsymbol{x}_n, \boldsymbol{x}_i)(i = 1, ... N)$ being the basis function for n-th training data point $(i = 1, ... N)$.

To achieve multi-class discrimination, auxiliary variables $Y \in R^{C\times N}$ are introduced. These auxiliary variables act as the regression targets of $\mathbf{W}^{\mathrm{T}}\boldsymbol{K}$ following a standardized model, and thus:

$$y_{nc}|\boldsymbol{w}_c, \mathbf{k}_n = \boldsymbol{N}_{y_{cn}}(\boldsymbol{w}_c^T\mathbf{k}_n, 1). \tag{35}$$

The relationship between the auxiliary variables and the class label is expressed by the following equation, which results in a standard multinomial probit function:

$$\mathrm{t}_n = i \; if \; y_{ni} > y_{nj} \; \forall \, j \neq i \; (i, j = 1, \dots, C) \tag{36}$$

which means that a sample n belongs to the i class if the value y_{ni} has the highest value of all elements in the n-th column.

After the auxiliary variables being introduced, the posterior class membership distribution is derived using the multinomial probit likelihood function and the method proposed in (T. Damoulas 2008):

$$p(t_n = i|\boldsymbol{W}, \boldsymbol{k}_n) = \xi_{p(u)}\left\{\prod_{j\neq i} \emptyset(u + (\boldsymbol{w}_i - \boldsymbol{w}_j)^T\boldsymbol{k}_n)\right\}, \tag{37}$$

where $u \sim N(0,1)$ and $\emptyset$ a Gaussian cumulative distribution function.

Based on the prior probability distributions over the weight parameters $\boldsymbol{w}$ and their hyperparameters $\boldsymbol{\alpha}$ defined by RVM, mRVM extends the equations (23) and (24) to the following two equations that are the corresponding probability distributions for the multi-class problem:

$$p(\boldsymbol{W}|\boldsymbol{A}) = \prod_{n=1}^{N}\prod_{c=1}^{C} N(0, \alpha_{nc}^{-1}) \tag{38}$$

$$p(\boldsymbol{A}|\tau, \upsilon) = \prod_{n=1}^{N}\prod_{c=1}^{C} Gamma(\tau, \upsilon) \tag{39}$$

where $\mathbf{A} \in \mathcal{R}^{\mathrm{N}\times\mathrm{C}}$ with elements α_{nc}^{-1} representing the variance of the distribution of the element w_{nc} in $\boldsymbol{W}$. Being similar to what is done in RVM, with sufficiently small hyper-parameters τ and v, the regression coefficients $\boldsymbol{W}$ posterior distribution is restricted around zero and a sparse solution can thus be achieved.

The training procedure of mRVM involves consecutive updates of the model parameters using an Expectation Maximization (E-M) procedure. The parameters $\boldsymbol{W}$ are updated based on the following equation:

$$\widehat{\boldsymbol{w}_c} = (\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{A}_c)^{-1}\boldsymbol{K}\boldsymbol{y}_c^T \tag{40}$$

where $\boldsymbol{A}_c$ is a diagonal matrix derived from the c-th column of $\mathbf{A}$.

Given a class i, $\boldsymbol{Y}$ are updated based on the following equation:

If $c \neq i$ $(c = 1, \ldots C)$, then

$$\widetilde{y_{nc}} \leftarrow \boldsymbol{k}_n\widehat{\boldsymbol{w}_c} - \frac{\xi_{p(u)}\left\{N_u(\boldsymbol{k}_n\widehat{\boldsymbol{w}_c} - \boldsymbol{k}_n\widehat{\boldsymbol{w}_\iota}, 1)\emptyset_u^{n,i,c}\right\}}{\xi_{p(u)}\left\{\emptyset(\boldsymbol{u} + \boldsymbol{k}_n\widehat{\boldsymbol{w}_\iota} - \boldsymbol{k}_n\widehat{\boldsymbol{w}_c}, 1)\emptyset_u^{n,i,c}\right\}}. \tag{41}$$

If $c = i$, then

$$\widetilde{y_{n\iota}} \leftarrow \boldsymbol{k}_n\widehat{\boldsymbol{w}_\iota} - \left(\sum_{j\neq i}\widetilde{y_{n\iota}} - \boldsymbol{k}_n\widehat{\boldsymbol{w}_J}\right). \tag{42}$$

The parameters $\mathbf{A}$ are updated using the following equations (Psorakis 2010):

$$\widehat{\alpha_{nc}} = \frac{2\tau + 1}{w_{nc}^2 + 2v}. \tag{43}$$

The learning process of mRVM (Psorakis 2010) follows a standard E-M scheme, subsequent updating the parameters from the above four equations until appropriate stop criteria is satisfied, except that the samples which are considered insignificant are explicitly removed. In other words, for an i sample, if $\alpha_{ic} > 10^5 \; \forall \, c \in \{1, \dots C\}$, then it is removed.

Note that the mRVM we discussed is mRVM2 in (Psorakis 2010). In fact, Tipping proposed an alternative way of construct the sparsity of RVM in (Tipping 2003), and (Psorakis 2010) extended it to a multi-class setting as well and called the extension mRVM1. However, mRVM1 is out of this paper's scope. Interested reader can find the details in (Psorakis 2010).

Since estimation of (37) cannot be computed analytically, a quadrature approximation approach was employed in (Psorakis 2010) to estimate it:

$$p(t_n = i|\boldsymbol{W}, \boldsymbol{k}_n) = \xi_{p(u)}\{F(u)\} = \frac{1}{\sqrt{2\pi}} \int F(u) e^{-u^2} \, d_u, \qquad (44)$$

where $u \sim N(0, 1)$ and the e^{-u^2} is the standard Guassian-Hermite weight function $W(x)$.

3.3 Integrated Feature Selection mRVM: IF-mRVM

As we stated earlier, due to the large number of features present for ED admission prediction problems, it would be advisable to implement feature selection to produce a parsimonious model. In this section, we propose an algorithm that integrates mRVM and feature selection, labeled integrated feature selection mRVM or IF-mRVM. Filtering based feature selection methods employ a sequential approach to feature selection and subsequent model building and the computationally tedious wrapper techniques build a prohibitively larger number of models with a variety of feature combinations for yielding a compact model with good classification accuracy. The proposed approach, in contrast, is a truly integrated efficient method that builds just one classifier and integrally eliminates any irrelevant or redundant features without compromising classification performance.

Suppose that the number of features of the dataset is d, then training data points are $\boldsymbol{x}_n = \{x_{n1}, x_{n2}, \dots x_{nd}\}$ $(n = 1, \dots N)$. The mechanism of IF-mRVM is that we impose a scale parameter to each feature of the dataset, estimate all these parameters, and remove those features with scale parameters smaller than a predefined threshold.

We denote the scale parameter corresponding to i-th $(i = 1, \dots N)$ feature of the c-th $(c = 1, \dots C)$ class by η_i^c, and then we impose the parameters to a input vectors $\boldsymbol{x}_n(n \in \{1, \dots N\})$) in the form $\{\eta_1^c x_{n1}, \eta_2^c x_{n2}, \dots \eta_2^c x_{nd}\}$. The kernel matrix for the c-th class is $\boldsymbol{K}_c \in \mathcal{R}^{\mathrm{N}\times\mathrm{N}}$ with the element n-th row $\boldsymbol{K}_{mn}^c = K_c(\boldsymbol{x}_n, \boldsymbol{x}_m, \boldsymbol{\eta}_c)(i = 1, \dots N)$ being the basis function between the n-th m-th training data point, and it takes the following form:

$$\boldsymbol{K}_c(\boldsymbol{x}_m, \boldsymbol{x}_n, \boldsymbol{\eta}_c) = K\big((\eta_1^c x_{m1}, \eta_2^c x_{m2}, \dots \eta_d^c x_{md}), (\eta_1^c x_{n1}, \eta_2^c x_{n2}, \dots \eta_d^c x_{nd})\big), \quad (45)$$

where $\boldsymbol{\eta}_c = \{\eta_1^c, \eta_2^c, \dots \eta_d^c\}$. Note that in RVM and mRVM, there is only one kernel matrix $\boldsymbol{K}$ because the numbers 'within' the kernel calculations are always the elements of the input vectors and do not change with the update of $\boldsymbol{W}, \boldsymbol{Y}$ and $\boldsymbol{A}$. However, as we can see from (45), in our kernel calculation, we use a unique set of feature parameters for each class because we assume that the importance of each feature to different classes could be different. Therefore, the number of kernel matrix in IF-mRVM is the same as the number of class.

Accordingly, we revise the log of marginal likelihood

$$\mathcal{L} = \sum_{c=1}^{C} -\frac{1}{2}[Nlog2\pi + \log|\boldsymbol{C}| + \boldsymbol{y}_c^T \boldsymbol{C}^{-1} \boldsymbol{y}^c], \quad (46)$$

which is given in (Psorakis 2010), to be in the following form:

$$\mathcal{L} = \sum_{c=1}^{C} -\frac{1}{2}[Nlog2\pi + \log|\boldsymbol{C}_c| + \boldsymbol{y}_c^T {\boldsymbol{C}_c}^{-1} \boldsymbol{y}^c]. \quad (47)$$

In (46),

$$\boldsymbol{C} = \boldsymbol{I} + \boldsymbol{K}^T \boldsymbol{A}^{-1} \boldsymbol{K}\,, \tag{48}$$

and we modify it to

$$\boldsymbol{C}_c = \boldsymbol{I} + \boldsymbol{K}_c{}^T \boldsymbol{A}_c{}^{-1} \boldsymbol{K}_c\,. \tag{49}$$

The gradient of the likelihood $\mathcal{L}$ with respect to η_i^c is then

$$\frac{\partial \mathcal{L}}{\partial \eta_i^c} = \sum_{m=1}^{N} \sum_{n=1}^{N} \frac{\partial \mathcal{L}}{\partial \boldsymbol{K}_{mn}^c} \frac{\partial \boldsymbol{K}_{mn}^c}{\partial \eta_i^c} \tag{50}$$

The first term in (50) is independent of the kernel function parameters, and we collect all the terms into a matrix $\boldsymbol{D}^c$ such that $\boldsymbol{D}_{mn}^c = \partial \mathcal{L} / \partial K_{mn}^c$, and from (47) and (49), we can obtain that

$$\boldsymbol{D}^c = (\boldsymbol{C}_c{}^{-1} \boldsymbol{y}^c \boldsymbol{y}_c^T \boldsymbol{C}_c{}^{-1} - \boldsymbol{C}_c{}^{-1}) \boldsymbol{K}_c \boldsymbol{A}_c{}^{-1}. \tag{51}$$

Regarding the kernel function, while the algorithm can incorporate a variety of kernel functions, we employ a radial basis function kernel (RBF), in particular the Gaussian kernel, which is very popular and widely used by support vector machines and RVMs and satisfies the requirements for the kernel function trick. Two other popular kernel functions are the Fisher kernel (Jaakkola 1998) and the polynomial kernel. The calculation of Fisher kernel involves a so called fisher score and fisher information matrix, so $\frac{\partial \boldsymbol{K}_{mn}^c}{\partial \eta_i^c}$ is not to be able to be calculated analytically. Polynomial kernel is defined as $K(\boldsymbol{x}_m, \boldsymbol{x}_n) = (\boldsymbol{x}_m^T \boldsymbol{x}_n + c)^d$ where $c \geq 0$ is a constant trading off the influence of higher-order versus lower-order terms in the polynomial and d is the degree of the polynomial kernel. The complexity of the calculation of $\frac{\partial \boldsymbol{K}_{mn}^c}{\partial \eta_i^c}$ depends on the degree of the polynomial kernel. One problem with the polynomial kernel is that it may suffer from numerical instability: when $\boldsymbol{x}_m^T\boldsymbol{x}_n + c < 1$, $K(\boldsymbol{x}_m, \boldsymbol{x}_n)$ tends to zero as the degree is increased, whereas when $\boldsymbol{x}_m^T\boldsymbol{x}_n + c > 1$, $K(\boldsymbol{x}_m, \boldsymbol{x}_n)a$ tends to infinity (Lin 2012).

In proposing the IF-mRVM, we recommend the Gaussian kernel function and add a scale parameter η_i^c to the i-th feature when calculating the kernel of the c-th class.

For a Gaussian kernel function, we have

$$\boldsymbol{K}_{mn}^c = e^{\{-\sum_{i=1}^d \eta_i^c (x_{mi} - x_{ni})^2\}} \tag{52}$$

Therefore, we have that

$$\frac{\partial \boldsymbol{K}_{mn}^c}{\partial \eta_i^c} = -\boldsymbol{K}_{mn}^c (x_{mi} - x_{ni})^2. \tag{53}$$

Thus

$$\frac{\partial \mathcal{L}}{\partial \eta_i^c} = \sum_{m=1}^N \sum_{n=1}^N -\boldsymbol{D}_{mn}^c \boldsymbol{K}_{mn}^c (x_{mi} - x_{ni})^2 \tag{54}$$

We then use the following equation to update $\boldsymbol{\eta}_c$ $(c = 1, \dots C)$:

$$\boldsymbol{\eta}_c^{(n+1)} = \boldsymbol{\eta}_c^{(n)} + step\ size * f(\nabla \mathcal{L}(\boldsymbol{\eta}_c)), \tag{55}$$

where

$$f(\nabla \mathcal{L}(\boldsymbol{\eta}_c)) = \{f(\frac{\partial \mathcal{L}}{\partial \eta_1^c}), f(\frac{\partial \mathcal{L}}{\partial \eta_2^c}), \dots f(\frac{\partial \mathcal{L}}{\partial \eta_d^c})\},$$

and $\frac{\partial \mathcal{L}}{\partial \eta_i^c}$ $(i = 1, \dots d)$ is the partial differentiation of $\mathcal{L}$ respecting to η_i^c.

Now we describe how we defined the function f. Let $\nabla\mathcal{L}(\boldsymbol{\eta}_c)^+$ denote the set of all positive $\frac{\partial\mathcal{L}}{\partial\eta_i^c}$ $(i = 1, \dots d)$ and $\nabla\mathcal{L}(\boldsymbol{\eta}_c)^-$ denotes the set of all negative $\frac{\partial\mathcal{L}}{\partial\eta_i^c}$ $(i = 1, \dots d)$, then we define $f\left(\frac{\partial\mathcal{L}}{\partial\eta_i^c}\right)$ as following:

$$f\left(\frac{\partial\mathcal{L}}{\partial\eta_i^c}\right) = \begin{cases} \frac{\left(\frac{\partial\mathcal{L}}{\partial\eta_i^c}\right) - \min\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^+\}}{\max\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^+ - \min\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^+} & if\ \frac{\partial\mathcal{L}}{\partial\eta_i^c} \in \nabla\mathcal{L}(\boldsymbol{\eta}_c)^+ \\ \frac{-\left(\frac{\partial\mathcal{L}}{\partial\eta_i^c}\right) + \min\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^-\}}{\max\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^- - \min\{\nabla\mathcal{L}(\boldsymbol{\eta}_c)^-} - 1 & if\ \frac{\partial\mathcal{L}}{\partial\eta_i^c} \in \nabla\mathcal{L}(\boldsymbol{\eta}_c)^- \end{cases} \tag{56}$$

n means the n-th iteration, and $step\ size$ is a predefined real number to adjust the speed of the change of $\boldsymbol{\eta}_c$. If a feature i has a parameter η_i^c that is smaller than a predefined threshold for $\forall\ c \in \{1, \dots C\}$, we remove it. We keep updating the $\boldsymbol{\eta}$ until the algorithm converges.

