UNIVERSITÉ DE MONTRÉAL

INTEGRATIVE PREDICTIVE SUPPORT SYSTEMS FOR HOSPITAL'S RESOURCE PLANNING AND SCHEDULING

ELNAZ KARIMI
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE INDUSTRIEL)
OCTOBER 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

INTEGRATIVE PREDICTIVE SUPPORT SYSTEMS FOR HOSPITAL'S RESOURCE PLANNING AND SCHEDULING

présentée par : KARIMI Elnaz
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. ROUSSEAU Louis-Martin, Ph. D., président
M. FRAYRET Jean-Marc, Ph. D., membre et directeur de recherche
M. GENDREAU Michel, Ph. D., membre et codirecteur de recherche
M. VERTER Vedat, Ph. D., membre et codirecteur de recherche
Mme BÉLANGER Valérie, Ph. D., membre
M. RUIZ Angel, Doctorat, membre externe

DEDICATION

In memory of my beloved grandfather, Hajibaba . . .

ACKNOWLEDGEMENTS

Countless number of hours were devoted to the development of this thesis, and this journey could not have been accomplished without the dedicated support of many individuals.

First and foremost, I would like to thank my supervisors Prof. Jean-Marc Frayret, Prof. Michele Gendreau and Prof. Vedat Verter for their guidance on professional and personal levels. Despite many hardships during the course of my doctoral study, I could always count on their selfless insight, feedback and advice. I would also like to thank my PhD jury members, Prof. Louis-Martin Rousseau, Prof. Valérie Bélanger and Prof. Angel Ruiz for their thoughtful comments.

Next, I would like to thank my parents for their unconditional love, support, and encouragement through the course of this PhD. They have always helped me overcome my frustrations, given me strength to push through difficult times, and together we have celebrated my joyful days. I would also like to thank my siblings, Behnaz and Mehdi for their unwavering emotional support, and my grandma for the calmness and happiness she presented in the darkest of days, and being the reason to smile despite the problems I faced.

Last but not least, I would like to specially thank Arshya Feizi for always being beside me during the rain and shine of this work, for encouraging me to never set a limit, for giving me confidence to believe in my abilities, and for motivating me to turn my ideas into reality.

RÉSUMÉ

Le système de santé du Canada a du mal à gérer le nombre croissant de patients ayant plusieurs maladies chroniques nécessitant l'accès à des soins de longue durée, et cela, principalement en raison de vieillissement de la population. Cela entraîne notamment de longs délais d'attente pour les patients et une augmentation des frais des soins de santé. Comme les hôpitaux représentent la plus grande part du budget de la santé, ils doivent améliorer leur efficacité opérationnelle en utilisant plus efficacement leurs ressources. En particulier, les hôpitaux qui fournissent aux patients des soins directs et un accès à des ressources coûteuses telles que les chirurgiens, les salles d'opération, les unités de soins intensifs et les salles d'opération, ont de la pression pour gérer leurs ressources efficacement.

Les chercheurs en recherche opérationnelle ont largement abordé les problèmes liés à la planification et l'ordonnancement des ressources dans les hôpitaux pendant de nombreuses années. Les modèles analytiques conventionnels visent ainsi à améliorer l'efficacité de la prise de décision de planification des ressources hospitalières à des fins stratégiques (à long terme), tactiques (à moyen terme) et opérationnelles (à court terme). Cependant, ces modèles ont du mal à adresser efficacement la complexité, la variabilité, et l'incertitude inhérentes aux opérations hospitalières, car ils utilisent souvent des distributions statistiques simplistes pour émuler ces opérations. Par conséquent, ils sont sous-optimaux dans des contextes réels d'utilisation.

Avec l'accroissement continu des quantités massives de données collectées dans les hôpitaux et les systèmes de santé, ainsi que les progrès dans le domaine de la modélisation prédictive, la communauté de la recherche opérationnelle a maintenant l'occasion de mieux analyser, comprendre et reproduire la complexité, la variabilité et l'incertitude des opérations hospitalières. À cette fin, l'objectif principal de cette thèse est de développer des cadres prédictifs intégrés, capables d'analyser et d'extraire des informations à partir de masses de données afin de mieux éclairer la planification et l'ordonnancement des ressources hospitalières aux niveaux stratégique, tactique, et opérationnel. Au meilleur des connaissances de l'auteur, cette thèse est une des premières à proposer des cadres pour la conception de systèmes prédictifs dans les hôpitaux.

Au niveau stratégique et tactique, le premier article (chapitre 4) développe un cadre hybride basé sur l'apprentissage machine et la simulation pour prédire la demande personnalisée des patients au niveau des ressources hospitalières. Le cadre reflète notamment la relation à long terme entre les hôpitaux et les patients ayant des maladies chroniques, couvrant ainsi

un horizon à long terme et intégrant le fait que les patients ont besoin, non pas d'une, mais de plusieurs visites à l'hôpital et accès à divers types de ressources. Dans cette thèse, nous proposons une approche novatrice basée sur l'apprentissage profond avec notamment un modèle de réseaux de neurones qui modélise les interactions complexes des patients chroniques avec les ressources hospitalières tout au long de leurs trajectoires de traitement. Cette nouvelle approche propose une série de réseaux de neurones où l'entrée de chaque réseau est définie comme la sortie de prédiction de son précédent. Les modèles proposés sont ainsi capables de prédire le traitement suivant du patient avec une précision (« recall ») allant de 68% à 79%. En plus de prévoir la prochaine étape des traitements des patients, nous proposons aussi une deuxième série de réseaux de neurones qui fournissent le temps prévu pour le prochain traitement. Ces trajectoires temporelles ainsi prédites sont ensuite incorporées dans une simulation à base d'agents capable de prédire la demande personnalisée et agrégée en ressources rares des hôpitaux à moyen et long terme en fonctions des profils des patients à traiter. Nous avons appliqué ce cadre intégratif à des données hospitalières réelles et montrons que le cadre proposé prédit efficacement la demande à moyen et à long terme de ressources rares dans les hôpitaux avec une précision de 77% (trajectoire) et de 64% (délai entre étapes), qui surpasse considérablement à la fois les méthodes traditionnelles de prévision demande et les techniques standard d'apprentissage automatique.

Au niveau tactique et opérationnel, l'article présenté au chapitre 5 propose un modèle intégratif pour la prédiction des durées d'intervention chirurgicale personnalisées. Ce cadre est le premier de ce genre, et permet d'incorporer des attributs opérationnels et temporels liés à la planification, en plus d'attributs liés aux patients, aux procédures et aux chirurgiens pour prévoir ainsi la durée des interventions chirurgicales. De plus, ce cadre illustre l'efficacité d'algorithmes d'apprentissage automatique, tels que « Random Forest » et « Support Vector Machine » pour capturer les relations complexes entre les prédicteurs de la durée des interventions chirurgicales. Nous avons appliqué ce cadre à des données hospitalières réelles et constaté une amélioration de 31% de la précision des prédictions par rapport à la pratique. De plus, les résultats montrent que les décisions liées à la planification telles que l'ordonnancement des procédures et l'affectation des blocs ont un impact significatif sur les durées d'intervention chirurgicale. Ce résultat a des implications importantes pour la littérature dédiée à la planification et l'ordonnancement des salles d'opération aux niveaux tactique et opérationnel. Autrement dit, ce résultat implique que la planification optimale des salles d'opération n'est possible que si l'on optimise conjointement la durée et l'ordre des chirurgies.

Au niveau opérationnel, l'article présenté au chapitre 6 propose un modèle intégratif pour la prédiction du risque de défaillance opérationnelle, et notamment du risque de temps supplémentaire. En pratique, même le plus précis des outils utilisés ne permet pas de prédire la

variabilité des processus hospitaliers avec une précision de 100%. Par conséquent, au niveau opérationnel, il est important d'éviter les décisions qui ont un risque élevé d'échec pouvant ainsi entraîner des conséquences négatives significatives, qui peuvent à leur tour impliquer des coûts supplémentaires, une qualité de soins inférieure, et causer une insatisfaction à la fois des patients et du personnel. Dans cette thèse, nous appliquons des techniques d'apprentissage machine probabiliste au problème des heures supplémentaires en salle d'opération. Plus précisément, nous montrons, en utilisant des données hospitalières réelles, que les algorithmes proposés sont capables de classer les horaires des salles d'opération qui entraînent des heures supplémentaires avec une précision de 88%. La performance des prédictions ainsi calculées est de plus améliorée grâce à l'utilisation de techniques d'étalonnage appliquées aux résultats d'algorithmes d'apprentissage automatique. Le modèle de risque proposé a ainsi des implications significatives à la fois pour la pratique de la gestion les ressources au niveau opérationnel, mais aussi pour la littérature académique. Tout d'abord, le modèle de risque proposé peut facilement être intégré dans les systèmes de planification des salles à l'hôpital afin d'aider les décideurs à éviter des horaires risqués. Deuxièmement, le modèle de risque proposé peut être utilisé conjointement avec les modèles existants d'ordonnancement des salles d'opération pour améliorer la performance opérationnelle des solutions.