For a new observation $\boldsymbol{x}_*$, the probability of its class label c is predicted using the following equation, which is based on a modification of (37):

$$p(t_* = c|\ \boldsymbol{x}_*, \boldsymbol{W}, \mathbf{k}_*^c\ ) = \xi_{p(u)}\left\{\prod_{j\neq c} \emptyset(u + (\boldsymbol{w}_c - \boldsymbol{w}_j)^T\mathbf{k}_*^c)\right\}, \tag{57}$$

where

$$\mathbf{k}_*^c = K^c(\boldsymbol{x}_*, \boldsymbol{x}_n) = K\big((\eta_1^c x_{*1}, \eta_2^c x_{*2}, \dots \eta_d^c x_{*d}), (\eta_1^c x_{n1}, \eta_2^c x_{n2}, \dots \eta_d^c x_{nd})\big)\ (n = 1, \dots N)$$

where $\eta_i^c(i = 1, \dots d)$ is the value of the parameter of feature i corresponding to class c at the end of algorithm training. Therefore, for those removed features, this parameter is 0.

3.4 Experiments

3.4.1 Performance of IF- mRVM on Benchmark Datasets

In this section, we discuss the results from applying the proposed algorithm on a range of benchmark datasets, which are all from real world problems. Our source of the datasets is the University of California Irvine (UCI) Machine Learning Repository. These datasets were also used in paper (Psorakis 2010), and we show the comparison of the results of IF-mRVM to be baseline mRVM. Table 2 shows the characteristics of the datasets used in our experiments. In Table 2, N is the sample size; C is the number of classes; D is the number of features of each sample.

Table 2 *Datasets used in the experiments*

Dataset	***N***	***C***	***D***
Breast Cancer	569	2	9
Ecoli	336	8	7
Iris	150	3	4
Wine	178	3	13
Soybean	47	4	35

In addition to the above original datasets, we created artificial datasets based on these datasets to check the ability of IF-mRVM to remove purely redundant/irrelevant noise features. The artificial datasets are created in the following way: we added D and $2D$ pure noise features (sampled from the standard Gaussian distribution) to each of the above original datasets. For instance, for Breast Cancer dataset, we work on the original dataset, which has 30 features as well as the following two artificial datasets: the dataset with 30 original features plus 30 pure dummy features and with 30 original features plus 60 pure noise features.

For each of the above datasets, we replicated both mRVM and IF-mRVM 500 times. For each run, we randomly split the data to 60% training set and 40% testing set. Initially, all elements in matrixes $\boldsymbol{W}$ and $\boldsymbol{\eta}$ are assigned to be 1, and we update $\boldsymbol{W}, \mathbf{A}, \mathbf{Y}$ and $\boldsymbol{\eta}$ alternatively. The step size in equation (55) is 0.1, and the threshold of dropping a feature varies across the datasets. In Table 3, and Table 4, we show the results of 500 runs of IF-mRVM and mRVM, respectively. During each run, we keep track of the number of removed features, number of relevance vectors, as well as the computation time. These experiments were conducted on a Windows 7 (64-bit) PC with two Intel Xeon(R) CPU processors (2.4GHz) and 8GB of RAM.

Table 3 *Results of IF-mRVM*

	Original Dataset	Original Dataset + 1D pure noise features	Original Dataset + 2D pure noise features
Soybean (TH=0.99)	0.9897	0.9713	0.9401
Iris (TH=0.8)	0.9440	0.9172	0.9137
Breast Cancer (TH=0.65)	0.9637	0.9640	0.9639
Winery (TH=0.9)	0.9624	0.9708	0.9672
Ecoli (TH=0.9)	0.9034	0.9162	0.9170

(a) Average classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	1.0000	1.0000	0.9474
Iris (TH=0.8)	0.9500	0.9333	0.9333
Breast Cancer (TH=0.65)	0.9634	0.9634	0.9634
Winery (TH=0.9)	0.9718	0.9718	0.9718
Ecoli (TH=0.9)	0.9104	0.9179	0.9179

(b) Median classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.0366	0.0477	0.0644
Iris (TH=0.8)	0.0332	0.0534	0.0588
Breast Cancer (TH=0.65)	0.0109	0.0108	0.0112
Winery (TH=0.9)	0.9104	0.0212	0.0237
Ecoli (TH=0.9)	0.0351	0.0302	0.0311

(c) Standard deviation of classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	3.996	4.046	4.246
Iris (TH=0.8)	4.704	5.226	5.846
Iris (TH=0.85)	5.084	5.194	5.886
Breast Cancer (TH=0.65)	7.610	5.522	8.124
Winery (TH=0.9)	4.888	6.222	6.766

(d) Number of relevance vectors

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	14.86	16.52	17.04
Iris (TH=0.8)	3.60	3.71	3.48
Iris (TH=0.85)	3.29	3.47	3.34
Breast Cancer (TH=0.65)	7.00	7.38	7.14
Winery (TH=0.9)	10.39	9.69	8.66

(e) Number of real features active at termination

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.00	2.84	7.60
Iris (TH=0.8)	0.00	1.43	1.84
Iris (TH=0.85)	0.00	0.67	1.10
Breast Cancer (TH=0.65)	0.00	2.69	3.02
Winery (TH=0.9)	0.00	0.12	0.06

(f) Number of dummy features active at termination

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	1.05	1.65	2.41
Iris (TH=0.8)	7.12	11.57	13.26
Iris (TH=0.85)	6.63	10.00	12.41
Breast Cancer (TH=0.65)	41.89	73.24	96.78
Winery (TH=0.9)	8.29	9.87	9.84

(g) Computation time

Table 4 *Results of Baseline mRVM*

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.9626	0.8604	0.7659
Iris (TH=0.8)	0.9485	0.8486	0.7954
Iris (TH=0.85)	0.9480	0.8459	0.7963
Breast Cancer (TH=0.65)	0.9684	0.9651	0.9637
Winery (TH=0.9)	0.9499	0.9149	0.8852
Ecoli (TH=0.9)	0.9245	0.8836	0.8534

(a) Average classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	1.0000	0.8947	0.7895
Iris (TH=0.8)	0.9500	0.8500	0.8000
Iris (TH=0.85)	0.9500	0.8500	0.8000
Breast Cancer (TH=0.65)	0.9670	0.9634	0.9634
Winery (TH=0.9)	0.9577	0.9155	0.8873

(b) Median classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.0621	0.0927	0.1113
Iris (TH=0.8)	0.0265	0.0529	0.0558
Iris (TH=0.85)	0.0269	0.0524	0.0542
Breast Cancer (TH=0.65)	0.0091	0.0093	0.0100
Winery (TH=0.9)	0.0274	0.0365	0.0396

(c) Standard deviation of classification accuracy

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	4.124	4.222	4.234
Iris (TH=0.8)	5.512	6.996	7.554
Iris (TH=0.85)	5.492	6.926	7.692
Breast Cancer (TH=0.65)	4.952	6.998	8.106
Winery (TH=0.9)	4.594	7.086	8.262

(d) Number of relevance vectors

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	20.99	21.00	21.00
Iris (TH=0.8)	4.00	4.00	4.00
Iris (TH=0.85)	4.00	4.00	4.00
Breast Cancer (TH=0.65)	9.00	9.00	9.00
Winery (TH=0.9)	13.00	13.00	13.00

(e) Number of real features active at termination

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.00	21.00	42.00
Iris (TH=0.8)	0.00	4.00	8.00
Iris (TH=0.85)	0.00	4.00	8.00

Breast Cancer (TH=0.65)	0.00	9.00	18.00
Winery (TH=0.9)	0.00	13.00	26.00

(f) Number of dummy features active at termination

	Original Dataset	1D pure noise features	2D pure noise features
Soybean (TH=0.99)	0.75	1.42	1.79
Iris (TH=0.8)	2.70	2.87	3.05
Iris (TH=0.85)	2.70	2.76	2.94
Breast Cancer (TH=0.65)	11.24	12.17	13.26
Winery (TH=0.9)	3.19	4.16	4.56

(g) Computation time

From Tables 4 (a) and (b), it is relatively clear that mRVM is highly affected by the presence of pure Gaussian noise features. The greater the number of noise features the higher the degradation in classification accuracy. On the contrary, the degradation in the performance of IF-mRVM is relatively minimum, if any. For example, in the case of Breast Cancer, Winery, and Ecoli datasets, there is no degradation in classification accuracy with IF-mRVM.

In the absence of any artificial pure Gaussian noise features (i.e., the original dataset), for one of the of the five datasets (Soybean), IF-mRVM yields more than 2% improvement in accuracy over mRVM; for another dataset (Ecoli), IF-mRVM decreases the mean accuracy by around 2%

over mRVM. For all other datasets, the performance of IF-mRVM and mRVM are about the same with difference generally less than 0.5%.

In terms of the standard deviation in the classification accuracy performance of the two methods, IF-mRVM reduces the standard deviation for soybean and vinery datasets, but increases for the other three datasets.

Sub-Table (d) in the two Table sets show the average number of relevance vectors present in the corresponding models. We see that IF-mRVM yields fewer relevance vectors than mRVM for Soybean, Iris, and Ecoli datasets, but has many more relevance vectors for Breast Cancer datasets and slightly more relevance vectors for Ecoli datasets.

Sub-tables (e) and (f) in both tables show the average number of real and pure noise features active (i.e., not discarded) at termination. These tables show that IF-mRVM is able to drop around 10% to 30% of the real features in the original datasets and is able to drop up to 99% of the pure noise features, without compromising much the classification accuracy. mRVM has no feature selection component, and hence, retains all the features present. One might wonder why the number of features listed in Table 3, and in particular Table 4, differ from the number of features listed in Table 2. As noted earlier, the datasets are randomly split into the training (60%) and testing (40%) datasets for building the models. Given that many of these datasets are relatively sparse (without much redundancy), partitioning yields many features that are non-discriminative (meaning feature values are constants for all the training records). These features are removed before building any RVM model. Hence, the difference in the number of features reported. Given that mRVM does not have any feature deletion feature, differences between Table 4 (e) and Table 2 are completely attributable to this pre-processing activity.

Lastly, sub-Table (g) of the two tables sets compare the computation time of the two methods. As expected, IF-mRVM is somewhat more computationally expensive than the baseline mRVM method. However, this increase in computation time is relatively minimal, if one were to compare the results to computation times that would be expected from a wrapper type approach to feature selection.

Overall, we see from the two Table sets that IF-mRVM can effectively drop the pure noise features as well as the non-important/redundant original features in the datasets. In addition, IF-mRVM can significantly improve the classification accuracy (at least in the presence of pure Gaussian noise features). One would expect these performance characteristics to be retained when IF-mRVM method is applied to other real-world datasets with noisy and redundant/irrelevant features.

3.4.2 Performance of IF- mRVM on Emergency Department Data

In this section, we discuss the results from applying IF-mRVM to the emergency department data from a collaborating VA Medical Center (VAMC) for making admission predictions of ED patients.

Data: VAMC employs a number of information systems to collect, store, and analyze patient flow data. The relevant system for this study is the ED Integration Software (EDIS) that incorporates several web-based systems to help healthcare professionals track and manage the flow of patient care (Technology 2010). The records of 7,532 patients that visited the VAMC sometime between July 25, 2011 to November 8, 2011 were analyzed in this work. In EDIS, each patient record included the following basic information: gender, age, complaint (as free text), acuity level (ESI index), and physician’s final disposition. EDIS also includes the following time stamps for each patient: time in (time at which ED check in is entered for patient), time out (time at which facility closed patient's ED visit), triage (elapsed time between patient's time in and initial acuity assessment), disposition (the patient's disposition), and admission time (time of inpatient ward admission).

Data Processing: In building our classification model, patient's age, gender, complaint, acuity and time in were used as covariate factors in our experiments. In our analysis, patient's age was discretized mostly by decades (< 30 years, 30-40, 40-50, ….., 80-90, >90 years) into 8 binary indicator variables and time-in was categorized into six four-hour time segments (e.g., Noon-4PM, 4-8PM …) as has been done in the literature (Sun, Heng et al. 2011, Peck, Benneyan et al. 2012). Primary patient complaint in the data was captured by the ED triage nurse and entered into the EDIS as free text. To handle the complaint text, we first remove all "stop" words (i.e., for, the, a etc.). Among the remaining words, we only retained words with a frequency ≥ 0.05% per patient record (i.e., the word should appear in at least one in 2,000 patient records on the average) to be used as a potential feature for the patient admission prediction model. We ended up retaining **588** unique complaint text words in our input. Including features derived from age, gender, time-in, ESI level, and complaint text words, the total number of features in the dataset came to be **590**. The models were expected to differentiate patients sent home from patients admitted to three wards (medicine, psychiatry, and surgery), resulting in four target classes.

We have built RVM models using both mRVM and IF-mRVM methods on the dataset, and the dataset was randomly split into 60% training data and 40% testing data. Given the size of the dataset and the relatively significant model building times, the runs were replicated five times to evaluate consistency.

Results: The mean accuracy of the original mRVM across all five replications is 88.4%. In terms of the accuracy of IF-mRVM, when the threshold is as high as 0.98, 210 features were discarded, and the mean accuracy is 83.9%; when the threshold is 0.97, 112 features were discarded, and the mean accuracy is 86.5%; when the threshold is 0.94, 32 features were discarded, and the mean accuracy is 87.15%; when the threshold is 0.925, 9 features were discarded, and the mean accuracy is 89.2%; When the threshold is 0.91, IF-mRVM did not discarded any features, and thus, there is no difference in the performance between mRVM and IF-mRVM. Therefore, IF-mRVM achieves the best balance of discarded features and maintaining accuracy.

3.5 Conclusion

Prediction of patient admissions during ED triage generates actionable information that can be exchanged between ED, inpatient wards, and different hospital departments. Timely and accurate predictions coupled with a reservation management system can help reduce the ED boarding times, improve patient flow, and reduce overcrowding. This Chapter proposes a novel multi-class classification method, based on a Bayesian framework, to predict the admission likelihood of ED patients. In particular, we extend the highly effective multi-class RVM classification technique to simultaneously consider feature selection. We also offer complete mathematical details for the implementation of the integrated feature selection mRVM, employing Gaussian kernels. Techniques for improving robustness of feature selection that exploit patterns within gradient histories and standardizing the gradients are also presented.

From the standpoint of classifiers, the proposed IF-mRVM has been shown to very effectively discard pure noise features when added to a variety of benchmark datasets, without compromising the classification accuracy. Application of IF-mRVM has also led to slight improvement in the performance of ED patient admission predictions.