ABSTRACT

Canada's health care system is struggling to manage the increasing demand of patients with multiple chronic issues who require access to long-term care, primarily due to Canada's aging population. This has resulted in long patient wait times and increasing healthcare costs. Since hospitals represent the largest share of the healthcare budget, they are required to improve their operational efficiency by making better use of their resources. In particular, hospitals that provide patients direct care and access to expensive resources such as surgeons, operating rooms, ICUs and wards are under scrutiny on whether or not they manage their resources effectively.

Operations research scholars have extensively addressed problems related to resource planning and scheduling in hospitals for many years. Conventional analytical models aim to improve the efficiency of decision-making in hospital resource planning at strategic (long-term), tactical (mid-term) and operational (short-term) levels. However, these models suffer from limited ability in effectively capturing the inherent complexity, variability and uncertainty of hospital operations because they often assume crude and simplistic statistical distributions to imitate these operations. Consequently, they are suboptimal in real-life settings.

With the massive amount of data gathered in the hospitals and healthcare systems and advances in the field of predictive modeling, the operations research community are now given the opportunity to better analyze, understand and replicate the complexity, variability and uncertainty of hospital operations. To this end, the main objective of this thesis is to develop integrate predictive frameworks that are capable of analyzing and extracting important patterns from large-scale data that better inform hospital resource planning and scheduling systems at the strategic, tactical and operational levels. To the best of the author's knowledge, this thesis is a pioneer in proposing frameworks for the design of hospital-wide integrative predictive support systems.

At the strategic and tactical level, the first article (Chapter 4) develops a hybrid machine learning-simulation framework for predicting personalized patient demand for hospital resources. The framework captures the long-term relationship between hospitals and chronic patients, which spans over a long-term horizon and incorporates the fact that patients will need, not one, but several visits to the hospital and access to various types of resources over a long time period. In this thesis, we propose a novel approach based on deep feedforward neural network model that models the complex interactions of chronic patients with hospital resources during their treatment pathways. The proposed novel approach does so

by developing a series of sequential individually trained deep feedforward neural networks, where each network's input is set as the prediction output of its preceding. The proposed models are capable of predicting patient's *next treatment* with an accuracy (measured by "recall") ranging from 68% to 79%. In addition to predicting the transition of patients between treatments in their clinical pathways, we propose a second series of temporal deep feedforward neural network models that provide the expected receiving time for the next treatment. The trained pathway and temporal predictive models are incorporated into an agent-based simulation which is capable of predicting personalized and aggregated demand for hospitals' scarce resources for the mid-term and long-term time horizon. We applied the proposed integrative framework to real hospital data and showed that proposed framework effectively predicts mid-term and long-term demand for hospital scarce resources with an accuracy of 77% and 64%, respectively, which dramatically outperforms traditional demand forecasting methods and standard machine learning techniques.

At the tactical and operational level, the article proposed in chapter 5 is an integrative predictive model for personalized surgical procedure durations. The framework is the first of its kind to incorporate scheduling-related, operational and temporal attributes in addition to patient specific, procedure specific and surgeon specific attributes to predict surgical procedure durations. Furthermore, the framework illustrates the effectiveness of machine learning algorithms such as Random Forest and Support Vector Machine to capture the complex relationships among the predictors of surgical procedure durations. We applied the proposed framework to real hospital data and found an improvement of 31% in the accuracy of our predictive model compared to its practice benchmark. Furthermore, the results show that scheduling-related decisions such as procedure sequencing and block assignment have a significant impact on surgical procedure durations. This result has significant implications for operating room planning and scheduling literature at both tactical and operational levels. Namely, it indicates that optimal operating room planning is achieved only through joint optimization of surgical duration procedures and schedules.

At the operational level, the article presented in chapter 6 proposes an integrative predictive model for operational failure risk assessment. Interestingly, even the most accurate predictive tools used in practice fall short in predicting variability in hospital processes with 100% accuracy. Therefore, at the operational level it is important to avoid decisions that have a high risk of failure which may subsequently result in significant adverse consequences, which, in turn, may incur additional costs, lower quality of care and cause patient and staff dissatisfaction. In this thesis, we apply probabilistic machine learning techniques to the operating room overtime problem. We show that the proposed algorithms are capable of classifying operating room schedules that result in overtime with an accuracy of 88% when applied to real hospital data. The predictive performance is further improved through the use of calibration techniques applied to the output of machine learning algorithms. The proposed risk model has significant implications for practice and operational level resource scheduling literature. First, the proposed risk model can easily be integrated into operating room scheduling systems at the hospital which ultimately assist decision makers in avoiding risky schedules. Second the proposed risk model may be used in conjunction with existing operating room scheduling models to improve the operational performance of commonplace solutions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

CP	Clinical Pathway
CRC	Colorectal Cancer
DRG	Diagnosis Related Group
FFNN	Feedforward Neural Network
MLR	Multivariate Linear Regression
NCCN	National Comprehensive Cancer Network
SBOS	Specialty-Block One-Surgeon
SBSS	Specialty-Block Shared-Surgeons
SDOS	Specialty-Day One-Surgeon
SDSS	Specialty-Day Shared-Surgeons
SVM	Support Vector Machine

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

As of 2017, the total health expenditure of Canada accounted for 11.5% of the gross domestic product (GDP) with $242 billion in total spending (Canadian Institute for Health Information, 2017). Remarkably, as shown in Figure 1.1, hospitals account for the highest share of this health spending (28%), placing them ahead of expenditure on drugs (16.4%) and physician services (15/4%). This trend is expected to continue in coming years. Moreover, Canada is now facing an urgent problem of a rising rate in the number of individuals with chronic diseases, with the aging population. In fact, more than one in five Canadian adults live with at least one chronic illness (Public Health Agency of Canada, 2015). Chronic diseases such as cardiovascular diseases, majority of cancers, diabetes and chronic respiratory diseases require long-term access to chronic care, which is in many cases, provided by hospitals. It is estimated that the direct cost of chronic diseases account for approximately 58% of total healthcare expenditure, rendering them the major source of healthcare costs (Public Health Agency of Canada, 2015).

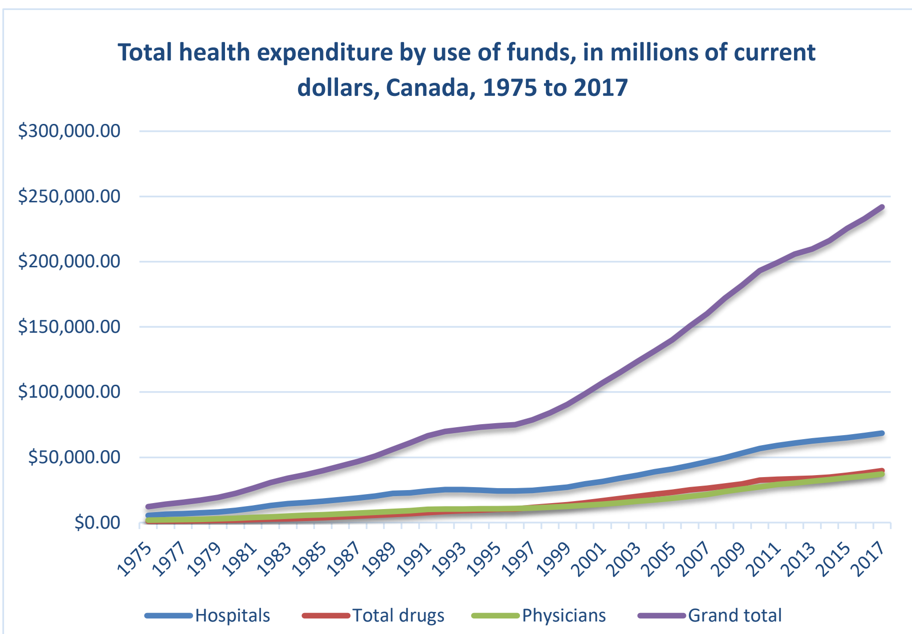


Figure 1.1 Total health expenditure by use of funds, in millions of current dollars, Canada, 1975 to 2017. Data source:National Health Expenditure Database, Canadian Institute for Health Information (CIHI)

Increasing healthcare costs are not the only challenge that Canada's health system faces. According to a report by Fraser Institute, patient wait time for surgical and other therapeutic treatments, mostly provided by hospitals i.e. elective surgeries, reached an all-time record high of 21.2 weeks in 2017 (a 1.2 week increase from 2016) (Fraser Institute, 2017). In most instances, these wait times exceed the wait times deemed "*clinically reasonable*" by physicians. The serious adverse medical consequences of prolonged waiting for treatment has been well-documented in the medical literature (Chen et al., 2008; Warlick et al., 2006). Moreover, in 2016 the economic burden of wait times thorough worker productivity loss was estimated to be approximately $1,759 per patient, when considering *only* the lost working hours of the week. Albeit, this number can reach a staggering $5,360 if we account for the total hours of a week lost, minus eight hours of sleep in the estimation (Barua and Ren, 2016).