From the standpoint of ED patient admission predictions, the proposed method overcomes several shortcomings of existing prediction methods: 1) All the models in the literature focus on binary admission predictions; they do not predict the target admission ward for the patient, 2) Majority of models in the literature do not offer probabilistic admission predictions (with the exception of the naïve Bayes classifiers), 3) None of the models exploit free-language complaint text collected during triage for admission predictions, and 4) Finally, the proposed method offers integrated feature selection to improve the parsimony of the admission prediction model.

There are several limitations to the current study. These limitations and their resolutions as part of future work are described next. First, as stated earlier, the ED data is highly imbalanced between different classes, but the classifier is unable to effectively address the imbalance. An approach like Synthetic Minority Over-sampling Technique (SMOTE) could be integrated into the method. An introduction to SMOTE can be found in a work by Chawla, Nitesh et al. (Chawla, Bowyer et al. 2002). Second, subject matter experts at the VAMC have indicated that additional patient information such as health history and frequency/outcomes of past ED visits are key factors affecting admission likelihood and target wards. Incorporation of this and other available information from the VA's CPRS system is part of a future study.

CHAPTER 4: SURVIVAL ANALYSIS MODELS FOR ESTIMATING LENGTH-OF STAY IN EMERGENCY DEPARTMENTS

4.1 Introduction

As stated in Chapter 2, our proposed approach for bed reservations explicitly accounts for the uncertainty associated with patient's ED length-of-stay (LOS). In this Chapter, we discuss how ED LOS can be effectively estimated. Before we discuss our methods, we introduce the definition of ED LOS and review the extant literature related to modeling of ED LOS.

Generally, ED LOS can be broken down into three distinct periods: waiting room time, treatment time, and boarding time (waiting for an inpatient bed). Waiting room time refers to the time elapsed between registration and initial contact with the physician. Treatment time refers to the time elapsed between initial contact with the physician and the disposition decision of the physician. Boarding time is the time elapsed between the admission disposition decision of the physician and the time of ED discharge (Ding, McCarthy et al. 2010). The lengths of these three stages depend on patient related factors, ED facility factors, and other hospital departments' factors. For instance, a study by Ding et al. (Ding, McCarthy et al. 2010) reports that acuity level and chief complaint were important predictors of all phases of care. Patients with psychiatric problems experienced the longest treatment times. Injured patients did not wait as long for an ED or inpatient bed. Patients who arrived by ambulance had shorter wait times but longer treatment times compared to those who did not. There are, of course, many other factors that affect the total ED LOS. Therefore, the ED LOS data naturally exhibits high variation and uncertainty.

In this work, when ED LOS refers to the time between the triage to the disposition decision of the physician, i.e, it is the sum of the first two phases in ED. The reason for doing this is that, in our proposed decision support framework, the model needs an estimate of this to consider ward-bed reservations in advance. Although we only focus on the length of the first two phases, this data remains highly varied. For instance, during the ED treatment process, the physician orders any necessary laboratory tests and imaging. The patient has to wait for the results and possibly interact with the physician several times during the process. The utilization, staff levels, and schedules of the labs, significantly affect the waiting time.

The information necessary for accurate estimation of these waiting segment lengths might not be readily available in the EDIS systems in real time. For example, the length of an imaging queue. Therefore, ED LOS is not only an ED issue, but also an inter-department issue, making it even harder to estimate the ED LOS using any one model. In this work, we investigate how effective survival analysis models are in estimating the ED LOS by including the near real-time information that is available at triage or right after as model covariates. Since survival analysis involves the modeling of time to event data, the ED patient's discharge time epoch is considered the "event" for the analysis.

In the remainder of this Chapter, we review the literature related to ED LOS estimation, and then discuss related survival analysis models. We then estimate the ED LOS using state-of-the-art survival analysis models and discuss the experimental results of ED data from a mid-west VA Medical Center.

4.2 Literature Review

Three streams of literature are closely related to our ED LOS estimation work. First stream focuses on identifying the factors affecting ED LOS, another investigates methods for modeling of data that is right skewed and long tailed (typical of ED LOS data), and the third stream focuses on survival analysis models.

4.2.1 Literature: Factors Affecting ED LOS

In terms of identifying the factors affecting ED LOS, a study by Wiler et al. (Wiler, Handel et al. 2012) indicated that LOS increased on days with higher percentage of daily admissions, higher elopements, higher periods of ambulance diversion, and during weekdays, whereas LOS decreased on days with higher numbers of discharges and weekends. Casalino et al. (Casalino, Wargon et al. 2013) conducted a prospective multicenter study evaluating the impact of age, patient's clinical acuity and complexity, and care pathways. Kocher et al. (Kocher, Meurer et al. 2012) assessed the contribution of testing and treatment to LOS, also stratified by disposition. Rathlev et al. (Rathlev, Chessare et al. 2007) measured the effect of various input, throughput, and output factors and concluded that hospital occupancy and the number of ED admissions are associated with ED LOS.

4.2.2 Literature: Statistical Modeling of Skewed & Long-Tailed Data

Another body of literature that is related to our work investigates the statistical modeling of skewed and long tailed data. Faddy et al. (Faddy, Graves et al. 2009) presented a phase-type method for modeling LOS data and assess the role of covariates. Arazzip et al. (Marazzi, Paccaud et al. 1998) assessed the adequacy of three widely used models - Lognormal, Weibull, and Gamma - for modeling LOS. Most recently, Gardiner (Gardiner 2013) demonstrated the application of several parametric and hazard rate models for fitting heavy tailed data: accelerated failure time (AFT) model, mixed proportional hazards model, and Coxian phase-type distribution model. Benefiting from all this prior work, we investigate hazard rate models that incorporate most of the factors identified above as well as some new covariates (in particular, lab work and medical imaging orders that might be ordered for the ED patient prior to physician’s disposition, ED patient census levels, staff levels etc.) and rely on historical data (for fitting the models) combined with near real-time data to estimate the ED LOS for individual patients.

4.2.3 Attempts to Estimate Parts of ED LOS

We found a few attempts to estimate parts of ED LOS. Ding et al. (Ding, McCarthy et al. 2010) used multivariate quantile regression model to conduct a retrospective cohort study to characterize the 10th, 50th, and 90th percentile of ED LOS using demographic, clinical and temporal characteristics in order to better inform patients and ED staff. Sun et al. (Sun, Teow et al. 2012) developed and validated a quantile regression model that predicts an individual patient's median and 95th percentile waiting time to be seen (i.e., time from triage completion to start time of emergency physician consultation) by using only data available at triage.

4.3 Survival Analysis Models

Survival analysis models are a family of statistical modeling approaches that deal with analysis of time to events. Suppose that the length of time to an event is represented by the random variable T, with a continuous probability distribution $f(t)$, then the cumulative distribution function, $F(t)$, for continuous distribution, is the probability that a random variable will have a survival time less than some stated value t.

Survival function is stated as $(t\ ) = 1 - F(t) = P(T \geq t)$, i.e., the probability of observing a survival time greater than or equal to some stated value t. In our case, T is the length of time elapsed from triage to physician's disposition, and is measured in minutes.

Hazard models predict LOS by investigating the probability of discharge at time $t + dt$ given that patient stayed until time t. The hazard function, $\lambda(t, x)$, is defined as the rate of failure (e.g., patient is discharged from ED) at a point in time t given survival (e.g., stay in ED) until that time:

$$\lambda(t, \boldsymbol{x}) = lim_{dt \to 0} \frac{p(t \leq T \leq t + dt | T \geq t, x)}{dt} \qquad (58)$$

where $\boldsymbol{x}$ is a vector of explanatory variables consisting of patient and hospital characteristics.

The relationship between the density function, the survival function, and the hazard function is expressed in the following two equations:

$$\lambda(t, \boldsymbol{x}) = \frac{f(t,\boldsymbol{x})}{S(t,\boldsymbol{x})} \tag{59}$$

$$\lambda(t, \boldsymbol{x}) = -ln\, S(t, \boldsymbol{x}). \tag{60}$$

Therefore, once we obtain an estimate for one of these three functions, we can estimate the other two through the above equations. The remainder of this section introduces the two popular types of hazard models: semi-parametric proportional hazard rate models and parametric hazard rate models.

4.3.1 Semi-parametric Proportional Hazard Model

The semi-parametric proportional hazard model was proposed by Cox (Cox 1972). He suggested a (partial) likelihood procedure to estimate the hazard function of a multiplicative and proportional form:

$$\text{Cox PH: } \lambda(t) = \lambda_0(t)e^{(\boldsymbol{x}'\boldsymbol{\beta})}\,, \tag{61}$$

where $\lambda_0(t)$ is an arbitrary unspecified base-line hazard function that specifies a continuous distribution, $\boldsymbol{\theta}$ is a vector of ancillary parameters characterizing the distribution of T, and $\boldsymbol{\beta}$ is a vector of unknown coefficients associated with the covariate vector x. The Cox proportional hazard (PH) model is the most commonly used semi-parametric duration model. It is easy to implement and has been reported to be robust in survival analysis (Buchman, Kubos et al. 1994). However, since the partial likelihood approach discards information regarding actual failure times and uses only their rank order, the efficiency of the estimates obtained by this approach is reduced.

4.3.2 Parametric Hazard Rate Models

Instead of using an unspecified function $\lambda_0(t)$ as in the Cox PH model, the parametric hazard rate models require restrictive assumptions regarding the functional form of the baseline hazard function $\lambda_0(t)$. In this work, we employ the Weibull distribution as the baseline function. The formulation of the model is as follows:

$$\text{Weibull: } \lambda(t) = p\lambda_0^p t^{p-1} e^{(\boldsymbol{x}'\boldsymbol{\beta})}\text{, where } p > 0. \tag{62}$$

Weibull distribution is a generalization of the exponential distribution and suitable for modeling data with monotonic hazard rates that either increase or decrease exponentially with time.

It should be noted that in this hazard model, $\boldsymbol{\beta}$ represents the effects of increases in $\boldsymbol{x}$ on the conditional probability of a termination of a stay, whereas in the standard regression analysis, β measures the effect of increases in $\boldsymbol{x}$ on the length of stay. The model parameter p is usually estimated by maximum likelihood estimation (MLE). See (Lee and Wang 2013) for a good introduction of hazard rate models for survival analysis.

4.3.3 Neural Network Models for Survival Analysis

In recent years, machine learning methods, particularly artificial neural networks (ANNs), are being widely used for survival analysis. Specifically, De Laurentiis et al. (De Laurentiis and Ravdin 1994) suggested situations in which ANNs are better than Cox's regression model: the relationship of variables to the outcome is complex and unknown. Several approaches have been proposed to employ ANNs for survival analysis.

Some studies treated the time interval as an input variable. The original vector is transformed into a set of data vectors, one for each possible follow-up time. Before the event time, the target value is set to 0, and to 1 at the time of event and all subsequent intervals (Biganzoli, Boracchi et al. 1998). Some other studies extended the Cox PH model (Faraggi and Simon 1995, Eleuteri, Tagliaferri et al. 2003, Ripley, Harris et al. 2004) using an ANN to allow non-linear predictors to be fitted implicitly and the effect of the covariates to vary over time. To do so, these models suggest replacing the linear predictor $\boldsymbol{x}'\boldsymbol{\beta}$ in Cox-PH with a neural network non-linear function.

Another family of models directly predicts the survival rate or the hazard rate, i.e., they set the survival rate or hazard rate of a subject as the target of the neural network. The outputs of networks were proved to be the survival or hazard probability. To do so, time is discretized to intervals and the hazard rate of each interval are estimated (Ravdin and Clark 1992, Liestbl, Andersen et al. 1994, Brown, Branford et al. 1997, Biganzoli, Boracchi et al. 1998).

4.4 Survival Analysis Models for Estimating ED LOS

In this section, we discuss the results of retrospectively estimating ED LOS for 28,809 patient records at a VAMC.

4.4.1 Data

In addition to the data from EDIS discussed in Chapter 3, the experiments in this Chapter incorporate more information from another VA database called Computerized Patient Record System (CPRS). In addition, we conducted experiments on more patient records than in Chapter 3. The records of 28,809 patients that visited the VAMC from July 25, 2011 to September 30, 2012 were analyzed in this work. While EDSI information remains similar to that described in Chapter 3, the following additional information was extracted from CPRS for each patient: triage vitals (Pulse, Temperature, Respiration Rate) and ordered items (includes list of laboratory tests and medical imaging the patients have undergone while in ED).

4.4.2 Models

We employed both non-parametric and parametric hazard models as well as ANN survival analysis models to incorporate the covariates into the analysis. The non-parametric approach is the Cox PH model and the parametric hazard approach is the model using Weibull distribution for the baseline hazard. Note that Weibull is the only continuous distribution with both the Accelerated Failure Time (AFT) and proportional hazard forms (Gardiner 2013), and in this analysis, we used the former since Cox model is already based on the proportional hazard assumption.

In terms of ANN survival analysis models, we estimate the complete survival curve and the hazard rate. To make the system tractable for ANNs, we take the approach proposed in (Brown, Branford et al. 1997): we discretize the elapsed time in units of width Δt. The inputs for the ANN are the patient's inputs and the jth output is the estimated hazard at time $j\Delta t$. During the model training process, the observed ED LOS is turned into a target hazard rate as follows: if a patient i leaves ED at time t_i , then

$$h_j^i = \begin{cases} 0 & if\ j < j_{out} \\ 1 & if\ j = j_{out} \end{cases}$$

where j_{out} is the smallest value of j such that $t_i < t_j$. There is no constraint on the hazard components at later times. When the network is being trained, the error at any output node presented with an undefined empirical hazard is set to zero, preventing the undefined hazards from updating the network weights. Choosing the structure of the ANN (e.g., number of hidden layers in a multi-layer perceptron, number of hidden nodes per hidden layer, activation function), associated training algorithms, and optimal discretization of time for survival analysis, is an active research area and it is beyond the scope of this work.

The ANN used in our experiments is the multi-layer perceptron (MLP) with a single hidden layer and sigmoidal activation function, and it was batch trained using a conjugate gradient-descent algorithm. We varied the number of hidden layer noes from 5 to 50 in steps of 5 and 20 hidden nodes worked best. In terms of discretizing ED LOS time, given the long right tail observed in the data (see Figure 10), the following scheme worked rather well: 0-330 minutes was segmented into 10 minutes intervals, 330-600 minutes into half-hour intervals; 600-1440 minutes into 1-hour intervals, and all times exceeding 1440 minutes were binned into one interval.