With the increase in demand for long-term care, rising costs of chronic conditions, and long wait times there is a growing pressure on healthcare providers to improve their operational efficiency by making better use of their resources. In fact, efficient use of available resources is a policy goal recognized by the federal government, provincial officials, and healthcare decision makers to ensure the sustainability of Canada's health system. The need for efficiency improvement is especially significant in hospitals since they typically utilize the most scarce and expensive of resources such as operating rooms, intensive care units, general wards, and surgeons and nurses. Implementing optimal resource planning and scheduling systems is known to achieve higher efficiency in hospital settings (Hans et al., 2012). However, such systems are effective only when they have a thorough realization of the complexity, variability and uncertainty that is inherent in hospital operations. Considering the large amount of patient data gathered by hospitals and advances in the field of predictive modeling, this thesis presents multiple integrative predictive support systems for hospitals that can be used to design more efficient and robust resource planning and scheduling systems. Using machine learning methods, developed predictive models are capable of realizing complex relationships between patient and hospital attributes – that are otherwise invisible to the eye – to enhance the prediction accuracy of these such support systems.

Resource planning and scheduling in hospital and inpatient settings has been an active subject of research in the field of operations research for many years (Roth and Van Dierdonck, 1995). Several analytical methods have been proposed to provide support for rational and effective decision making at strategic, tactical and operational levels. Despite their different technical approaches and pros and cons, the general consensus among all is to provide insights on how to better utilize and allocate existing resources to improve health outcomes whilst reducing costs. However, conventional optimization methods are limited in their ability to design effective systems, primarily due to the inherent variability and uncertainty of healthcare operations. In fact, as we later discuss in detail in Chapter 2, conventional models must rely on somewhat crude assumptions about these uncertainties. At the strategic level, hospital-wide resource planning decisions that their impact spans over a long-term horizon such as capacity acquisition, budget allocation and hospital case mix, rely on aggregated demand forecast of patients who are heterogeneous in their needs. However, aggregated forecasts fail to consider that, for example, two patients admitted to a hospital for the same diagnosis may require different care procedures due to the differences in their characteristics; and in turn, differ in the amount of specialty care resources (such as general nursing wards or intensive care units) they would demand and ultimately, differ their length of stay (LOS) in the hospital. Therefore, resource planning techniques that solely focus on crude patient classification schemes such as "diagnosis related groups (DRGs)" to determine long-term resource needs of hospital case mix are suboptimal in nature.

At the tactical level, medium-term resource planning decisions focus on issues such as operating room scheduling, staff scheduling, temporary capacity expansions such as bed reallocation and hiring of part-time nurses and staff. Such mid-term resource planning decisions rely on forecasting future demand for patients who are already in their clinical care pathway (or are expected to begin their clinical care pathway), and estimating the service time associated with each unit in demand. For example, a decision planner must anticipate and prepare resources for a cancer patient who will, for example, undergo a 2-hour surgery followed by a one day stay in the ward, and four 1-hour sessions of chemotherapy administrations, three months apart, for the next year.

At the operational level, resource planning decisions are concerned with issues such as surgical case sequencing for an operating block (i.e. where to place each patient in the block) and rescheduling elective surgical cases when the risk of operating room overtime runs high. These decisions rely on not only predicting service times accurately, but also predicting the uncertainty and risks associated with the operations, and preparing accordingly.

The increasing amount of data gathered in healthcare systems, in conjunction with the advances in the field of artificial intelligence and machine learning methods provide a unique opportunity to transform the traditional classification and prediction tools in healthcare into intelligent predictive tools. The main objective of this thesis is to propose integrative predictive frameworks that are capable of discovering informative relationships within the massive amount of knowledge in the data and translate it into intelligent decision support tools that better inform resource planning and scheduling systems. In this thesis, this research objective is addressed through developing integrative predictive support systems that can help improve resource planning efficiency at the strategic, tactical and operational level.

At both strategic and tactical levels, hospital-wide resource planning systems rely on patient classification techniques to predict the long-term and medium term demands for hospital resources. Operations research literature use crude patient classification schemes such as "diagnosis related groups (DRGs)" and "structured clinical pathways" to predict the long-term resource needs of patients. However, these schemes are gravely inaccurate demand prediction tools and not robust in practice because they fail to realize the heterogeneity and variances in patients' extensive (and complex) interactions with various resources and facilities within the hospital during the course of their care pathway (refer to Chapter 2). Therefore, the first specific objective of this paper is to propose and develop an integrative predictive tool for creating a dynamic and personalized clinical pathways for patients, which in turn improves resource planning efficiency at the strategic and tactical level. The tool enables a resource planner to anticipate the flow of each patient within hospital units on an individual level, and predict the specific long and medium-term demand for different hospital resources.

At the tactical level, the efficiency and efficacy of decisions regarding medium-term resource planning and scheduling are limited by the ability of the planning system and decision makers to account for service time variability as well as making accurate demand predictions. Patient service times such as surgical procedure duration, lengths of stay in the ICU or wards, number of visits to specialty resources such as radiology, radiotherapy and chemotherapy, considerably vary based on patients' diagnoses and other medical and operational characteristics. The operations research literature has either ignored this variability entirely or, at best, developed sub-optimal solutions. Some scholars have ignored variations by developing deterministic scheduling systems which assume constant (or diagnosis-dependent constant) service times (refer to Chapter 2). Alternately, others have considered the stochasticity of patient service times by fitting a probabilistic distribution based on analyzing historical data with no attention to the underlying heterogeneity within the population. In any case, conventional prediction tools are unable to handle the complex (and often *invisible* to the eye) relationships among a great number of static and dynamic (time varying) variables. Therefore, they are inaccurate service time prediction tools, produce low quality decisions on resource planning and scheduling, and are not robust in practice and hence, unimplementable. The second specific objective of this paper is to propose an integrative predictive tool for patient service times is proposed. This tool identifies the drivers of variability, and generates effective and practical surgical scheduling decisions.

At the operational level, resource planning and scheduling systems are expected to efficiently and effectively make short-term decisions related to the care delivery operations at individual resource level. The decisions can be made *"offline"* which reflects decisions in advance of the event. Decisions such as scheduling surgical cases to each operating room block, sequencing surgical procedures within each operating room block, scheduling patients to chemotherapy and radiotherapy slots fall into this category. Decisions may also be *"online"*; these decisions are made in real-time due to unexpected events such as rescheduling of surgical procedures to avoid operating room overtime or early transfer of low-risk patients from the ICU to a ward due to arrival of emergency cases. Unfortunately, the inherent uncertainty within these operations contaminates both online and offline decisions, which ultimately takes a severe toll on the operating costs and in some cases, quality of care. For example, premature discharge of patients may result in their readmission with even more severe conditions, or over-scheduling operating room blocks with surgical cases may result in operating room overtime or rescheduling of surgical cases which in turn, increases operational costs and damages both staff and patient satisfaction. Hence, it is crucial for the resource planning and scheduling systems to be able to assess the risks associated with every plausible decision and schedule resources while finding the optimal balance in the trade-off between direct and indirect costs of under and over-utilization of resources. Therefore, the third specific objective of this thesis is to propose an integrative predictive tool that identifies risks associated with decisions at the operational level is proposed and developed. This tool may help decision makers understand the risks associated with any scheduling decision they consider and make necessary adjustments to lower the risk of adverse outcomes to ultimately, obtain a higher operational performance.

The work presented in this dissertation makes the following contributions:

1. This thesis uses artificial intelligence for demand prediction designed specifically for hospital resource planning at strategic and tactical level to address the shortcomings of existing demand prediction methods. Specifically, unlike conventional methods that utilize crude patient classification schemes, this thesis proposes a hybrid machine learning simulation demand prediction framework that is trained using the clinical pathway of previous patients and predicts time-variant aggregated demand for scarce resources in a hospital. The framework is designed to be integrated with hospitals' resource planning and scheduling tools to develop robust and practical policies.

2. This thesis utilizes supervised learning algorithms to predict patient service times for hospital resource planning and scheduling systems at tactical level. The proposed framework is designed to address the shortcomings of the conventional approach of fitting distribution functions to historical data. Our framework incorporates medical and

operational attributes to capture the inherent variability and heterogeneity in patients' service times. The framework is then used to design an integrative surgery duration prediction tool which enables more efficient operating room planning and scheduling.

3. This thesis uses probabilistic machine learning techniques to propose a risk model framework for predicting adverse outcomes for hospitals' resource planning and scheduling at an online/offline operational level. Specifically, this work focuses on predicting the risk of costly overtime operating room schedules. Scheduling systems that rely on conventional probability distributions to capture the uncertainty and variability of surgical procedure durations often face the risk of overtime and in turn, result in suboptimal efficiency. The integrative predictive tool is designed for integration with hospitals' master operating room scheduling systems to identify and flag schedules with high risks of overtime. This allows for offline and online operational decisions that can lower the risk of overtime and lead to better operating room operational performance.

The proposed integrative predictive tools will help hospitals to improve their resource planning and scheduling systems. The proposed tools are expected to enhance the practicability and robustness of existing operations research methods, help to improve the quality of care delivered at the hospital level, and aid hospitals to manage the costs of care delivery by more efficiently and effectively utilizing their scarce resources. We expect that this research will facilitate the design of better resource planning systems for hospitals by leveraging and resurfacing the hidden knowledge in the massive amount of medical and operational data gathered by hospitals using advances in the field of predictive modeling.