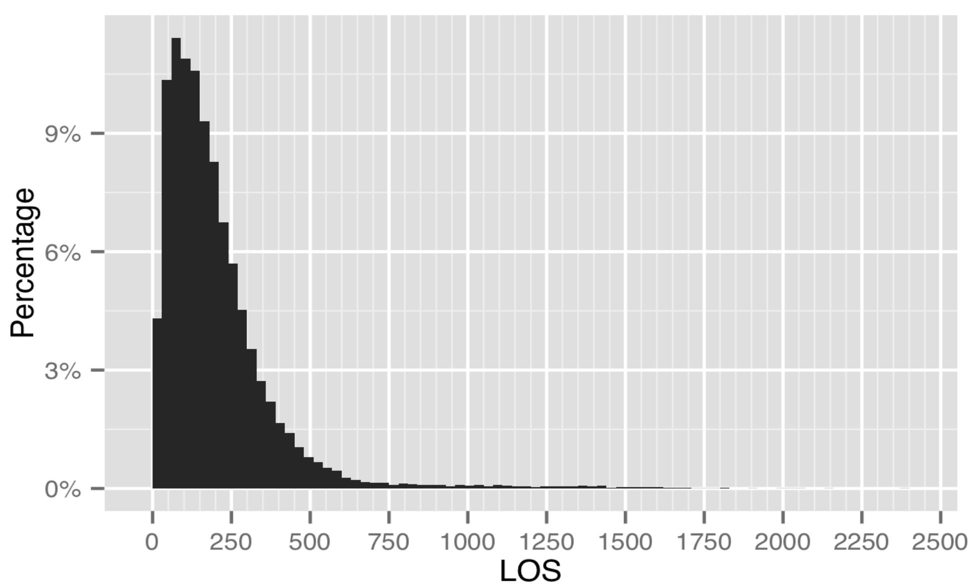


Figure 10. Histogram of ED LOS of the patients in our data

4.4.2 Experiments & Results

In all our experiments, we have split the data into 60% training data and 40% testing data. The experiments were done in R software version 3.0.2 .

In the experiments, we first included the following information that is collected in triage as the covariates: patient age, gender, patient acuity level, the crowding level and staff level in ED upon the patient's arrival, and patient triage vitals such as pulse, temperature, and respiration rate. All factors other than crowding level and the staff level are obtained directly from the VAMC dataset. Crowding level in our work is defined as the number of patients with the same or more severe acuity levels ahead of a patient when he/she arrives at ED (i.e., the number of patients whose arrival time (time in) is earlier whereas disposition time is later than the current patient's time in). The staff level refers to the number of physicians in ED. Our data does not include this number, but includes the name of the physician for each patient. We used the number of distinct physician names that appear within the historical records within an hour after a patient's arrival time as the staff level. The staffing level is generally available real-time in many EDIS systems. We also included the target ward as a covariate. Overall, we had eight covariates.

According to Cox PH model, all covariates other than gender and respiration rate were statistically significant. Weibull hazard model identified respiration rate also as a statistically significant covariate. For MLP model, we used those variables which were identified as statistically significant by both Cox PH model and Weibull hazard model. Since none of the Cox PH, Weibull hazard or the MLP model outputs a standard R^2 value for assessing the model fit, which is an indicator of the proportion of total variation of outcomes explained by the model, we report the R^2 of linear fitting of the scatter plot of the mean LOS predicted against actual LOS. The R^2 of Cox PH model, Weibull Hazard model, and the MLP model are 0.18, 0.19 and 0.21, respectively.

We then investigated how much of the total variation can be explained by the model if the lab-work (e.g., blood work) and medical imaging (e.g., chest X-rays) undergone by the patients are also considered as covariates. In our data, the 'orderable item' covariates include 210 medical imaging items and 373 lab-work items. We used all of these covariates and the eight covariates in the initial models to estimate the ED LOS. According to Cox PH model, 81 of these items were statistically significant, whereas Weibull hazard model identified 88 of them as statistically significant covariates. Again, for MLP model, we used those variables which were identified as statistically significant by both Cox PH model and Weibull hazard model. By adding the orderable items as the input, the R^2 of Cox PH and Weibull hazard model nearly doubled from before to 0.334 and 0.349, respectively. The R^2 of the MLP method was improved to 0.342.

While these predictions are still not satisfactory, we are able to demonstrate that ED LOS prediction models for individual facilities show great promise and are good topics for further research.

4.5 Conclusion

In this Chapter, we propose three survival analysis models to estimate the ED LOS: Cox proposal hazard model, parametric survival analysis model using Weibull distribution as the baseline, and ANN survival analysis model. We conduct the experiments using only information collected during the triage as the covariates as well as using triage information and all lab-work (e.g., blood work) and medical imaging (e.g., chest X-rays) undergone by the patients covariates. When we only used triage information as the covariates, three models have similar results, but the ANN method led to slightly better estimation. However, when we add all the lab-work and medical imaging as the covariates as well, both Cox hazard model and parametric survival analysis model using Weibull distribution as the baseline almost doubles the R^2 value, whereas ANN is not able to improve its performance as much. The reason is the number of neural network inputs also increases proportionately with the increasing of the number of covariates, and this condition naturally increases the likelihood of ANN training algorithm convergence problems as well as the challenge for ANN to handle large number of inputs (Muknahallipatna and Chowdhury 1996). In fact, it has been suggested that when the dimension of inputs of ANN is increased, features selection is desirable to create an appropriate combination of inputs in order to obtain better generalization capabilities with the models (Satizábal M and Pérez-Uribe 2007).

Overall, we see that the variation in ED LOS was not well explained by these models. This can be attributed to significant variation that stems from all the “sub-queues” that form from any laboratory tests and medical imaging ordered by the physician as well as the natural variation present in ED from serving a variety of patients and with finite resources. There are a number of patient related factors, system factors as well as factors from other departments affecting the ED LOS. Since different factors affect different stages of the ED LOS, as noted at the beginning of the Chapter, it might be desirable to split the ED LOS into its major constituent components and estimate them separately with improved near real-time information (e.g., exploit information regarding cycle lead times for recently ordered lab/imaging orders and/or queue lengths to improve lab result estimates).

We also find that the estimation of ED LOS is improved if the information that is available after triage is added to the models. This observation verifies the value of ‘triage faculty’ or allowing triage staff to order certain tests that are likely to be ordered by the physician. In fact, research (Partovi, Nelson et al. 2001, Russ, Jones et al. 2010, Nestler, Fratzke et al. 2012) has reported that the strategy of ‘triage faculty’, i.e. placing a physician at triage to begin patient assessment and speeding up ordering of lab work etc. may reduce ED LOS for patients. Research has also shown that triage staff are capable of predicting some of the significant lab and imaging work necessary for the patients. In addition, a lot of these orders are placed long before patient's disposition decision, and hence, this information might be available for updating the ward-bed reservation model decisions.

CHAPTER 5: CASE STUDY: IMPLEMENTATION OF THE PROPOSED WARD-BED RESERVATION MODELING FRAMEWORK

In this section, we report the results from retroactively applying our proposed cost sensitive ward-bed reservation model to data collected from the ED and inpatient wards of a VAMC in the U.S. Mid-West.

5.1 ED and Ward Data

The ED data we used in the experiments is the same as the data discussed earlier in Chapters 3 and 4. The dataset is made up of records from 28,857 patients that visited the VAMC from July 25, 2011 to September 30, 2012 that are stored CPRS and EDIS systems.

The three main inpatient wards of the VAMC are medicine, surgery and psychiatry wards. The data for these three wards in this work came from a FY 2012 report of the VAMC. According to this report, the medicine ward is relatively busy with 84 beds and an average utilization of 84.6%. However, the surgery and psychiatry wards have relatively low utilization, and hence, these patients do not experience much delay and there is no benefit in considering ward-bed reservations in advance. For these reasons, we limit our case study experiments and discussions to the medicine ward. Like many hospitals, the VAMC uses average LOS as a key performance indicator in its inpatient flow management. According to the same report, the average patient LOS for FY 2012 in the medicine ward is 4.07 days in the VAMC.

Tables 5 and 6 show some statistics of the historical data. Table 5 shows the distribution of the demographic information. Table 6 shows the minimum, 1st quartile, median, 3rd quartile and the maximum ED LOS of the patients who took the orderable items with the top 20 highest frequencies. Note that the orderable item in each row of the first column of Table 6 is not the only item that has been taken by the corresponding patients. It is one of the many items that has been taken by the patients. Therefore, we cannot assume that the differences in the ED LOS between different columns are solely due to the difference between the corresponding items.

Table 5 *Statistics of patient information in our data*

Term	Count	Percentage
Acuity (ESI)		
1	48	0.17%
2	3392	11.75%
3	12486	43.27%
4	10804	37.44%
5	2127	7.37%
Destination		
Home	23837	82.60%
Medicine	4004	13.88%
Psychiatry	384	1.33%
Surgery	632	2.19%
Gender		
F	2627	9.10%
M	26230	90.90%
Age		
<10	2	0.01%
10<20	1	0.00%
20<30	1121	3.88%
30<40	2131	7.38%
40<50	3117	10.80%
50<60	7109	24.64%
60<70	8860	30.70%
70<80	3337	11.56%
80<90	2530	8.77%
90+	649	2.25%
Arrival Time		
0-2	966	3.35%

2-4	599	2.08%
4-6	706	2.45%
6-8	1411	4.89%
8-10	2867	9.94%
10-12	4091	14.18%
12-14	4241	14.70%
14-16	4144	14.36%
16-18	3553	12.31%
18-20	2756	9.55%
20-22	2121	7.35%
22-24	1402	4.86%

Table 6 *Statistics of ED LOS by Orderable items with highest frequencies*

Name of Orderable Items	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Freq.
nothing	0.2	50.93	84.87	109.6	138.2	3816	12164
cbc/5	15.93	185.8	260	325.8	368.9	2155	11569
comprehensive metabolic panel	15.93	191.1	265.7	329.4	376.1	2013	10015
troponin-poc	15.93	171.1	243.7	330.9	370.5	2046	8329
pt w/inr -top	15.93	189	267.3	342.5	387.9	2155	5760
magnesium	15.93	190	271.9	356	401.2	2046	4871
ptt -top	15.93	186.5	265.8	340.2	387.1	2155	4501
blood culture	47.08	208.7	279.9	332.8	384.4	1677	3987
chest 2 views pa,lat	16.33	156.7	225	279.3	328.2	1697	3670
urinalysis (reflex microscopic)	29.57	186.8	262.7	300.5	360	1619	3646
ck-mb screen	23.35	203.6	311.9	456.3	533.9	2046	3566
bnp (new)	15.93	182.1	257.6	331.1	371.2	2155	3499
culture & susceptibility	15.12	182	262.6	309	369.4	1905	3401
chest single view	15.93	178	258.5	339.8	385.8	2046	3150
urinalysis, (dcd 2-6-12)	15.12	181.8	256	310.4	362.8	1905	3125
troponin-i-ultra	15.93	249.8	379.5	536.4	721.7	2046	2654
basic metabolic panel	29.57	175.9	247.6	334.3	368.5	2395	2221
lipase	41.45	207	282.8	339.7	398.4	1749	2191
glucose (bst)	13.35	179.7	263.6	333.1	388.4	2046	2182
drugs of abuse	15.62	195.8	279.5	348.1	404.4	2395	2061

5.2 ED Patient Admission Prediction

While we still employ the mRVM model to estimate the probability of ED patient admission, we take a different approach in this Chapter from the approach discussed in Chapter 4. In this Chapter, we use the multi-class RVM model which is built into the Pattern Recognition Toolbox (PRT) for Matlab (The Pattern Recognition Toolbox (PRT) for MATLAB) . In PRT, instead of one mRVM model that is able to conduct the multiclass classification, a one-against-one approach, which we introduced in Chapter 4, was used to extend the binary RVM to conduct the multiclass classification. The main advantage of PRT method is that it is faster than both original mRVM and IF-mRVM. Since we need to investigate the results of ED patient admission prediction when we enforce different level of pre-defined 'model confidence', i.e., we assign a patient to a specific class only if the predicted probability of this patient belonging to the class is higher than the predefined 'confidence', we prefer a faster method.

While we will employ the full dataset to estimate ED LOS, as we state in the next section, we only employed a part for the dataset (data from 7,535 patients) to build and test the admission probability prediction performance of the RVM. In building our classification model, patient's age, gender, complaint, acuity, and time in were used as covariate factors in our experiments. In our analysis, patient's age was discretized mostly by decades and time in was categorized into twelve two-hour time segments (e.g., Noon-2PM, 2-4PM) as has been done in the literature (Sun, Heng et al. 2011, Peck, Benneyan et al. 2012). To further improve the accuracy of the admission prediction, we handled the free language primary patient complaint in a way that differs from the approach discussed in Chapter 3. Here, we manually matched the free language complaint codes initially to a standard list. A list of complaint codes generated by Aronsky at al. (Aronsky, Kendall et al. 2001) (57 codes in total) was used as the standard codes list in our work. We first manually mapped the free-text complaint in our data to the standard codes. For those patient records with no matching codes to the standard list due to local and regional differences in phrasing and abbreviating, we generated another 55 codes to match. Both these code sets are listed in the Appendix. In an effort to improve the confidence in our experimental results, we split the record dataset by randomly allocating 60% to be used as the training dataset and 40% to be used as the testing dataset.

A polynomial kernel with default parameters in PCI (a degree of 2 and an offset 0) were used in this work (further refinement of the kernel and associated parameters might improve performance but is not the focus of the research). For any patient record from the testing dataset, RVM outputs the probability that the patient belongs to each class (the sum of the probabilities across all classes, including discharge, is 1). By default, the class with the greatest probability is assigned to the testing point. In Table 7, we show the confusion Table of RVM for the 3,014 patients in the testing data, with a total accuracy of 90.8%, and the area under the curve (AUC) is calculated as 0.807 using the generalized multi-class procedure from (D.J. Hand 2001). Since RVM provides probabilistic predictions, it also allows us to set a predefined threshold for assigning class membership, i.e., each class is assigned a threshold. The higher the threshold of a class, the stricter we are in assigning the class membership (improving confidence). We investigated the accuracy of the RVM by setting increasingly higher threshold levels, and the results are shown in Table 8. Clearly, as the threshold increases, we are less likely to make reservation decisions for the patients with low confidence (posterior probability). By increasing the threshold, we are "losing" relatively few patients from consideration for advance reservation decision. The threshold determination for each class is thus equivalent to discriminating between the admission to a single target ward or being discharged.