The rest of this dissertation is organized as follows: Chapter 2 provides a comprehensive literature review, which includes an overview on the predictive techniques used in healthcare resource planning and scheduling operations and the application of artificial intelligence and machine learning techniques in healthcare. The reviewed articles are classified on the basis of their key applications in the healthcare resource planning and scheduling context. Chapter 3 presents the structure of the dissertation and reviews the methodologies used to accomplish the aforementioned objectives. Chapters 4, 5, and 6 presents the three articles, addressing the aforementioned objectives respectively as follows:

- Chapter 4: **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) Patient Demand Prediction Using a Hybrid Machine Learning-Simulation Approach. Submitted to *Manufacturing and Service Operations Management*.

- Chapter 5: **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) An integrative Framework for Surgery Duration Prediction: A Supervised Learning Approach.

Submitted to *Production and Operations Management*.

- Chapter 6: **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) Risk Of Operating Room Overtime: A Probabilistic Learning Approach. Submitted to *Health Care Management Science*.

Chapter 7 will provide a general discussion of thesis and Chapter 8 will conclude the dissertation and proposes possible avenues for future study.

CHAPTER 2 LITERATURE REVIEW

We begin this chapter with providing a comprehensive review of the predictive models used for strategic, tactical and operational (online/offline) resource planning and scheduling of healthcare operations. The review includes predictive models used in conjunction with other optimization/simulation models developed for resource planning and scheduling purposes. The second section of the chapter provides an overview of artificial intelligence and machine learning predictive models and their application in healthcare operations management, and decision-making tools.

2.1 Healthcare Resource Planning and Scheduling

We rely on the time-dependent taxonomy of resource planning and scheduling to categorize the reviewed articles. The taxonomy was proposed by Zijm (2000) and later applied to healthcare operations by Hans et al. (2012). The framework posits that long-term decisions are made based on aggregated information at hand, and as the decision's time horizon shortens, the information at hand is more granular, allowing for more detailed decision-making. For example, whereas the decision to purchase an MRI machine may be based on the aggregated anticipated demand for the machine on average, the decisions for how to schedule the device over the next month to maximize its utilization requires more detailed demand information such as patient characteristics. Interestingly, the taxonomy also represents the top-down hierarchy of decision making of managers and planners. In the MRI machine for example, the decision to invest in such expensive machine is made by upper management while the decisions on its utilization are made by planners on lower levels. This taxonomy has also used by other scholars to classify the decision-making processes of resource planning in healthcare (Hulshof et al., 2012a; May et al., 2011). Depending on the decision's time horizon, planning can be made for the long-term, mid-term, or short-term which are labeled as "strategic level", "tactical level", "operational level", respectively. In the remainder of this section we provide an overview of the literature who aimed to optimize decision-making on each of these levels.

2.1.1 Predictive models for Hospital's Strategic Level Resource Planning

Long-term planning horizon, also denoted as *"strategic level"* planning, addresses the decisions regarding the hospital's mission as well as the development of the processes of healthcare delivery. Most prominent decisions at this level for hospitals are determining: 1) the case mix which stipulates which patient types will be served in the hospital (patient portfolio), 2) locations of the facility, 3) resource capacity planning such as the number of beds, staff, equipment and etc. The hospital's case mix and resource capacity planning decisions are often jointly considered. This is because resource acquisition decisions are made in response to the anticipated future demand of patients within the case mix; and the hospital case mix is decided according to the available resources, such that patients are provided with timely access to care (Ma and Demeulemeester, 2013; Vissers et al., 2001). Decision making processes at the strategic level require access to highly aggregated information, most importantly, the demand forecast of patients within the case mix.

Joint case mix and resource capacity optimization models have received much attention over the past decade due to fact that many governments across the globe have changed their reimbursement systems such that hospitals are reimbursed for patients, not according to the service provided by the hospital, but by their diagnosis (Ma and Demeulemeester, 2013). Therefore, an optimal case mix will impact a hospital's revenue significantly. To serve the optimal case mix, a hospital is forced to manage its costs by: 1) efficient planning of their most scarce and expensive resources such as nurses, operating rooms and etc. and 2) selecting the case mix of patients such that they are provided the required care, efficiently (Hof et al., 2017). Therefore, the accurate prediction of patient needs within a case mix is an essential input to any optimal optimization model for hospital resources. Underestimation of demand for resources results in low quality of care (Harper and Shahani, 2002), long wait times (VanBerkel and Blake, 2007) and over-utilization of resources (Adan et al., 2009; Gallivan et al., 2002) and hence inefficiency in the system, while overestimation of demand can result in under-utilization of expensive resources (Adan et al., 2009; Green, 2005) and hence loss of revenue (Butler et al., 1996). Nonetheless, due to the complexity of demand forecasting at the strategic level, few papers in the field of operations research have attempted to address the stochasticity of demand in their modeling with respect to case mix optimization and resource planning at a strategic level (Ma and Demeulemeester, 2013). Specifically, three approaches exist that capture demand variability (heterogeneity) across patients. The general approach is to group similar patients together and assume identical trajectories for all patients in each group. Clearly, the implicit assumption is that patients within each group are fully homogenous. These approaches include using: 1) Patient flow simulation 2) Diagnosis-Related Groups classification and 3) Structured clinical pathways.

Patient flow simulation: Simulation of patient flow within the hospital has been used in a few studies to predict aggregated demand for capacity planning purposes at the strategic level (Bekker et al., 2017). VanBerkel and Blake (2007) develop a simulation model for the flow of patients within a surgical unit from the point of diagnosis to the point of discharge. They grouped the elective patients based on their diagnosis code and estimated the average wait time for surgery, and the rate at which new patients enter the system from historical data. Similarly, to support hospital bed capacity planning decision making, Harper and Shahani (2002) propose a simulation model for the flow of patients with eight different diagnoses to forecast the demand of admitted patients for hospital beds using average rate of new patients entering the system in each category.

Diagnosis-related-groups classification: The other approach employed is to estimate the aggregated demand based on classifying patients within the case mix by major diagnostic categories and use the classification scheme to determine an aggregated demand forecast for required capacities for a hospital's most important and scarce resources (Gartner and Kolisch, 2014). The Diagnosis-Related Groups (DRG) classification scheme proposed by the World Health Organization is the most well-known and applied approach, which is also the most used classification method for hospital reimbursement purposes (Fetter and Freeman, 1986).

Structured clinical pathways: Grouping based on patients' clinical pathways is the other method used to predict the demand of patients for hospital resources. Clinical pathways are structured multidisciplinary care plans that determine the medical procedures as well as the sequence in which the patients have to received them during the course of their treatment. Similar to DRG-based grouping, clinical pathways classify patients according to their diagnosis. However, they differ in that clinical pathway classification is more granular in terms of medical accuracy and more comprehensive than the DRG-based classification or patient flow simulation (Chow et al., 2011; Gartner and Kolisch, 2014). On this front, Ozcan et al. (2017) propose a simulation-optimization decision support based on clinical pathways to predict the patient requirements for down-stream hospital resources upon admission. Nevertheless, despite its improvements in granularity, clinical pathway classification fails to capture the heterogeneity in patients' extensive (and complex) interactions with various resources and facilities within the hospital, even within patients with the same diagnosis. Therefore, it results in suboptimal strategic decision making (Leeftink et al., 2018).

2.1.2 Predictive models for Hospital's Tactical Level Resource Planning

The mid-term planning horizon, also denoted as *"tactical level"* planning, addresses the decisions regarding the execution of the processes in the hospital. Specifically, planning on this level involves making decisions regarding questions pertaining what, where, how, when of the delivery of care (Hulshof et al., 2012a). Therefore, on this level the challenge is to optimally decide how to provide care for the determined case mix of patients efficiently, considering the limited available resource capacities determined during strategic level planning. Most prominent tactical level planning decisions for hospitals include: 1) capacity allocation for the hospital's most scarce resources which include operating rooms, beds, nurses and etc. among departments and specialties caring for each category of the case mix and 2) staff scheduling.

Decision making processes at the tactical level deal with the uncertainty of service-time durations and demand fluctuations which cause major complications for optimal resource planning. Therefore, predictive models are required to provide accurate information on: 1) predicting demand of newly admitted patients and downstream demand for patients currently undergoing treatment, 2) service-time of patients scheduled for treatments such as surgeries or their length of stay in a ward or ICU (Gupta and Denton, 2008).