Table 7 *Confusion Table for the RVM classifier on the testing dataset*

		Predicted Outcome			
		Discharged	Medicine	Psychiatry	Surgery
Actual Outcome	Discharged	2,426	57	16	0
	Medicine	134	270	0	5
	Psychiatry	16	0	20	0
	Surgery	30	21	1	18

Table 8 *Results of RVM classifier on the testing dataset when using different minimum probability thresholds for classification*

Probability Threshold:	0.25	0.4	0.6	0.8	0.9
Base Volume:	3014	3014	2943	2804	2684
Correct Predictions:	2734	2734	2686	2581	2480
Accuracy:	90.8%	90.8%	91.3%	92.1%	92.5%
Discharged Volume:	2499	2499	2470	2409	2337
Correct Predictions:	2426	2426	2398	2341	2271
Accuracy:	97.1%	97.1%	97.1%	97.2%	97.2%
Medicine Volume:	409	408	377	317	274
Correct Predictions:	270	270	252	212	183
Accuracy:	66.0%	66.2%	66.8%	66.9%	66.8%
Psychiatry Volume:	36	36	31	24	21
Correct Predictions:	20	20	18	15	13
Accuracy:	55.6%	55.6%	58.1%	62.5%	61.9%
Surgery Volume:	70	70	65	54	52
Correct Predictions:	18	18	18	13	13
Accuracy:	25.7%	25.7%	27.7%	24.1%	25.0%

We note the presence of misclassifications between the medicine and surgery wards. Conversations with ED physicians in the VAMC have confirmed that this could be caused by the “overflow” between these two wards among other known factors. Therefore, we investigated whether it is possible to incorporate this into the classifier and whether the "overflow" attribute for medicine and surgery specialty influences the prediction significantly. In order to do so, we combined the surgery ward and medicine ward into a single ward, called Medicine-Surgery ward, and then trained the classifier, and the result is shown in Table 9. The total accuracy slightly improves to 91.9%, and the AUC is calculated as 0.825 using the generalized multi-class procedure from (D.J. Hand 2001). In addition, we see little to no target ward misclassifications. All the errors stem from misclassifications between discharge and admission to target ward. Table 10 is similar to Table 7, it shows the accuracy of the RVM by setting increasingly higher threshold levels when combining the surgery ward and medicine ward into one ward.

Table 9 *Confusion results of RVM classifier with medicine and surgery ward combined*

		Predicted Outcome		
		Discharged	Psychiatry	Medicine-Surgery
Actual Outcome	Discharged	2422	16	61
	Psychiatry	16	20	0
	Medicine-Surgery	149	1	329

Table 10 *Results of RVM classifier on the testing dataset when using different minimum probability thresholds for classification*

Prob. Threshold:	0.25	0.4	0.6	0.8	0.9
Base Volume:	3014	3014	2959	2808	2685
Correct Predictions:	2771	2771	2739	2645	2546
Accuracy:	91.9%	91.9%	92.6%	94.2%	94.8%
Discharged Volume:	2499	2499	2474	2402	2330
Correct Predictions:	2422	2422	2408	2359	2298
Accuracy:	96.9%	96.9%	97.3%	98.2%	98.6%
Medicine Surgery Volume:	479	479	455	382	336
Correct Predictions:	329	329	315	271	238
Accuracy:	68.7%	68.7%	69.2%	70.9%	70.8%
Psychiatry Volume:	36	36	36	24	19
Correct Predictions:	20	20	16	15	10
Accuracy:	55.6%	55.6%	44.4%	62.5%	52.6%

We are not endorsing any specific technique for making patient admission probability predictions. We are only discussing one effective model, RVM classifier, to demonstrate the utility and practicality of the proposed cost-sensitive ward-bed reservation model.

5.3 Ward-bed Lead-time Estimation

As noted earlier, the average patient LOS for FY 2012 in the medicine ward is 4.07 days in the VAMC. From the VAMC's ED Information System (EDIS), the mean ward-bed lead-time delay for the entire medicine ward patient pool in our dataset is relatively small at just 9.6 minutes, attributable to the fact that it is a relatively large ward with 84 beds and only nominal utilization of 84.6%. However, another collaborating VAMC center has a 42 bed medicine ward with a utilization of 93.7%, so the situation there is much worse with ward-bed lead-times in hours. If one were to employ Kingman's Expression (18) to estimate the ward-bed lead-time for our VAMC by substituting the ward utilization level of 84.6%, 84 ward-beds, and average LOS of 4.07 days, it yields a variability component of ${C_a}^2 + {C_e}^2 \approx 0.32$. We note once again that the choice of generating expected ward-bed lead-times using the Kingman's expression is not restrictive. To study the impact of our proposed advance reservation decision support model on a variety of medical center situations, as reported in Table 6, we created several possible and realistic settings while employing the same variability component of ${C_a}^2 + {C_e}^2 = 0.32$. Table 11 also reports the predicted ward-bed lead-times from the Kingman's expression under these different settings.

Table 11 *Predicted ward-bed lead-time under different possible and realistic scenarios*

Setting	**Ward Utilization, *U***	**Beds in Ward, *m***	**Predicted TL**
1	0.846	84	9.6
2	0.875	84	17.7
3	0.9	84	31.2
4	0.925	84	57.7
5	0.846	64	16.6

6	0.875	64	28.9
7	0.9	64	48.6
8	0.925	64	86.1
9	0.846	44	33.2
10	0.875	44	54.4
11	0.9	44	86.4
12	0.925	44	145.4
13	0.846	24	91.1
14	0.875	24	137.8
15	0.9	24	204.4
16	0.925	24	321.8

5.4 Application of Proposed Cost-sensitive Ward-bed Reservation Model

In this section, we report on the results from applying the proposed cost-sensitive ward-bed reservation model to the testing data from VAMC.

5.4.1 Optimal Reservation Threshold

For reasons outlined earlier, we limit our focus to the medicine ward. Since Cox PH and Weibull hazard model yielded very similar performance in terms of R^2, and Weibull hazard model is able to output explicit estimation of the model parameters and the mean of the predicted ED LOS, we only focus on the results of Weibull model for the different settings in Table 11. Sample reservation thresholds for few select patients and some of their inputs are shown in Table 12.

5.4.2 Impact of Proposed Ward-Bed Reservation Model

We investigated the potential impact of making reservation according to the policy proposed by our model. If the patient is predicted as going to be admitted into the medicine ward and the admission probability is greater than the reservation threshold according to the multi-class classification model, the reservation is made for this patient. In the testing data, 348 patients were predicted as going to be admitted to medicine ward and 196 of them have valid ED LOS estimation (others don't have valid ED LOS estimation due to missing values of triage vitals).

Table 12 *Sample of proposed optimal policy results for select medicine ward patients*

MODEL INPUTS									RESULT
Setting	**Age**	**ESI**	**Pulse**	**Temperature**	**Respiration Rate**	**Crowding Level**	**Staff Level**	**Orderable Items**	**Threshold**
11	78	3	74	98	14	10	2	CT ABD/PELVIS W/O CONTRAST (PARENT), CULTURE & SUSCEPTIBILITY, URINALYSIS	0.81
14	56	3	89	99	18	9	6	BLOOD CULTURE,CBC/5,CK-MB SCREEN,COMPREHENSIVE METABOLIC,PANEL,MAGNESIUM, TROPONIN-I-ULTRA, WOUND CULTURE PANEL	0.26
15	68	2	82	98	19	24	5	ABDOMEN 2 VIEWS,BNP (NEW),CBC/5,CHEST 2 VIEWS PA,LAT,CK-MB SCREEN, COMPREHENSIVE METABOLIC PANEL,CULTURE & SUSCEPTIBILITY,D-DIMER HS, MAGNESIUM,PT W/INR -TOP,PTT -TOP,TROPONIN-I-ULTRA,URINALYSIS	0.41

Table 13 *Savings from making advance reservations according to the proposed policy*

	Wrong Reservations			Correct Reservations		
Test Setting	# of Reservations	Total Cost ($)	Ave. Cost per Patient ($)	# of Reservations	Total Savings ($)	Ave. Savings per Patient ($)
11	3	6,346	453	0	0	0
12	14	1,969	394	3	198	66
13	5	6,262	447	37	7,034	190
14	14	9,453	430	98	43,588	445
15	22	7,620	224	148	135,401	915
16	34	6,346	453	39	8,008	205

Out of these 196 patients, 148 of them were indeed medicine ward patients, and the other 48 patients were misclassified as medicine ward patients: 38 discharged patients and 10 surgery patients. The reservations for the misclassified patients were labeled wrong reservations, and the reservations for the patients who were admitted to the medicine wards were labeled correct reservations. The wrong reservations lead to bed wastage costs, and as discussed in chapter 2, the bed wastage cost is calculated using equation (3). The savings from reservations were still calculated as the cost of no reservation subtracted by the cost of making the reservation. Again, we calculated the bed wastage costs for each of the 16 test settings, and results are shown in Table 13.

For the first 10 settings, our policy did not suggest any reservations. For setting 11, it suggested 3 reservations, but all 3 reservations were for misclassified patients, so the policy led to some losses. For settings 12, 13 and 14, there are both wrong and correct reservations. Although for settings 12 and 14, the total savings from correct reservations exceeds the losses from wrong reservations. For settings 15 and 16, both the total and the average of the savings from correct reservations are much higher than the costs of the wrong reservations. This indicates that the reservation policy leads to significant savings for smaller wards with higher levels of utilization.

5.5 Conclusion

In this Chapter, using historical data, we retroactively tested the proposed decision support framework using data from a VAMC to investigate potential savings.

We see that the reservation threshold depends on a lot of factors, and the value of it could be anywhere between 0 and 1. Although the threshold values in Table 10 are identified as the x-axis value of the intersection point of the curve of costs of making reservation and no reservation as we showed in Figure 7 and Figure 8, for a specific patient, we do not have to create such curves and find the intersection of them. In fact, the threshold can be created for ‘groups’ of patients, i.e. if a future patient arrives at ED and his or her characteristics belong to one of the predefined ‘groups’, then we compare this patient’s admission probability and the threshold to decide whether a reservation should be made for this patient. However, in real life, it is not likely for an ED to have patients who have exactly the same age, the same ESI level, the same vitals, the same orderable items, and the same system factors (arrival time, crowding level, staff level etc.). Therefore, we only need to directly calculate the cost of making reservation and no reservation and make the decision.

The results indicate that for larger wards that benefit from various forms of risk pooling, the reservation system is not likely to create much savings. However, it results in significant cost saving potential and reduced patient boarding times for smaller inpatient wards with relatively high levels of utilization. Although the proposed reservation is not always true due to the errors of the admission probability prediction, and those wrong reservations cause loss, the savings of the correct reservations far exceed loss, resulting in potential net savings.

CHAPTER 6: CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

Overcrowding in emergency departments (EDs) is an increasing global problem that affects the quality and access of health care. It has prompted researchers to investigate a number of questions and seek solutions to this pressing problem. Although many methods have been proposed and being deployed across a variety of hospitals, there is no evidence that any set of solutions work best.

This dissertation develops a decision support framework for streamlining ED patient flow. Instead of getting into system re-design or capacity expansion, this work focuses on exploiting the near real-time information in ED to enhance patient flow coordination between ED and its down-stream departments (i.e., inpatient wards). Specifically, a novel cost sensitive ward-bed reservation policy based on the prediction of admission likelihood of ED patients is proposed. The policy identifies an admission probability as the threshold for making the reservation decision. It also recommends an optimal bed reservation time slot based on a modified News-Vendor model to minimize the cost of patient waiting and bed wastage. This model and some results are published in a Health Care Management Science journal (Qiu, Chinnam et al. 2014)

The proposed ward-bed reservation model relies on a probabilistic classifier for estimating the likelihood of ED patient admission. Given the free text nature of chief patient complaints recorded during triage and the relatively large vocabulary, effective feature selection becomes a critical pre-requisite for developing a robust and effective classifier. In this context, we propose a novel extension to the highly effective relevance vector machine (RVM) to simultaneously address feature selection during model learning, leading to the so-called integrated feature selection multi-class RVM (IF-mRVM). Results from testing the approach on a variety of benchmarking datasets are very promising.

The proposed ward-bed reservation model also relies on an estimate for ED patient’s length-of-stay. We have investigated a number of promising statistical and neural network based hazard rate models for arriving at these estimates.

Steps are currently being taken to pilot the methods for real-time ED patient admission predictions and patient flow coordination at a VA Medical Center (VAMC). The case study using historical data from a VAMC demonstrates that applying the proposed reservation policy might lead to some savings associated with reduced boarding times, in particular, for smaller wards with high levels of utilization.

There are several limitations to the current work. These limitations and their resolutions as part of future work are described next. First, we assumed that the lead-time of an admission ward-bed is deterministic within the model. As discussed earlier, this is practical for a growing number of hospitals employing IT bed management systems (BMS), including all the VA medical centers. However, for hospitals not equipped with BMS systems, the lead-times have to be modeled using historical data and will remain uncertain. Our sensitivity analysis however revealed that the impact of this uncertainty on expected cost is somewhat minimal. Still, an extended model considering this lead-time as a distribution will be explored in the future.

Second, the error rate of the target ward predictions has affected the number of reservations and the extent of cost savings reported in this study. One reason is that the free text patient complaints cause a challenge for pure data drive classifiers. In the future, semantic text mining methods will need be investigated in conjunction with data driven methods to handle the free text complaints (Katarzyniak 2011).

Subject matter experts at the VAMC indicated that additional patient information such as the history of the patients, i.e., whether the patient has visited ED in the last 24 hours, are key factors affecting the patient's admission likelihood and target wards. Incorporation of this and other available information from the VA's CPRS system is part of a future study.

Lastly, the difficulty in achieving greater savings in many of the case study test settings, in most part, stems from the inability of the ED LOS models to more accurately estimate patient LOS and should be the target for future research. The difficulties can be attributed to significant variation that stems from all the "sub-queues" that form from any laboratory tests and medical imaging ordered by the physician as well as the natural variation present in ED from serving a variety of patients and with finite resources. There are a number of patient related factors, system factors as well as factors from other departments affecting the ED LOS. Since different factors affect different stages of the ED LOS, it might be more desirable to split the ED LOS into its major constituent components and estimate them separately with improved near real-time information (e.g., exploit information regarding cycle lead times for recently ordered lab/imaging orders and/or queue lengths to improve lab result estimates). We believe that this research is critical to further enhancing the cost effectiveness of the proposed ward-bed reservation policies. In addition, given the significant differences across medical centers in terms of ED LOS variability, the identification of root causes for the variability in individual facilities should yield opportunities for improvement of ED processes, and in turn, savings from the proposed advanced inpatient bed reservation models. While cost savings have been the main focus in this study, the patient satisfaction and improved health outcomes from reduced wait times might be deemed more significant in many healthcare facilities.

APPENDIX: COMPLAINT CODES USED BY OUR STUDY

The appendix is the list of ‘complaint codes’ which we used to match the free text complaints in our data to. All these codes are used as binary input on our models. Part 1 includes the codes from the reference (Aronsky, Kendall et al. 2001), and part 2 includes the codes we generated.

Part 1: Existing complaint codes from (Aronsky, Kendall et al. 2001)

Abdominal pain, flank pain, overdose (intentional), abdominal problems, fluid/nutrition alteration, peripheral vascular/leg pain , allergies/hives/med reaction/sting, foreign body, procedure, assault/rape, follow-up , psychiatric/social problems, back pain, genito-urinary problem, respiratory problems, bites, gun-shot wound ,skin complaint/trauma, body aches, gynecological problem, stabbing, burns, headache, stroke/CVA, cardiac arrest, hemorrhage, substance abuse, cardio-vascular complaint, industrial/machinery accidents, fainting/syncope, chest pain, infection, temperature related convulsions, seizures, ingestion (accidental), traffic injury, dental toothache, laceration, traumatic injuries specific (FT), diabetic problems, medication refill , unconsciousness, (specific) diagnosis (FT), neck pain, unknown problem (man down), dizzy, needle stick, vaginal bleeding, ear/nose/throat problems, neurological complaint , weakness, eye problem, obstetrical problem, fall, orthopedic injury, fever, other (FT).