Resource planning at the tactical level requires access to accurate prediction of patient pathway (flow) within the hospital to be maximally effective. However, as discussed previously, due to the complexity of patients' interaction with hospital resources and heterogeneity of patients' needs, most conventional resource planning optimization tools must inevitably assume identical pathways (trajectories) for patients with similar diagnoses; and subsequently, use simple methods to estimate patients' transitions between resources in their pathways. The estimated trajectory is used in queuing models (Armony et al., 2015; Green, 2006; Hall, 2013) and mathematical models (Adan et al., 2009; Zhang et al., 2009) to optimize resource planning. However, since these methods force patients into predefined trajectories solely based on their diagnosis, they are unable to accurately predict patients' pathway and hence are not scalable. For example, the clinical pathway of a heart attack patient with the comorbidity of diabetes may significantly differ from that of an *ordinary* heart attack patient. Moreover, the pathways of patients that require long-term care from the hospital such as chronic patients are highly complexity. Therefore, most of these methods focus on the transition of patients between resources only in a *single* visit to hospital (for example from the emergency room to operating room or ward) (Helm and Van Oyen, 2014). In other words, a patient's return after three months as an example, as part of the long-term treatment plan is not considered. Simulation-based models have been more flexible in modelling the variability of patient flows within a hospital (Harper, 2002; Zeltyn et al., 2011). Nevertheless, despite their higher flexibility, these models still rely on the same simple statistical techniques to estimate patients' transitions between resources and hence, are not scalable.

Regression techniques and time series modeling have been used in a few studies to predict the occupancy level of hospital units. Littig and Isken (2007) propose patient in-flow and out-flow equations using time series and multinomial regression models to estimate the occupancy level of different units of hospital in real-time. However, this work does not account for heterogeneity of patients' trajectories and hence is not generalizable to real hospital settings. Abraham et al. (2009) use a combination of regression and seasonal autoregressive integrated moving average (ARIMA) models to predict occupancy up to a week ahead at the hospital level, while Earnest et al. (2005) use ARIMA to make real-time prediction of ward bed occupancy levels during a SARS outbreak. However, neither of these models account for the interaction between hospital units (resources) and therefore, are not optimal.

Markov models and phase-type distributions have also been used to model patient flow's flow within the hospital. Weiss et al. (1982) develop a continuous-time semi-Markov model to predict the flow of obstetric patients in a hospital. However, the model is only applicable to the flow of patients in a single visit to the ward and lacks generalizability. Marshall and McClean (2003) categorized patients based on their length of stay in the hospital using a Bayesian belief network and propose phase-type distributions to model the flow of geriatric patients within the ward and estimate the bed occupancy level. Their proposed Markov model and phase-type distributions ignore the heterogeneity of patient needs and interaction of resources within the hospital and are designed to model the flow of patient for only a single visit.

Finally, certain patient classification techniques have been used to address patient heterogeneity in patient flow modeling (Harper, 2005). Helm and Van Oyen (2014) use classification techniques such as CART, Kmeans and neural network to determine the DRG and clinical pathway of patients based on their initial diagnosis, and other patient attributes such as age and sex; they classify patients into homogeneous clusters of "patient trajectories". Although this work classifies patients in to the appropriate DRG or clinical pathway using more individual-level patient attributes than previous studies, ultimately, they still suffer from the assumption that patients within the same DRG require the exact same level of care and access to resources.

Capacity allocation and scheduling of expensive and scarce resources with high patient demand and competition among specialties is especially difficult when considering service stochasticity. In the case of operating room scheduling, service times (surgical procedure duration) vary greatly across type of procedure and specialty. Even for a specific type of surgery the service times vary considerably from one patient to another and from one surgeon to another. Employing statistical predictive models that assume a rightly-skewed distribution (in most cases log-normal distribution for surgical duration) are the most well-known approach in the stochastic operating room scheduling literature (Cardoen et al., 2010; Spangler et al., 2004). These models assumed that the surgical procedure durations are independent and identically distributed within each type of surgery, which has proven to not hold in practical settings (Stepaniak et al., 2009). Regression-based methods such as linear and log-linear regressions have been proposed to study the impact of attributes such as surgeon-related, temporal and operational elements on the duration of surgery (Kayis et al., 2012; Stepaniak et al., 2009; Zhou et al., 2016). However, conventional approaches of fitting distribution functions to historical data or using simple linear regressions are not sophisticated enough to predict the surgical durations accurately, rendering operating room scheduling based on such methods severely suboptimal (Cardoen et al., 2010).

Accurate prediction of length of stay of patients in the ICU or in the surgical ward is a vital component of bed allocation and bed capacity planning (Harper, 2002). Similar to surgical procedure durations, conventional predictive methods used to predict a patient's length of stay rely on fitting a log-normal distribution to historical data (Marshall et al., 2005). Linear regression (Littig and Isken, 2007), negative binomial regression (Mallor and Azcárate, 2014) and phase-type (PH) distributions (Gu et al., 2018) are among the other techniques used to predict or classify the length of stay of patients. More recently, machine learning algorithms such as classification and regression tree (Li et al., 2013) and decision trees (Barnes et al., 2015) have been successfully applied to predict the length of stay and have proven to perform compare to their conventional counterparts.

2.1.3 Predictive models for Hospital's Operational Level Resource Planning

Short-term planning horizon, also denoted as *"operational level"* planning, addresses the decisions regarding the execution of the processes in the hospital. Specifically, planning at this level compromises of 1) in-advance detailed decisions at individual resource levels, also known as "offline operational planning", which occurs when elective demand is completely known and the only uncertainty arises from emergency cases, and 2) real-time decisions regarding dealing with unplanned events that may arise during the execution of processes, also known as "online operational planning" (Hulshof et al., 2012a). Most prominent offline operational level planning decisions for hospitals include: 1) patient appointment scheduling which assign individual patients to available time-slots of resources using the schedules developed at the tactical level and 2) staff to shift assignment, which details the day-to-day scheduling of nurses and staff at the hospital. On the other hand, online operational level planning decisions deal with unplanned events such as 1) patient no-shows and cancellation, 2) rescheduling of surgical cases due to arrival of emergency cases to the operating room 3) rescheduling of surgical cases to avoid operating room overtime and 4) dealing with risk of readmission due to early discharge.

Uncertainties at the operational level include 1) service time variability and 2) risk of unplanned events such as patient no-shows and cancellation and emergency demand. At the operational level the main objective is to utilize the available capacity as effectively as possible while considering these uncertainties. Appointment scheduling without consideration of emergency demand as well as the uncertainty of service time, may result in over-utilization of resources and lower quality of care (Cardoen et al., 2010; González-Arévalo et al., 2009). Conversely, appointment scheduling without consideration of risk of no-shows and cancellation of scheduled patients as well as the uncertainty of service time may result in under-utilization of available scarce resources and hence loss of revenue for the hospital and limit accessibility to other patients (Kheirkhah et al., 2015; Moore et al., 2001). Therefore, operations research literature has attempted to tackle this trade-off by incorporating predictive models into their proposed scheduling systems to capture the uncertainty in processes (Gupta and Denton, 2008; Liu et al., 2010).

A rich body of literature exists in operations research dedicated to designing optimal appointment scheduling in the presence of patient no-shows and cancellation and emergency demand. These studies propose offline operational level strategies such as overbooking (Liu and Ziya, 2014; Parizi and Ghate, 2016; Zacharias and Pinedo, 2014) and reserving dedicated resources for emergency patients (Ferrand et al., 2014; van Veen-Berkx et al., 2016a) to mitigate the risk of unplanned events and to avoid succumbing to undesirable online operational level policies such as rescheduling patients at the last minute or having the staff work overtime. The effectiveness of the proposed strategies is then tested by predicting the probabilities of no-shows and cancellation and emergency demand with simple statistical models. Interestingly however, in an empirical study, Norris et al. (2014) show that no-shows and cancellation are in fact, not random, and factors such as lead time, financial payer, patient age, and the patient's prior attendance history impact the probabilities of no-shows and cancellations of patients. Alaeddini et al. (2011) develop a predictive model using logistic regression and Bayesian inference to predict the probability of no-shows using patient characteristics and prior appointments attendance history on an individual patient level. Clearly, effective and efficient decision-making on the operational level requires knowledge of service times. Therefore, similar to tactical level decisions previously discussed, if not accurately predicted and incorporated in the decision-making process, decisions on the operational level are negatively impacted by uncertainty in service times. On the operational level, service time uncertainty can complicate and disrupt day-to-day operations. Due to this uncertainty, many resource scheduling models have been developed in the operations research literature with the objective of minimizing the adverse effect of service time uncertainty. The most prominent example is the operating room's block scheduling methods developed to minimize operating room overtime in the presence of uncertain surgical procedure durations (Adan et al., 2011; Choi and Wilhelm, 2014; Lamiri et al., 2008a). Similar to tactical level resource planning models, at the operational level, most of the previous approaches use statistical models to predict the service time (Rais and Viana, 2011). However, the proposed statistical models fail to incorporate the predictors of service time such as patient characteristics and temporal and operational attributes by assuming that they are random in nature.

2.2 Application of Machine learning and Artificial Intelligence in Healthcare

Considering electronic health records (EHR) and other prominent medical databases, and internal hospital databases, massive amounts of complex, context-dependent and heterogeneous data exist in the healthcare domain which offers a promising ground for improvements in data-driven decision-making. Knowledge discovery on this scale is beyond the abilities of conventional empirical and statistical methods to their overly simple nature, but present an excellent opportunity for machine learning and artificial intelligence algorithms to display their spectacular strength of analysis.