Part 2: Newly generated complaint codes

Abnormal lab, blood pressure, cold/flu, ankle pain, foot pain, hip pain, knee pain, hand pain, arm pain, groin pain, constipation, consultation, cyst/lump, decreased responsiveness, diarrhea, digestion related, drowsiness/lethargy, lib/extremity related, medical side effect, mental status, multiple complaints, muscle/skeletal related, nausea/vomiting, procedure in , procedure out, shoulder pain, tumor/cancer related, internal organ related/pathological issue, acute organ/pathological issue, abnormal behavior, acute/severe, arm pain, abscess, chest discomfort, anal/rectum abscess, congestion, blood clot, dementia, face related, discharge, fatigue, metabolic problem, skin related, short of breath, rectum/anal bleeding, groin pain, epigastria pain, detox, cough, blood sugar, acute infection, confusion, gout, uti, general discomfort.

REFERENCES

Afilalo, J., A. Marinovich, M. Afilalo, A. Colacone, R. Leger, B. Unger and C. Giguere (2004). "Nonurgent emergency department patient characteristics and barriers to primary care." Academic Emergency Medicine **11**(12): 1302-1310.

Ahmed, M. A. and T. M. Alkhamis (2009). "Simulation optimization for an emergency department healthcare unit in Kuwait." European Journal of Operational Research **198**(3): 936-942.

American College of Emergency Physicians (2010). ACEP policy statements: Triage scale standardization. Dallas, TX, American College of Emergency Physicians.

Andrulis, D. P., A. Kellermann, E. A. Hintz, B. B. Hackman and V. B. Weslowski (1991). "Emergency departments and crowding in United States teaching hospitals." Annals of Emergency Medicine **20**(9): 980-986.

Ardagh, M. W., J. E. Wells, K. Cooper, R. Lyons, R. Patterson and P. O'Donovan (2002). "Effect of a rapid assessment clinic on the waiting time to be seen by a doctor and the time spent in the department, for patients presenting to an urban emergency department: a controlled prospective trial." New Zealand Medical Journal **115**(1157): U28.

Aronsky, D., D. Kendall, K. Merkley, B. C. James and P. J. Haug (2001). "A comprehensive set of coded chief complaints for the emergency department." Academic Emergency Medicine **8**: 980-989.

Asplin, B. R., D. J. Magid, K. V. Rhodes, L. I. Solberg, N. Lurie and C. A. Camargo (2003). "A conceptual model of emergency department crowding." Annals of Emergency Medicine **42**(2): 173-180.

Atzema, C., G. Bandiera and M. J. Schull (2005). "Emergency department crowding: the effect on resident education." Annals of Emergency Medicine **45**(3): 276-281.

Baker, D. W., C. D. Stevens and R. H. Brook (1991). "Patients who leave a public hospital emergency department without being seen by a physician. Causes and consequences." JAMA **266**(8): 1085-1090.

Barthell, E. N., K. Coonan, J. Finnell, D. Pollock and D. Cochrane (2004). "Disparate systems, disparate data: integration, interfaces, and standards in emergency medicine information technology." Academic Emergency Medicine **11**(11): 1142-1148.

Bayley, M. D., J. S. Schwartz, F. S. Shofer, M. Weiner, F. D. Sites, K. B. Traber and J. E. Hollander (2005). "The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission." Annals of Emergency Medicine **45**(2): 110-117.

Bazarian, J. J., S. M. Schneider, V. J. Newman and J. Chodosh (1996). "Do admitted patients held in the emergency department impact the throughput of treat-and-release patients?" Academic Emergency Medicine **3**(12): 1113-1118.

Begley, C. E., Y. Chang, R. C. Wood and A. Weltge (2004). "Emergency department diversion and trauma mortality: evidence from houston, Texas." Journal of Trauma **57**(6): 1260-1265.

Biganzoli, E., P. Boracchi, L. Mariani and E. Marubini (1998). "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach." Statistics in Medicine **17**(10): 1169-1186.

Bindman, A. B., K. Grumbach, D. Keane, L. Rauch and J. M. Luce (1991). "Consequences of queuing for care at a public hospital emergency department." JAMA **266**(8): 1091-1096.

Biyu He, F. D., Alex Macario, Stefanos Zenios (2012). "The Timing of Staffing Decisions in Hospital Operating Rooms: Incorporating Workload Heterogeneity into the Newsvendor Problem " Manufacturing & Service Operations Management **14**(1): 99-114.

Bond, P. A. (2001). "A staffed ED assessment room: impact on wait times for nonurgent patients at a Saudi Arabian hospital." Journal of Emergency Nursing **27**(4): 394-395.

Boyle, A., K. Beniuk, I. Higginson and P. Atkinson (2012). "Emergency department crowding: time for interventions and policy evaluations." Emergency Medicine International.

Breiman, L. (1984). Classification and regression trees. Belmont, Calif., Wadsworth International Group.

Broomhead, D. S. L. D. R. S. and E. Radar (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks. Malvern, Worcs., Royals Signals & Radar Establishment.

Brown, S. F., A. J. Branford and W. Moran (1997). "On the use of artificial neural networks for the analysis of survival data." Neural Networks, IEEE Transactions on **8**(5): 1071-1077.

Bucheli, B. and B. Martina (2004). "Reduced length of stay in medical emergency department patients: a prospective controlled study on emergency physician staffing." European Journal of Emergency Medicine **11**(1): 29-34.

Buchman, T. G., K. L. Kubos, A. J. Seidler and M. J. Siegforth (1994). "A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit." Critical Care Medicine **22**: 750-762.

Burges, C. J. C. (1998). "A Tutorial on Support Vector Machines for Pattern Recognition." Data Min. Knowl. Discov. **2**(2): 121-167.

Burt, C. W., L. F. McCaig and R. H. Valverde (2006). "Analysis of ambulance transports and diversions among US emergency departments." Annals of Emergency Medicine **47**(4): 317-326.

Cameron, A., K. Rodgers, A. Ireland, R. Jamdar and G. A. McKay (2014). "A simple tool to predict admission at the time of triage." Emergency Medicine Journal.

Casalino, E., M. Wargon, A. Peroziello, C. Choquet, C. Leroy, S. Beaune, L. Pereira, J. Bernard and J. C. Buzzi (2013). "Predictive factors for longer length of stay in an emergency department: a prospective multicentre study evaluating the impact of age, patient's clinical acuity and complexity, and care pathways." Emergency Medicine Journal.

Chawla, N. V., K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer (2002). "SMOTE: synthetic minority over-sampling technique." Journal of Artificial Intelligence Research **16**(1): 321-357.

Cochran, J. K. and K. T. Roche (2009). "A multi-class queuing network analysis methodology for improving hospital emergency department performance." Computers & Operations Research **36**(5): 1497-1512.

Connelly, L. G. and A. E. Bair (2004). "Discrete event simulation of emergency department activity: a platform for system-level operations research." Academic Emergency Medicine **11**(11): 1177-1185.

Considine, J., M. Kropman, E. Kelly and C. Winter (2008). "Effect of emergency department fast track on emergency department length of stay: a case-control study." Emergency Medicine Journal **25**(12): 815-819.

Cooke, M. W., S. Wilson and S. Pearson (2002). "The effect of a separate stream for minor injuries on accident and emergency department waiting times." Emergency Medicine Journal **19**(1): 28-30.

CoVar Research,The Pattern Recognition Toolbox for MATLAB.

Cox, D. R. (1972). "Regression Models and Life-Tables." Journal of the Royal Statistical Society. Series B (Methodological) **34**: 187-220.

Cristianini, N. and J. Shawe-Taylor (2000). An introduction to support Vector Machines: and other kernel-based learning methods, Cambridge University Press.

Csiszár, I. (1996). Maxent, Mathematics, and Information Theory. Maximum Entropy and Bayesian Methods. K. Hanson and R. Silver, Springer Netherlands. **79:** 35-50.

D.J. Hand, R. J. T. (2001). "A simple generalization of the area under the ROC curve to multiple class classification problems." Machine Learning **45**(2): 171-186.

Darrab, A. A., J. Fan, C. M. Fernandes, R. Zimmerman, R. Smith, A. Worster, T. Smith and K. O'Connor (2006). "How does fast track affect quality of care in the emergency department?" European Journal of Emergency Medicine **13**(1): 32-35.

De Laurentiis, M. and P. M. Ravdin (1994). "A technique for using neural network analysis to perform survival analysis of censored data." Cancer Letters **77**(2-3): 127-138.

Dickson, E. W., S. Singh, D. S. Cheung, C. C. Wyatt and A. S. Nugent (2009). "Application of lean manufacturing techniques in the Emergency Department." Journal of Emergency Medicine **37**(2): 177-182.

Ding, R., M. L. McCarthy, J. S. Desmond, J. S. Lee, D. Aronsky and S. L. Zeger (2010). "Characterizing waiting room time, treatment time, and boarding time in the emergency department using quantile regression." Academic Emergency Medicine **17**(8): 813-823.

Donald, K. J., A. N. Smith, S. Doherty and V. Sundararajan (2005). "Effect of an on-site emergency physician in a rural emergency department at night." Rural and Remote Health **5**(3): 380.

Eckstein, M. and L. S. Chan (2004). "The effect of emergency department crowding on paramedic ambulance availability." Annals of Emergency Medicine **43**(1): 100-105.

Eitel, D. R., S. E. Rudkin, M. A. Malvehy, J. P. Killeen and J. M. Pines (2010). "Improving service quality by understanding emergency department flow: a White Paper and position statement prepared for the American Academy of Emergency Medicine." Journal of Emergency Medicine **38**(1): 70-79.

Eleuteri, A., R. Tagliaferri, L. Milano, S. De Placido and M. De Laurentiis (2003). "A novel neural network-based survival analysis model." Neural networks : the official journal of the International Neural Network Society **16**: 855-864.

Faddy, M., N. Graves and A. Pettitt (2009). "Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions." Value in Health **12**(2): 309-314.

Falvo, T., L. Grove, R. Stachura, D. Vega, R. Stike, M. Schlenker and W. Zirkin (2007). "The opportunity loss of boarding admitted patients in the emergency department." Academic Emergency Medicine **14**: 332-337.

Faraggi, D. and R. Simon (1995). "A neural network model for survival data." Statistics in Medicine **14**: 73-82.

Fatovich, D. M., Y. Nagree and P. Sprivulis (2005). "Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia." Emergency Medicine Journal **22**(5): 351-354.

Fernandes, C. M., J. M. Christenson and A. Price (1996). "Continuous quality improvement reduces length of stay for fast-track patients in an emergency department." Academic Emergency Medicine **3**(3): 258-263.

FitzGerald, G., G. A. Jelinek, D. Scott and M. F. Gerdtz (2010). "Emergency department triage revisited." Emergency Medicine Journal **27**(2): 86-92.

Friedman, J. (1989). "Regularized Discriminant Analysis." Journal of the American Statistical Association **84**(405): 165-175.

Fromm, R. E., Jr., L. R. Gibbs, W. G. McCallum, C. Niziol, J. C. Babcock, A. C. Gueler and R. L. Levine (1993). "Critical care in the emergency department: a time-based study." Critical Care Medicine **21**(7): 970-976.

Furnkranz, J. (1999). "Separate-and-Conquer Rule Learning." Artificial Intelligence Review **13**(1): 3-54.

Gardiner, J. C. (2013). "Modeling heavy-tailed distributions in healthcare utilization by parametric and Bayesian methods." Journal of Statistical Theory and Practice

Gilboy, N. (2012). Emergency severity index. implementation handbook Version 4. Rockville, MD, United States Agency for Healthcare Research and Quality.

Gilboy, N., D. Travers and R. Wuerz (1999). "Re-evaluating triage in the new millennium: A comprehensive look at the need for standardization and quality." Journal of Emergency Nursing **25**(6): 468-473.

Green, L. V., S. Savin and N. Savva (2013). "“Nursevendor Problem”: Personnel Staffing in the Presence of Endogenous Absenteeism." Management Science **59**(10): 2237-2256.

Green, L. V., J. Soares, J. F. Giglio and R. A. Green (2006). "Using queueing theory to increase the effectiveness of emergency department provider staffing." Academic Emergency Medicine **13**(1): 61-68.

Guyon, I. and A. Elisseeff (2003). "An introduction to variable and feature selection." Journal of Machine Learning Research **3**: 1157-1182.

Haas, S. W., D. Travers, J. E. Tintinalli, D. Pollock, A. Waller, E. Barthell, C. Burt, W. Chapman, K. Coonan, D. Kamens and J. McClay (2008). "Toward vocabulary control for chief complaint." Academic Emergency Medicine **15**(5): 476-482.

Han, J. H., D. J. France, S. R. Levin, I. D. Jones, A. B. Storrow and D. Aronsky (2010). "The effect of physician triage on emergency department length of stay." Journal of Emergency Medicine **39**(2): 227-233.

Handler, J. A., J. G. Adams, C. F. Feied, M. Gillam, J. Vozenilek, E. N. Barthell and S. J. Davidson (2004). "Emergency medicine information technology consensus conference: executive summary." Academic Emergency Medicine **11**(11): 1112-1113.

Hastie, T. and R. Tibshirani (1998). Classification by pairwise coupling. Proceedings of the 1997 conference on Advances in neural information processing systems**:** 507-513.

Haupert, J., E. K. Lee, H. Atallah, D. Wu, L. L. H. Jr, M. Wright, C. T. IV and E. T. Post (2014). Grady Health System and Georgia Institute of Technology: Transforming E.D. Workflow and Patient Care. Analytics Conference 2014, Informs. **Edelman Award Presentations**.

Health Management Associates (2011). Impact of Proposed Budget Cuts to Community- Based Mental Health Services. Texas Conference of Urban Counties.

Holden, R. J. (2011). "Lean Thinking in emergency departments: a critical review." Annals of Emergency Medicine **57**(3): 265-278.

Holroyd, B. R., M. J. Bullard, K. Latoszek, D. Gordon, S. Allen, S. Tam, S. Blitz, P. Yoon and B. H. Rowe (2007). "Impact of a triage liaison physician on emergency department overcrowding and throughput: a randomized controlled trial." Academic Emergency Medicine **14**(8): 702-708.

Holweg, M. (2007). "The genealogy of lean production." Journal of Operations Management **25**(2): 420-437.

Hoot, N. R. and D. Aronsky (2008). "Systematic review of emergency department crowding: causes, effects, and solutions." Annals of emergency medicine **52**(2): 126-136.

Howard, M. S., B. A. Davis, C. Anderson, D. Cherry, P. Koller and D. Shelton (2005). "Patients' perspective on choosing the emergency department for nonurgent medical care: a qualitative study exploring one reason for overcrowding." Journal of Emergency Nursing **31**(5): 429-435.

Hsu, C.-W. and C.-J. Lin (2002). "A comparison of methods for multiclass support vector machines." IEEE Transactions on Neural Networks **13**(2): 415-425.