Machine learning and artificial intelligence algorithms have matured in the past decade and are recently used extensively in many fields, including healthcare for complex decision-making problems. Generally, the algorithms are categorized into three large domains depending on how they train their data and what is to be expected of them:

- **Supervised learning:** In this method a labeled set of training data is used to estimate unknown response values. If the response values are continuous the problem is identified as *"regression"* and if the response values are two or more discrete categories the problem is identified as *"classification"*. Supervised learning algorithms include, but not limited to, linear regressions, decision trees such as random forest, support vector machines and artificial neural network.

- **Unsupervised learning:** In this method an unlabeled set of training data is used to either discover and divide the dataset into cluster of similar categories also known as *"clustering"*, or to determine the distribution of data also known as *"density estimation"*.

- **Reinforcement learning:** In this method the algorithm is not provided a fixed labeled but rather interacts with the environment to learn the optimal output through a series of feedback loops of trial and error. In simple terms, the algorithms in this case choose an *"action"* from a set of available actions to arrive at a *"state"* in a sequence of states. Each action taken receive an immediate reward and the series of actions result in an ultimate reward. Examples of application of reinforcement learning are the recent algorithms developed to learn and play chess or "Go" (Silver et al., 2016).

A thorough description of these methods is provided in Bishop (2006).

Recently, scholars have utilized machine learning methods for medical and healthcare data for various purposes. In this section a thorough review of the applications of machine learning methods in healthcare is discussed. The first stream of research has focused on the development of clinical decision support systems which are able to perform comparative effectiveness analysis of different treatments to find personalized optimal treatment for patients (Oztekin et al., 2018; Zhao et al., 2015), predict personalized medical diagnoses (Esteva et al., 2017; Jordan and Mitchell, 2015; Wang et al., 2018) and perform clinical risk assessment for prognosis of specific diseases (Casanova et al., 2013; Kong et al., 2012; Kourou et al., 2015).

A second stream of literature, which is more relevant to our work, has focused on integrating machine learning predictive models with optimization techniques to improve decision-making in healthcare settings. Helm et al. (2016a) develop an integrative density estimation algorithm (unsupervised learning algorithm) to predict the personalized risk of readmission for discharged patients, and incorporate the predictive model in a stochastic optimization model of staff scheduling for post-discharge monitoring of patients. Ang et al. (2015a) proposes a novel method for prediction of emergency room wait times by integrating supervised learning methods with principles of queuing systems. The proposed method, called Q-Lasso, combines the Lasso regression model with fluid model estimators and is capable of predicting emergency department wait times for low-priority patients with greater accuracy than both machine learning methods and fluid model estimators, when used individually. Anjomshoa et al. (2018) use a clustering algorithm to group patients based on surgery duration and length of stay in the hospital. They use the predictive model in combination with a mixed-integer programming model to optimally allocate operating room blocks to surgical specialties. Finally, Ranjan et al. (2017) integrate a clustering algorithm with semi-Markov models to group patients based on their treatment trajectory. The predictive model is used to estimate the transition of patients between multiple wards in the hospital and integrate the model with mixed-integer programming to find the optimal resource schedule in the hospital.

The third stream of literature explores the application of artificial intelligence and deep learning algorithms in analyzing medical history databases such as electronic health records (EHRs) which contain information such as patient demographics, laboratory tests results, diagnosis codes, treatments, prescriptions and clinical notes. Choi et al. (2016a) apply a recurrent neural network model (RNN) called *"DoctorAI"* to sequential and temporal EHR data to predict the diagnosis(s), medication(s) and timing of a subsequent visit of patient. Esteban et al. (2016a) develop a recurrent neural network that embeds both static and dynamic medical information of patients to predict the risk of kidney failure after a kidney transplant. Meanwhile, Pham et al. (2017a) developed a deep neural network based on the long short-term memory (LSTM) method to predict future medical outcomes and use this method to model disease progression, recommend optimal treatment and predict future risk of diabetes and mental health patients.

2.3 Literature Gap

In the previous sections we discussed the most recent and relevant literature on predictive models developed to support hospital resource planning at strategic, tactical, operational levels and provided an overview of the applications of machine learning and artificial intelligence in the healthcare domain. On strategic and tactical level planning, prior literature incorporated variability of demand and by developing heuristics that grouped patients solely based on their diagnosis. Similar approaches have been used to account for variability of service time for tactical and operational level planning.

Conventional predictive models for service time group patients solely based on the type of treatment performed. Other more sophisticated service time prediction models have been developed over the past few years which rely on machine learning algorithms. Although these modules have improved the prediction power significantly compare to traditional models, they are limited to one or few type(s) of diagnosis(es) or procedure(s) and therefore, are not generalizable to all conditions. On the operational level, simple statistical distributions have been commonly used to predict the risk of unplanned events such as patient no-shows, cancellations, and emergency demands. However, the proposed distributions assume that the unplanned events are random in nature. This assumption has been refuted by many empirical studies who indicated the significant impact of patient attributes on such events. Moreover, the strategies that have been developed by stochastic models that aim to mitigate these risks, either reserve more than required capacity (dedicated vs pooled capacities for elective and emergency patients) or increase wait times for patients (overbooking in face of no-shows and cancellation) with the refuted assumption that risks are random; and therefore, are suboptimal.

To summarize, there still exists a wide gap in the literature for optimizing resource planning and scheduling tools on the strategic and tactical levels, risk prediction models on the operational level. Although stochastic models have improved decision making for resource planning and scheduling compared to their static counterparts, neither of these models account for the heterogeneity of patients and service times within the same “diagnostic group”. With the advances in machine learning and databases containing the medical history of patient, operations research scholars have a unique opportunity to take advantage of data and tools at their disposal to develop more accurate resource planning and scheduling systems by incorporating accurate, effective, integrative and comprehensive predictive models. In the next chapters address these gaps and present frameworks for such predictive models.

CHAPTER 3 RESEARCH METHODOLOGY AND STRUCTURE OF THESIS

This dissertation focuses on exploring the use of artificial intelligence and machine learning predictive models to improve resource planning and scheduling in hospitals. Generally, machine learning methods have matured in the past decade and are now capable of extracting much more information from data than conventional regression models. In this thesis, the general objective is to exploit these methods to develop efficient, accurate predictive models to be used in resource planning and scheduling tools. An overview of the specific objectives studied in the dissertation is as follows:

- To develop a predictive modeling tool that may be used as an integrative support system to improve resource planning efficiency at the strategic level;
- To develop a predictive modeling tool that may be used an integrative support system to improve resource planning and scheduling efficiency at the tactical level;
- To develop a predictive modeling tool that may be used as an integrative support system to improve resource planning and scheduling efficiency at the operational level;

In this thesis, we aim to explore and analyze the efficacy of using artificial intelligence and machine learning methods to develop integrative predictive tools for hospital resource planning and scheduling. To do so, this research was conducted in collaboration with Jewish General Hospital (JGH) in Montreal, Canada. This collaboration enables us to thoroughly study the resource planning and scheduling decision making processes in the hospital through numerous interviews with multiple stakeholders such as surgeons, staff and resource planners. We also reviewed the various information technology and analytical decision support tools used in this hospital to facilitate the process of decision making by providing relevant information to managers. The objectives defined in this thesis and the methodologies used to achieve them, are aligned with the actual needs and best interest of the hospital. In addition, JGH provided the data required to build, train and test the proposed integrative predictive tools.

JGH is one of Montreal's largest and busiest acute-care teaching hospitals, with 637 beds and more than 23,000 patient admissions per year. JGH provides care to a diverse patient demographic. It is one of Canada's referral centers for cancer and neonatology diagnoses and hence, admits a large number of patients from across Quebec and other Canadian provinces

(20% of hospitalized patients) as well as providing care for Montreal residence (80% of hospitalized patients). A growing majority of admitted patients in this hospital suffer from chronic illnesses and require a high level of care which is often provided by a multidisciplinary team of health professionals. Hence, the hospital-wide strategic and tactical level resource planning and scheduling systems are of great importance for the hospital. Moreover, one of JGH's most critical resources is its surgical pavilion. It consists of 13 operating rooms, a 40-bed ward and an ICU that performs operation procedures over 14 specialties. In this hospital, high congestion in the surgical pavilion results in surgical procedure cancellations and/or operating room overtimes, which is one major concerns in the hospital. This results from suboptimal and inefficient operating room scheduling at both tactical and operational levels.

The rest of this chapter provides an overview of how artificial intelligence and machine learning predictive models were used in this dissertation to achieve the aforementioned objectives.