Hwang, U., L. D. Richardson, T. O. Sonuyi and R. S. Morrison (2006). "The effect of emergency department crowding on the management of pain in older adults with hip fracture." Journal of the American Geriatrics Society **54**(2): 270-275.

Ieraci, S., E. Digiusto, P. Sonntag, L. Dann and D. Fox (2008). "Streaming by case complexity: evaluation of a model for emergency department Fast Track." Emergency Medicine Australasia **20**(3): 241-249.

Institute of Medicine (2007). Hospital-based emergency care : at the breaking point. Washington, D.C., Committee on the Future of Emergency Care in the United States Health System, National Academies Press.

Jaakkola, T. a. D. H. (1998). "Exploiting Generative Models in Discriminative Classifiers." Advances in Neural Information Processing Systems: 487-493.

Jackson, T., R. Beale, T. Jackson and R. Beale (1990). Neural Computing: An Introduction. Bristol, Adam Hilger.

Jelinek, G. A., T. J. Weiland and C. Mackinlay (2010). "Supervision and feedback for junior medical staff in Australian emergency departments: findings from the emergency medicine capacity assessment study." BMC Medical Education **10**: 74.

Jensen, F. V. (1996). Introduction to Bayesian Networks, Springer-Verlag New York, Inc.

Johnston, M. (2008). Hundreds die because of hospital crowding. New Zeland Herald. Wellington, New Zealand.

Katarzyniak, R. C. T.-F. H. C.-F. N. N. T. (2011). Semantic methods for knowledge management and communication. Berlin, Springer.

Kelen, G. D., J. J. Scheulen and P. M. Hill (2001). "Effect of an emergency department (ED) managed acute care unit on ED overcrowding and emergency medical services diversion." Academic Emergency Medicine **8**(11): 1095-1100.

Kelly, A. M., M. Bryant, L. Cox and D. Jolley (2007). "Improving emergency department efficiency by patient streaming to outcomes-based teams." Australian Health Review **31**(1): 16-21.

Kendall, J., B. Reeves and M. Clancy (1998). "Point of care testing: randomised controlled trial of clinical outcome." BMJ **316**(7137): 1052-1057.

Khouja, M. (1999). "The Single-period (news-vendor) problem: literature review and suggestions for future research." Omega The International Journal of Management Science **27**: 537-553.

Kilic, Y. A., F. A. Agalar, M. Kunt and M. Cakmakci (1998). "Prospective, double-blind, comparative fast-tracking trial in an academic emergency department during a period of limited resources." European Journal of Emergency Medicine **5**(4): 403-406.

King, D. L., D. I. Ben-Tovim and J. Bassham (2006). "Redesigning emergency department patient flows: application of Lean Thinking to health care." Emergency Medicine Australasia **18**(4): 391-397.

Kingman, J. F. C. (1962). "On Queues in Heavy Traffic." Journal of the Royal Statistical Society. Series B (Methodological) **24**(2): 383-392.

Kocher, K. E., W. J. Meurer, J. S. Desmond and B. K. Nallamothu (2012). "Effect of testing and treatment on emergency department length of stay using a national database." Academic Emergency Medicine **19**(5): 525-534.

Kohavi, R. and G. H. John (1997). "Wrappers for feature subset selection." Artificial Intelligence **97**(1-2): 273-324.

Kolker, A. (2008). "Process modeling of emergency department patient flow: effect of patient length of stay on ED diversion." Journal of Medical Systems **32**(5): 389-401.

Kotsiantis, S. B. (2007). "Supervised Machine Learning: A Review of Classification Techniques." Informatica **31**: 249-268.

Krochmal, P. and T. A. Riley (1994). "Increased health care costs associated with ED overcrowding." American Journal of Emergency Medicine **12**(3): 265-266.

Kulkarni, S. S., M. J. Magazine and A. S. Raturi (2004). "Risk Pooling Advantages of Manufacturing Network Configuration." Production and Operations Management **13**(2): 186-199.

Kulstad, E. B. and K. M. Kelley (2009). "Overcrowding is associated with delays in percutaneous coronary intervention for acute myocardial infarction." International Journal of Emergency Medicine **2**(3): 149-154.

Kwa, P. and D. Blake (2008). "Fast track: has it changed patient care in the emergency department?" Emergency Medicine Australasia **20**(1): 10-15.

Kyriacou, D. N., V. Ricketts, P. L. Dyne, M. D. McCollough and D. A. Talan (1999). "A 5-year time study analysis of emergency department patient care efficiency." Annals of Emergency Medicine **34**(3): 326-335.

Lambe, S., D. L. Washington, A. Fink, M. Laouri, H. Liu, J. Scura Fosse, R. H. Brook and S. M. Asch (2003). "Waiting times in California's emergency departments." Annals of Emergency Medicine **41**(1): 35-44.

Lee-Lewandrowski, E., D. Corboy, K. Lewandrowski, J. Sinclair, S. McDermot and T. I. Benzer (2003). "Implementation of a point-of-care satellite laboratory in the emergency department of an academic medical center. Impact on test turnaround time and patient emergency department length of stay." Archives of Pathology and Laboratory Medicine **127**(4): 456-460.

Lee, E. K. and H. Y. Atallah (2013). Modeling and Optimizing Emergency Department Workflow Informs. **Wagner Prize Presentations**.

Lee, E. T. and J. W. Wang. (2013). "Statistical methods for survival data analysis." from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=642570.

Leegon, J. and D. Aronsky (2006). "Impact of different training strategies on the accuracy of a Bayesian network for predicting hospital admission." AMIA Annual Symposium Proceedings **2006**: 474-478.

Leegon, J., I. Jones, K. Lanaghan and D. Aronsky (2005). "Predicting hospital admission for Emergency Department patients using a Bayesian network." AMIA Annual Symposium Proceedings 1022.

Li, J., L. Guo and N. Handly (2009). Hospital Admission Prediction Using Pre-hospital Variables. 2009 IEEE International Conference on Bioinformatics and Biomedicine, IEEE**:** 283-286.

Li, J., L. Guo, N. Handly, A. A. Mai and D. A. Thompson (2012). "Semantic-enhanced models to support timely admission prediction at emergency departments." Network Modeling Analysis in Health Informatics and Bioinformatics **1**: 161-172.

Liestbl, K., P. K. Andersen and U. Andersen (1994). "Survival analysis and neural nets." Statistics in Medicine **13**(12): 1189-1200.

Lin, C.-J. (2012). Machine learning software: design and practical use

Lindley-Jones, M. and B. J. Finlayson (2000). "Triage nurse requested x rays--are they worthwhile?" Journal of Accident and Emergency Medicine **17**(2): 103-107.

Littlestone, N. and M. K. Warmuth (1994). "The weighted majority algorithm." Information and Computation **108**(2): 212-261.

Liu, S., C. Hobgood and J. H. Brice (2003). "Impact of critical bed status on emergency department patient flow and overcrowding." Academic Emergency Medicine **10**(4): 382-385.

MacKay, D. J. (1992). "Bayesian interpolation." Neural Computation **4**(3): 415-447.

Majumder, S. K., Nirmalya Ghosh, and Pradeep K. Gupta (2005). "Relevance vector machine for optical diagnosis of cancer " Lasers in Surgery and Medicine **36**(4): 323-333.

Marazzi, A., F. Paccaud, C. Ruffieux and C. Beguin (1998). "Fitting the distributions of length of stay by parametric models." Medical Care **36**(6): 915-927.

Marcot, B. G., J. D. Steventon, G. D. Sutherland and R. K. McCann (2006). "Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation." Canadian Journal of Forest Research **36**(12): 3063-3074.

Markovitch, S. and D. Rosenstein (2002). "Feature Generation Using General Constructor Functions." Machine Learning **49**(1): 59-98.

McConnell, K. J., C. F. Richards, M. Daya, C. C. Weathers and R. A. Lowe (2006). "Ambulance diversion and lost hospital revenues." Annals of Emergency Medicine **48**(6): 702-710.

McHugh, M., K. V. Dyke and M. McClelland (2011). Improving Patient Flow and Reducing Emergency Department Crowding: A Guide for Hospitals. Rockville, MD, Agency for Healthcare Research and Quality.

McHugh, M. and P. Tanabe (2011). The Emergency Severity Index is the most commonly used triage system in the U.S. . Boston, MA, Society of Academic Emergency Medicine Annual Meeting.

Mika, S., G. Ratsch, J. Weston, B. Scholkopf and K. Muller (1999). Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.

Miro, O., M. T. Antonio, S. Jimenez, A. De Dios, M. Sanchez, A. Borras and J. Milla (1999). "Decreased health care quality associated with emergency department overcrowding." European Journal of Emergency Medicine **6**(2): 105-107.

Mitchell, T. M. (1997). Machine Learning, McGraw-Hill, Inc.

Moloney, E. D., K. Bennett, D. O'Riordan and B. Silke (2006). "Emergency department census of patients awaiting admission following reorganisation of an admissions process." Emergency Medicine Journal **23**(5): 363-367.

Moskop, J. C., D. P. Sklar, J. M. Geiderman, R. M. Schears and K. J. Bookman (2009). "Emergency department crowding, part 1--concept, causes, and moral consequences." Annals of Emergency Medicine **53**(5): 605-611.

Moskop, J. C., D. P. Sklar, J. M. Geiderman, R. M. Schears and K. J. Bookman (2009). "Emergency department crowding, part 2--barriers to reform and strategies to overcome them." Annals of Emergency Medicine **53**(5): 612-617.

Mu-Chun, S., J. Woung-Fei and C. Hsiao-Te (1996). A static hand gesture recognition system using a composite neural network. Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on.

Muknahallipatna, S. and B. H. Chowdhury (1996). Input dimension reduction in neural network training-case study in transient stability assessment of large systems. Intelligent Systems Applications to Power Systems, 1996. Proceedings, ISAP '96., International Conference on.

Murray, R. P., M. Leroux, E. Sabga, W. Palatnick and L. Ludwig (1999). "Effect of point of care testing on length of stay in an adult emergency department." Journal of Emergency Medicine **17**(5): 811-814.

National Center for Injury, P. and Control (1997). Data elements for emergency department systems. Atlanta, Ga. (4770 Buford Hwy. NE, Atlanta 30341-3724), National Center for Injury Prevention and Control.

Nestler, D. M., A. R. Fratzke, C. J. Church, L. Scanlan-Hanson, A. T. Sadosty, M. P. Halasy, J. L. Finley, A. Boggust and E. P. Hess (2012). "Effect of a physician assistant as triage liaison provider on patient throughput in an academic emergency department." Academic Emergency Medicine **19**(11): 1235-1241.

Ng, D., G. Vail, S. Thomas and N. Schmidt (2010). "Applying the Lean principles of the Toyota Production System to reduce wait times in the emergency department." Canadian Journal of Emergency Medicine **12**(1): 50-57.

Nilsson, N. J. (1965). Learning machines. New York, McGraw-Hill.

Niska, R., F. Bhulya and J. Xu (2010). National Hospital Ambulatory Medical Care Survey: 2007 Emergency Department Summary. National Health Statistics Reports. Hyattsville, MD, National Center for Health Statistics.

Nocedal, J. and S. Wright (2006). Numerical Optimization, Springer.

O'Brien, D., A. Williams, K. Blondell and G. A. Jelinek (2006). "Impact of streaming "fast track" emergency department patients." Australian Health Review **30**(4): 525-532.

Olivares, M., C. Terwiesch and L. Cassorla (2008). "Structural Estimation of the Newsvendor Model: An Application to Reserving Operating Room Time." Management Science **54**(1): 41-55.

Olshaker, J. S. (2009). "Managing emergency department overcrowding." Emergency Medicine Clinics of North America **27**(4): 593-603.

Oredsson, S., H. Jonsson, J. Rognes, L. Lind, K. E. Goransson, A. Ehrenberg, K. Asplund, M. Castren and N. Farrohknia (2011). "A systematic review of triage-related interventions to improve patient flow in emergency departments." Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **19**: 43.

Parris, W., S. McCarthy, A. M. Kelly and S. Richardson (1997). "Do triage nurse-initiated X-rays for limb injuries reduce patient transit time?" Accident and Emergency Nursing **5**(1): 14-15.

Partovi, S. N., B. K. Nelson, E. D. Bryan and M. J. Walsh (2001). "Faculty triage shortens emergency department length of stay." Academic Emergency Medicine **8**(10): 990-995.

Partovi, S. N., B. K. Nelson, E. D. Bryan and M. J. Walsh (2001). "Faculty triage shortens emergency department length of stay." Academic Emergency Medicine **8**(10): 990-995.

Parvin, C. A., S. F. Lo, S. M. Deuser, L. G. Weaver, L. M. Lewis and M. G. Scott (1996). "Impact of point-of-care testing on patients' length of stay in a large emergency department." Clinical Chemistry **42**(5): 711-717.

Patel, P. B. and D. R. Vinson (2005). "Team assignment system: expediting emergency department care." Annals of Emergency Medicine **46**(6): 499-506.

Peck, J., J. Benneyan, S. Gaehde, D. Nightingale and V. H. A. Boston (2012). Models for Using Predictions to Facilitate Hospital Patient Flow. Healthcare Systems Process Improvment Conference 1-6.

Peck, J. S., J. C. Benneyan, D. J. Nightingale and S. A. Gaehde (2012). "Predicting Emergency Department Inpatient Admissions to Improve Same-day Patient Flow." Academic Emergency Medicine **19**: E1045-E1054.

Pines, J. M., R. J. Batt, J. A. Hilton and C. Terwiesch (2011). "The financial consequences of lost demand and reducing boarding in hospital emergency departments." Annals of Emergency Medicine **58**(4): 331-340.

Pines, J. M. and J. E. Hollander (2008). "Emergency department crowding is associated with poor care for patients with severe pain." Annals of Emergency Medicine **51**(1): 1-5.

Polevoi, S. K., J. V. Quinn and N. R. Kramer (2005). "Factors associated with patients who leave without being seen." Academic Emergency Medicine **12**(3): 232-236.

Pollino, C. A., O. Woodberry, A. Nicholson, K. Korb and B. T. Hart (2007). "Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment." Environmental Modelling & Software **22**(8): 1140-1152.

Powell, E. S., R. K. Khare, A. K. Venkatesh, B. D. Van Roo, J. G. Adams and G. Reinhardt (2012). "The relationship between inpatient discharge timing and emergency department boarding." The Journal of emergency medicine **42**(2): 186-196.

Psorakis, I., Theodoros Damoulas, and Mark A. Girolami 21.10 (2010): 1588-1598. (2010). "Multiclass relevance vector machines: sparsity and accuracy " IEEE Transactions on Neural Networks **21**(10): 1588-1598.

Qiu, S., R. B. Chinnam, A. Murat, B. Batarse, N. H. and W. Jordan (2014). "A cost sensitive inpatient bed reservation approach to reduce emergency department boarding." Health Care Management Science **Accepted**.

R Foundation for Statistical Computing,R: A Language and Environment for Statistical Computing.

Rathlev, N. K., J. Chessare, J. Olshaker, D. Obendorfer, S. D. Mehta, T. Rothenhaus, S. Crespo, B. Magauran, K. Davidson, R. Shemin, K. Lewis, J. M. Becker, L. Fisher, L. Guy, A. Cooper and E. Litvak (2007). "Time series analysis of variables associated with daily mean emergency department length of stay." Annals of Emergency Medicine **49**(3): 265-271.

Ravdin, P. and G. Clark (1992). "A practical application of neural network analysis for predicting outcome of individual breast cancer patients." Breast Cancer Research and Treatment **22**(3): 285-293.

Reeder, T. J., D. L. Burleson and H. G. Garrison (2003). "The overcrowded emergency department: a comparison of staff perceptions." Academic Emergency Medicine **10**(10): 1059-1064.

Richardson, J. R., G. Braitberg and M. J. Yeoh (2004). "Multidisciplinary assessment at triage: a new way forward." Emergency Medicine Australasia **16**(1): 41-46.

Rifkin, R. and A. Klautau (2004). "In Defense of One-Vs-All Classification." The Journal of Machine Learning Research **5**: 101-141.

Ripley, R. M., A. L. Harris and L. Tarassenko (2004). "Non-linear survival analysis using neural networks." Statistics in Medicine **23**: 825-842.

Robert, J. H. and C. J. Lakhmi, Eds. (2001). Radial basis function networks 2: new advances in design. Studies in Fuzziness and Soft Computing, Physica-Verlag GmbH.

Rodi, S. W., M. V. Grau and C. M. Orsini (2006). "Evaluation of a fast track unit: alignment of resources and demand results in improved satisfaction and decreased length of stay for emergency department patients." Quality Management in Health Care **15**(3): 163-170.

Rogers, T., N. Ross and D. Spooner (2004). "Evaluation of a 'See and Treat' pilot study introduced to an emergency department." Accident and Emergency Nursing **12**(1): 24-27.

Rondeau, K. V. and L. H. Francescutti (2005). "Emergency department overcrowding: the impact of resource scarcity on physician job satisfaction." Journal of Healthcare Management **50**(5): 327-340; discussion 341-322.

Rosenblatt, F. (1962). Principles of neurodynamics: perceptrons and the theory of brain mechanisms, Spartan Books.

Ross, M. A., S. Naylor, S. Compton, K. A. Gibb and A. G. Wilson (2001). "Maximizing use of the emergency department observation unit: a novel hybrid design." Annals of Emergency Medicine **37**(3): 267-274.

Rowe, B. H., P. Channan, M. Bullard, S. Blitz, L. D. Saunders, R. J. Rosychuk, H. Lari, W. R. Craig and B. R. Holroyd (2006). "Characteristics of patients who leave emergency departments without being seen." Academic Emergency Medicine **13**(8): 848-852.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. E. R. David, L. M. James and C. P. R. Group, MIT Press**:** 318-362.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition. E. R. David, L. M. James and C. P. R. Group, MIT Press. **1:** 318-362.

Russ, S., I. Jones, D. Aronsky, R. S. Dittus and C. M. Slovis (2010). "Placing physician orders at triage: the effect on length of stay." Annals of Emergency Medicine **56**(1): 27-33.

Russell, S. and P. Norvig Artificial intelligence: a modern approach (2nd edition), Prentice Hall.

Saghafian, S., G. Austin and S. Traub (2014). "Operations Research Contributions to Emergency Department Patient Flow Optimization: Review and Research Prospects." Available at SSRN 2420163.

Sanchez, M., A. J. Smally, R. J. Grant and L. M. Jacobs (2006). "Effects of a fast-track area on emergency department performance." Journal of Emergency Medicine **31**(1): 117-120.

Satizábal M, H. and A. Pérez-Uribe (2007). Relevance Metrics to Reduce Input Dimensions in Artificial Neural Networks. Artificial Neural Networks – ICANN 2007. J. Sá, L. Alexandre, W. Duch and D. Mandic, Springer Berlin Heidelberg. **4668:** 39-48.

Schneider, S. M., M. E. Gallery, R. Schafermeyer and F. L. Zwemer (2003). "Emergency department crowding: a point in time." Annals of Emergency Medicine **42**(2): 167-172.

Schull, M. J., M. Vermeulen, G. Slaughter, L. Morrison and P. Daly (2004). "Emergency department crowding and thrombolysis delays in acute myocardial infarction." Annals of Emergency Medicine **44**(6): 577-585.

Shaw, K. N. and J. M. Lavelle (1998). "VESAS: a solution to seasonal fluctuations in emergency department census." Annals of Emergency Medicine **32**(6): 698-702.

Shmueli, G. P. N. R. B. P. C. (2010). Data mining for business intelligence : concepts, techniques, and applications in Microsoft Office Excel with XLMiner. Hoboken, N.J., Wiley.

Singer, A. J., P. Viccellio, H. C. Thode, Jr., J. L. Bock and M. C. Henry (2008). "Introduction of a stat laboratory reduces emergency department length of stay." Academic Emergency Medicine **15**(4): 324-328.

Spaite, D. W., F. Bartholomeaux, J. Guisto, E. Lindberg, B. Hull, A. Eyherabide, S. Lanyon, E. A. Criss, T. D. Valenzuela and C. Conroy (2002). "Rapid process redesign in a university-based emergency department: decreasing waiting time intervals and improving patient satisfaction." Annals of Emergency Medicine **39**(2): 168-177.

Sprivulis, P. C., J. A. Da Silva, I. G. Jacobs, A. R. Frazer and G. A. Jelinek (2006). "The association between hospital overcrowding and mortality among patients admitted via

Western Australian emergency departments." Medical Journal of Australia **184**(5): 208-212.

Stover-Baker, B., B. Stahlman and M. Pollack (2012). "Triage nurse prediction of hospital admission." Journal of Emergency Nursing **38**(3): 306-310.

Subash, F., F. Dunn, B. McNicholl and J. Marlow (2004). "Team triage improves emergency department efficiency." Emergency Medicine Journal **21**(5): 542-544.

Sun, Y., B. H. Heng, S. Y. Tay and E. Seow (2011). "Predicting hospital admissions at emergency department triage using routine administrative data." Academic Emergency Medicine **18**: 844-850.

Sun, Y., K. L. Teow, B. H. Heng, C. K. Ooi and S. Y. Tay (2012). "Real-time prediction of waiting time in the emergency department, using quantile regression." Annals of Emergency Medicine **60**(3): 299-308.

T. Damoulas, M. A. G. (2008). "Probabilistic multi-class multi-kernel learning: On protein fold recognition and remote homology detection." Bioinformatics **24**(10): 1264-1270.

Tan, P.-N., M. Steinbach and V. Kumar (2005). Introduction to Data Mining, (First Edition), Addison-Wesley Longman Publishing Co., Inc.

Technology, V. H. I. (2010). VistA Emergency Department Intergration Software User Guide Version 1.0, Department of Veteran Affairs Health Systems Desing and Development.

Thurston, J. and S. Field (1996). "Should accident and emergency nurses request radiographs? Results of a multicentre evaluation." Journal of Accident and Emergency Medicine **13**(2): 86-89.

Tipping, M. E. ((2001)). "Sparse Bayesian Learning and the Relevance Vector Machine." The Journal of Machine Learning Research 1: 211-244

Tipping, M. E., and Anita C. Faul (2003). "Fast marginal likelihood maximisation for sparse Bayesian models." Proceedings of the ninth international workshop on artificial intelligence and statistics **1**(3).

Travers, J. P. and F. C. Lee (2006). "Avoiding prolonged waiting time during busy periods in the emergency department: Is there a role for the senior emergency physician in triage?" European Journal of Emergency Medicine **13**(6): 342-348.

Tsai, W. W., D. B. Nash, B. Seamonds and G. J. Weir (1994). "Point-of-care versus central laboratory testing: an economic analysis in an academic medical center." Clinical Therapeutics **16**(5): 898-910; discussion 854.

United States General Accounting Office (2003). HOSPITAL EMERGENCY Crowded Conditions Vary among Hospitals and Communities.

Uusitalo, L. (2007). "Advantages and challenges of Bayesian networks in environmental modelling." Ecological Modelling **203**(3–4): 312-318.

Vapnik, V. N. (1995). The nature of statistical learning theory, Springer-Verlag New York, Inc.

Wachtel, R. E. and F. Dexter (2010). "Review article: review of behavioral operations experimental studies of newsvendor problems for operating room management." Anesthesia and Analgesia **110**(6): 1698-1710.

Weiss, S. J., A. A. Ernst, R. Derlet, R. King, A. Bair and T. G. Nick (2005). "Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED." American Journal of Emergency Medicine **23**(3): 288-294.

Welch, S. J., B. R. Asplin, S. Stone-Griffith, S. J. Davidson, J. Augustine, J. Schuur and A. Emergency Department Benchmarking (2011). "Emergency department operational metrics, measures and definitions: results of the Second Performance Measures and Benchmarking Summit." Annals of Emergency Medicine **58**(1): 33-40.

Wiler, J. L., D. A. Handel, A. A. Ginde, D. Aronsky, N. G. Genes, J. L. Hackman, J. A. Hilton, U. Hwang, M. Kamali, J. M. Pines, E. Powell, M. Sattarian and R. Fu (2012). "Predictors of patient length of stay in 9 emergency departments." The American journal of emergency medicine **30**(9): 1860-1864.

Wu, T.-F., C.-J. Lin and R. C. Weng (2004). "Probability Estimates for Multi-class Classification by Pairwise Coupling." The Journal of Machine Learning Research **5**: 975-1005.

Wuerz, R. C., L. W. Milne, D. R. Eitel, D. Travers and N. Gilboy (2000). "Reliability and validity of a new five-level triage instrument." Academic Emergency Medicine **7**(3): 236-242.

Xie, B. (2013). "Development and Validation of Models to Predict Hospital Admission for Emergency Department Patients." International Journal of Statistics in Medical Research **2**(1): 55-65.

Yang, Y. and J. O. Pedersen (1997). A Comparative Study on Feature Selection in Text Categorization. Proceedings of the Fourteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc.**:** 412-420.

Yang, Y. and G. Webb (2003). On Why Discretization Works for Naive-Bayes Classifiers. AI 2003: Advances in Artificial Intelligence. T. Gedeon and L. Fung, Springer Berlin Heidelberg. **2903:** 440-452.

Yu, L. and H. Liu (2004). "Efficient Feature Selection via Analysis of Relevance and Redundancy." Journal of Machine Learning Research **5**: 1205-1224.

Zhang, G. P. (2000). "Neural networks for classification: a survey." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **30**(4): 451-462.

ABSTRACT

IMPROVING EMERGENCY DEPARTMENT PATIENT FLOW THROUGH NEAR REAL-TIME ANALYTICS

by

SHANSHAN QIU

August 2014

Advisor: Dr. Ratna Babu Chinnam

Co-Advisor: Dr. Alper Murat

Major: Industrial Engineering

Degree: Doctor of Philosophy

This dissertation research investigates opportunities for developing effective decision support models that exploit near real-time (NRT) information to enhance the "operational intelligence" within hospital Emergency Departments (ED). Approaching from a systems engineering perspective, the study proposes a novel decision support framework for streamlining ED patient flow that employs machine learning, statistical and operations research methods to facilitate its operationalization.

ED crowding has become the subject of significant public and academic attention, and it is known to cause a number of adverse outcomes to the patients, ED staff as well as hospital revenues. Despite many efforts to investigate the causes, consequences and interventions for ED overcrowding in the past two decades, scientific knowledge remains limited in regards to strategies and pragmatic approaches that actually improve patient flow in EDs.

Motivated by the gaps in research, we develop a near real-time triage decision support system to reduce ED boarding and improve ED patient flow. The proposed system is a novel variant of a newsvendor modeling framework that integrates patient admission probability

prediction within a proactive ward-bed reservation system to improve the effectiveness of bed coordination efforts and reduce boarding times for ED patients along with the resulting costs. Specifically, we propose a cost-sensitive bed reservation policy that recommends optimal bed reservation times for patients right during triage. The policy relies on classifiers that estimate the probability that the ED patient will be admitted using the patient information collected and readily available at triage or right after. The policy is cost-sensitive in that it accounts for costs associated with patient admission prediction misclassification as well as costs associated with incorrectly selecting the reservation time.

To achieve the objective of this work, we also addressed two secondary objectives: first, development of models to predict the admission likelihood and target admission wards of ED patients; second, development of models to estimate length-of-stay (LOS) of ED patients. For the first secondary objective, we develop an algorithm that incorporates feature selection into a state-of-the-art and powerful probabilistic Bayesian classification method: multi-class relevance vector machine. For the second objective, we investigated the performance of hazard rate models (in particual, the non-parametric Cox proportional hazard model, parametric hazard rate models, as well as artificial neural networks for modeling the hazard rate) to estimate ED LOS by using the information that is available at triage or right after as the covariates in the models.

The proposed models are tested using extensive historical data from several U.S. Department of Veterans Affairs Medical Centers (VAMCs) in the Mid-West. The Case Study using historical data from a VAMC demonstrates that applying the proposed framework leads to significant savings associated with reduced boarding times, in particular, for smaller wards with high levels of utilization.

For theory, our primary contribution is the development of a cost sensitive ward-bed reservation model that effectively accounts for various costs and uncertainties. This work also contributes to the development of an integrated feature selection method for classification by developing and validating the mathematical derivation for feature selection during mRVM learning. Another contribution stems from investigating how much the ED LOS estimation can be improved by incorporating the information regarding ED orderable item lists.

Overall, this work is a successful application of mixed methods of operation research, machine learning and statistics to the important domain of health care system efficiency improvement.

AUTOBIOGRAPHICAL STATEMENT

Shanshan Qiu was born in Enshi, an autonomous prefecture of the two Chinese minority ethnic groups, Tujia (which Shanshan belongs to) and Miao in western Hubei province of China. She received the Bachelor of Science and Master in Science in Applied Mathematics in 2002 and 2005, respectively, from Huazhong Normal University (HZNU), Wuhan, China. She was a lecturer of Linear Algebra and Calculus for the undergraduate students in HZNU when she was a master student. After she obtained her M.S. degree, she started her work as a student advisor in the department of computer science of HZNU until she came to the United States in 2007.

She got her M.A. in Mathematical Statistics in December, 2009 from the Department of Mathematics of Wayne State University. During her Master's program, she was a summer research assistant for a terrain characterization project supported by the U.S. Army TARDEC-NAC, and she developed mathematical models that appropriately characterize terrain topology for different applications.

During her studies the Industrial & Systems Engineering Department of Wayne State University, she participated in a number of research projects funded by agencies such as the National Science Foundation and the U.S. Department of Veterans Affairs (VA). She made a number of technical presentations at INFORMS, POMS, INFORMS HEALTHCARE and several other conferences. An article based on this dissertation research has been published in the journal of Healthcare Management Science. Her research areas are Analytics and Informatics, Applied Statistics, Data Mining, Computational Intelligence, and Supply Chain Management.

She joined Global Analytics of Ford Motor Credit Company as a financial analyst in March, 2014 and is planning to continue her job at the same department after graduation.