3.1 Strategic and Tactical-Level Demand Prediction

The first specific objective of this dissertation is to propose an integrative support system that is capable of accurately predicting the aggregated demand for resource planning systems at the strategic and tactical level. On these levels, effective resource planning requires knowledge of hospital-wide demand forecasts of the case mix of patients for whom the hospital is providing care in the long-term and mid-term. To this end, this study proposes an integrative framework using a hybrid artificial intelligence-simulation approach that learns from the clinical pathways of previous patients to predict the pathways for newly admitted and existing patients. The proposed framework uses historical medical and personal data to predict which patients are most likely to use which resources, and when, in the course of their treatment. We extend the literature on demand predictive modeling by improving patient classification schemes from the traditional DRG-based or structured clinical pathway grouping to personalized and dynamic clinical pathway prediction schemes. First, two networks of feed-forward neural network learning algorithms are developed to 1) build classification models for the clinical pathway of patients and 2) build regression models for the timing of each treatment epoch within the clinical pathway. Second, the trained predictive models are integrated into an agent-based simulation to predict aggregated demand for hospital scarce resources in the medium-term (less than six months) and long-term (beyond six months). Noteworthy is that the proposed sequential temporal neural network model was the ultimate result of an exhaustive search in the pursuit of the most effective predictive model. Many models were designed and redesigned using various machine learning and deep learning techniques such as, but not limited to, vanilla recurrent neural network, short long term memory and hidden markov model . A list of packages used in this work is provided in the Appendix. The framework is applied to colorectal patients under care at JGH. We show that our framework can effectively predict medium-term and long-term demand for hospital scarce resources (surgery, chemotherapy and radiology) with high accuracy.

The proposed integrative framework has been submitted for publication in the Journal of "*Manufacturing and Service Operations Management*" and is presented in Chapter 4 of this dissertation:

- **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) Patient Demand Prediction Using a Hybrid Machine Learning-Simulation Approach. Submitted to *Manufacturing and Service Operations Management.*

3.2 Patient Service Times Prediction

The second objective of this dissertation is to propose an integrative support system that is capable of accurately predicting patient service times for resource planning systems on the tactical level. Generating efficient schedules on the tactical level requires accurate prediction of patient service times in addition to demand prediction. Scheduling systems narrow the scope of decision-making at the tactical level to a single hospital resource such as a ward, or small number of inter-dependent hospital resources such as operating rooms and ICUs. On this level the significance of producing optimal scheduling decisions is especially pronounced for the scarcest of resources such as operating rooms. Therefore, in this study we focus on predicting surgical procedure durations because 1) operating rooms are the largest cost and revenue centers in the hospital, 2) due to high variability in surgical procedure durations, optimal operating room scheduling on a tactical level is extremely challenging. In this study we develop an integrative framework using advanced supervised learning algorithms that model surgical procedure durations. We incorporate patient, surgeon, scheduling and operational attributes to train our model to capture much of the variances. The supervised learning algorithms were developed in R and a list of publicly available packages used in this work is provided in the Appendix. Compared to the current practice of using linear regression, our framework improves the accuracy of the surgery prediction by an average of 31%. This corresponds to a decrease of $\approx$30 minutes in the mean squared error.

The proposed integrative framework has been submitted for publication in the Journal of "*Production and Operations Management*" and is presented in Chapter 5 of this dissertation:

- **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) An integrative Framework for Surgery Duration Prediction: A Supervised Learning Approach. Submitted to *Production and Operations Management*.

3.3 Risk Prediction for Adverse Events

The final objective of this thesis is to propose an integrative support system that is capable of accurately predicting the risks of adverse events associated with the planning and scheduling systems at the operational level. On the operational level, decisions mostly relate to the execution of processes in the short-term and require reacting to and mitigating unplanned events. The uncertainty and variability of processes renders this type of decision-making extremely difficult, especially for processes where any miscalculation may lead to grave consequences such as complicating hospital operations, increasing hospital's tangible and intangible costs and lowering the quality of care provided to the patients. The risk of an operating room running overtime is one of the most important instances in which unplanned events may result in inefficiencies in downstream resources, high labor costs and staff and patient dissatisfaction. Therefore, in this study, we develop an integrative framework that identifies operating room schedules with a high risk of overtime. For this we use probabilistic learning algorithms and calibration techniques to propose a system that learns from previous operating room schedules to identify ones that run a high risk of overtime. The proposed risk prediction system is designed to be integrated with a hospital's operating room scheduling system to identify and flag schedules with high risks of overtime. Applying the framework to the data from JGH, our model is trained to ascertain schedules that run a high risk of overtime. We show that the proposed framework is able to discriminate the high-risk schedules from low-risk schedules with an excellent accuracy of 97.85%. The proposed risk model can provide operating room planners with the overtime risk associated with any operating room schedule they devise. Subsequently, this enables them to improve the operational performances of the operating room scheduling system by making the necessary adjustments to lower the risk of overtime.

The proposed integrative framework has been submitted for publication in the journal of "*Health Care Management Science*" and is presented in Chapter 6 of this dissertation:

- **Karimi, E**., Frayret, J.M., Gendreau, M., Verter, V. (2018) Risk of Operating Room Overtime: A Probabilistic Learning Approach. Submitted to *Health Care Management Science*.

CHAPTER 4 ARTICLE 1: PATIENT DEMAND PREDICTION USING A HYBRID MACHINE LEARNING-SIMULATION APPROACH

Elnaz Karimi[1], *Jean Marc Frayret*[1], *Michel Gendreau*[1], *Vedat Verter*[2]

Submitted to journal of *"Manufacturing and Service Operations Management"*

Abstract: Existing literature on resource planning and scheduling problems rely on crude patient classification schemes such as 'diagnosis-based groups' and 'structured clinical pathways' to predict patient flow and demand for resources within a healthcare setting. However, these methods fail to account for heterogeneity within each 'group' of patients. This paper addresses this shortcoming by incorporating personal and medical patient data to predict personalized patient clinical pathways. The proposed framework captures the long-term relationship between hospitals and chronic patients, which spans over a long-term horizon and incorporates the fact that patients will need, not one, but several visits to the hospital and access to various types of resources over a long-time period. We propose a novel approach based on deep feedforward neural network model that models the complex interactions of chronic patients with hospital resources during their treatment pathways. The proposed novel approach does so by developing a series of sequential individually trained deep feedforward neural networks, where each network's input is set as the prediction output of its preceding. The proposed models are capable of predicting patient's *next treatment* with an accuracy (measured by 'recall') ranging from 68% to 79%. In addition to predicting the transition of patients between treatments in their clinical pathways, we propose a second series of temporal deep feedforward neural network models that provide the expected receiving time for the next treatment. The trained pathway and temporal predictive models are incorporated into an agent-based simulation which is capable of predicting personalized and aggregated demand for hospitals' scarce resources for the mid-term and long-term time horizon. We applied the proposed integrative framework to real hospital data and showed that proposed framework effectively predicts mid-term and long-term demand for hospital scarce resources with an accuracy of 77% and 64%, respectively.

Keywords: Patient Flow Prediction; Clinical Pathways, Demand forecasting; Hospital Resource Planning; Deep Learning; Agent-Based Simulation

---

[1]Department of Mathematics and Industrial Engineering, Polytechnique Montreal University, Montreal, Canada

[2]Desautels Faculty of Management, McGill University, Montreal, Canada

4.1 Introduction

Chronic diseases such as cancer account for 67% of the healthcare budget in Canada. Specifically, they cost Canadians about $65 billion in treatment and $135 billion in lost productivity, adding up to a total $200 billion (Public Health Agency of Canada, 2013). According to the Public Health Agency of Canada (PHCA), approximately half (51.6%) of Canadians over the age of twenty live with at least one chronic disease and four out of five are at risk of at least one chronic illness (Public Health Agency of Canada, 2015). Consequently, healthcare costs threaten to overwhelm provincial budgets across the country (Public Health Agency of Canada, 2013). Hospitals are a major component of healthcare systems that are directly impacted by the economic burden of patients with chronic diseases. Therefore, it has become increasingly critical for Canadian hospitals to better manage their resources, especially considering the increasing rate of aging within the Canadian population.

Increase in healthcare demand has forced healthcare systems and hospitals to manage their capacity holdings with utmost efficiency to remain financially viable (Gupta and Denton, 2008). Moreover, hospital-wide efficiency improvements through matching available capacity and resources with demand have proven to deliver better hospital-wide operational performance, reduce capacity cost, and improve patient outcomes (Harper and Shahani, 2002). Therefore, it is crucial for healthcare systems to implement hospital-wide efficiency improvements in light of the increasing demand which can be achieved through design of robust resource planning and scheduling tools (Butler et al., 1996; Roth and Van Dierdonck, 1995).

One key challenge in developing such sophisticated tools is the accurate prediction of demand for hospitals' most scarce and expensive resources such as operating rooms, beds, radiology and imaging units, etc (Gupta and Denton, 2008; Harper and Shahani, 2002). In fact, "anticipation of needs" has been identified as the 8^{th} rule of the "new rules" for redesigning and improving care (Baker, 2001). Since Harper and Shahani (2002) introduced the idea, and demonstrated the effectiveness of patient classification in accurate demand prediction, many hospitals are using such schemes to anticipate the resources they would need.

Currently, most simulation/optimization models rely on crude patient classification schemes such as "diagnosis related groups (DRGs)" to determine the resource needs of patients (Fetter and Freeman, 1986). Although utilization of these methods is straightforward, they are gravely inaccurate due to the high variances even within each group. In other words, patients - even within the same DRG - are heterogeneous and current classification systems are ineffective in capturing the variances in patients' extensive (and complex) interactions with various resources and facilities within the hospital that are important for simulation and optimization purposes (Isken and Rajagopalan, 2002). In fact, Leeftink et al. (2018) argue that scheduling and resource planning models that do not incorporate variability of care pathways, both across patients, and changes that may occur during the course of pathway, are not robust in practice. In this paper, we propose an integrative framework using artificial intelligence techniques that learns from previous patients' clinical pathways to predict mid-term and long-term demand for hospital resources. We do so by improving patient classification schemes from the traditional DRG-based grouping to a more personalized level. Doing so, our proposed model can provide a significant contribution to the field of healthcare scheduling and resource planning by providing a robust way of incorporating personalized and dynamic clinical pathway prediction.

Currently, much of the literature on scheduling and resource planning in hospitals have limited the scope of its work to a single or small number of similar DRG(s) or type(s) of hospital resource(s) for patients on single hospital visits due to the complexity of handling the large heterogeneity in patients needs (Drupsteen et al., 2013; Gupta and Denton, 2008; Hulshof et al., 2012b; White et al., 2011). Nonetheless, these methods are suboptimal demand prediction schemes at the hospital-wide level for two reasons: 1) many chronic patients with multidisciplinary needs require access to multiple resources during the course of their treatment. For example, while a "typical" heart attack patient may require a routine surgery followed by a two-day stay in the intensive care unit (ICU), the dynamics of this clinical pathway may change if this patient has a chronic co-morbidity such as diabetes, which may complicate the stages of treatment or require further testing resources, which renders DRG-based patient grouping inaccurate in demand prediction; 2) Often a patient's needs are beyond what they would require on that particular hospital visit. For example, a colon cancer patient's treatment may be performed on two (or more) stages separated by a few months, where, on each stage, the patient may require one or more hospital visits. Current demand prediction methods fail to incorporate such long-term outlooks on the patient's treatment stages. Our framework also solves these two issues by incorporating medical and personal data from patients in addition to their clinical pathways and treatments to predict who, with what medical and personal characteristics, are likely to use which resource, and when, in the course of their treatment.

Clinical pathways (CPs)[3] are broadly defined as *"Integrated care pathways are structured multidisciplinary care plans which detail essential steps in the care of patients with a specific clinical problem"* (Campbell et al., 1998, p.1). Nonetheless, due to the complexity of conditions and variances within patients, adherence to such officially structured pathways may not

[3]Clinical pathways are also referred to as "integrated care pathways", "coordinated care pathways" or "care maps."

be practical. Predicting personalized CPs enables a resource planner to anticipate the flow of each patient within the hospital's units, and the time-specific demand for different resources. However, conventional prediction tools such as regression and time series are unable to handle the complex (and sometimes *invisible* to the eye) relationships among a great number of static and dynamic (time varying) variables. However, machine learning algorithms capture this complexity and process large amounts of data using multiple layers of regressions and statistical analysis (Esteban et al., 2016b).

The proposed framework is a hybrid machine learning-simulation predictive tool. In the learning stage, we train a series of feed-forward neural network learning algorithms to build classification models for the clinical pathway of patients and regression models for the timing of each treatment epoch within the clinical pathway. In the simulation stage, we integrate the trained predictive models into an agent-based simulation to predict aggregated demand for hospitals' scarce resources in the mid-term (less than six months) and long-term (less than one year). We design the simulation by first predicting pathway of patients at individual level and then we aggregate their demand for resources during the planning horizon. Using data from the Montreal Jewish General Hospital, one of the leading hospitals in Canada, we demonstrate the effectiveness of our framework in predicting the needs of interdisciplinary patients with chronic diseases that have long-term relationships with hospitals.[4]

We apply our tool to the colorectal patients under care at the Montreal Jewish General Hospital. We show that our hybrid prediction tool can effectively predict mid-term and long-term demand for hospital scarce resources (surgery, chemotherapy and radiology) with an average recall[5] of 75% and 63%, respectively. The rest of the paper is organized as follows. Section 5.2 provides a literature review which covers the state-of-the-art methods for predicting patients' clinical pathways and demand for hospital resources. Section 6.3 explains the fundamentals of feed-forward neural network learning algorithm. Section 5.4 presents the framework for our proposed hybrid integrative prediction model and demonstrates its application on real-world data. Section 6.5 discusses the results of this application. Finally, section 6.6 provides our concluding remarks.

4.2 Literature Review

Various classification techniques have been used in the healthcare scheduling and resource planning literature to group patients and predict their demand for hospital resources. Marynis-

[4]In this paper we focus solely on predicting patients' demand for elective procedures and do not incorporate emergency visits to the hospital

[5]Also known as "sensitivity", or "true positive rate"

sen and Demeulemeester (2016) provide a comprehensive review of these methods. In this section, we focus solely on reviewing the papers that are closely related to our work.

Considering the advances in artificial intelligence and machine learning algorithms over the past decade, a great opportunity has emerged for healthcare operations management scholars to enhance the accuracy of their demand prediction algorithms by extracting more information from data. In fact, several computer science scholars have been able to demonstrate the effectiveness of machine learning algorithms in predicting future clinical events in recent years (Choi et al., 2016a; Esteban et al., 2016b; Funkner et al., 2017).

Casemix patient classification is frequently used, which can be either based on Diagnosis-Related Groups (DRG) or clinical pathways. DRG-based casemix is grounded on the international classification of diseases (ICD-9-CM), specified by the World Health Organization. This casemix categorizes patients based on the diagnosis and treatment procedures (Fetter and Freeman, 1986). On this front, Gartner et al. (2015) use mixed-integer programming to schedule scarce hospital resources for admitted elective patients assuming that the patient DRG and clinical pathway can be predicted in the early stages of treatment. They highlight the importance of early prediction of patients' DRG and clinical pathway in scheduling and how it can significantly improve the hospital's performance and the utilization of its scarce resources. Our proposed method enables such early prediction schemes since our prediction model satisfies their key assumption that clinical pathways can be predicted in early stages of patient admission to the hospital.

The second casemix method is based on the notion of clinical pathways. Predefined clinical pathways have been used in a several papers for the purpose of scheduling and resource planning of elective patients in a multidisciplinary setting. Gartner and Kolisch (2014) propose a discrete optimization model for scheduling elective patients in a hospital-wide setting with the objective to maximize the contribution margin, using both DRG and standardized clinical pathways. Upon applying to real-world data, they show a significant improvement of the contribution margin. However, due to the high complexity of clinical pathways, they only apply their model to patients with clinical pathways that are simple enough to be classified with sufficient accuracy. Ozcan et al. (2017) propose a simulation-optimization decision support tool to improve the performance of hospitals using clinical pathways by better aligning patient requirements with down-stream hospital resources. Their proposed model was tested on a standard surgery-based clinical pathway of thyroid surgical treatment at a hospital. Evidently, variances in clinical pathways of patients have hardly been taken into account in the resource planning and scheduling literature.

Another related stream of research is the recent advances in modeling sequential Electronic

Health Record (EHR) data to predict diagnoses and disease progression within the computer science literature. Choi et al. (2016a) use a recurrent neural network model to develop an intelligent clinical decision support tool called "Doctor AI" to predict the diagnosis codes and medication prescriptions of patients in their *subsequent* visit based on their current visit in an outpatient clinical setting. Using longitudinal patient visit records in the EHR database as the input, they show that the proposed tool is able to achieve above 64% recall in predicting diagnoses of patients' next visit to the clinic and is effective in serving as a diagnosis assistance. Esteban et al. (2016b) apply a recurrent neural network that uses both static and dynamic patient information from patients who undergo a kidney transplant to predict the probability of endpoint clinical events (rejection of the kidney, loss of the kidney and death). Finally, Pham et al. (2017b) propose "DeepCare", an end-to-end deep dynamic neural network that uses the EHR database to predict disease diagnosis, intervention recommendations and future risks.

All three proposed tools are designed to serve as a personalized recommendation system tool for medical decision-making purposes by predicting future medical risks for specific diseases, but are not concerned with the treatment stages and their timing. Our method predict patients' clinical pathway to serve as an integrative predictive tool for hospital resource planning and scheduling.

4.3 Deep Feed-forward Neural Network Structure

Feed-forward neural network, also known as the "multilayer perceptron" (MLPs), was inspired by the efforts to mathematically represent the information processing abilities of biological systems (McCulloch and Pitts, 1943; Rumelhart et al., 1986). Feed-forward neural networks are the quintessential deep learning models in the field of pattern recognition. As presented in Figure 4.1 a neural network is comprised of a network of processing units, arranged in layers (called hidden layers) and connected through weight vectors that map the input vector (x) to the output vector ($\tilde{y}$). Therefore, each neural network model consists of three or more layers: 1)one input layer, 2) one output layer, and 3) at least one hidden layer. The goal of a feed-forward neural network is to approximate a classifier (or regression) function which involves the solution of a nonlinear optimization problem. Below, we provide a description of the two-layer deep feed-forward neural network presented in Figure 4.1.

Neural networks can be described using a series of functional transformations that begins by

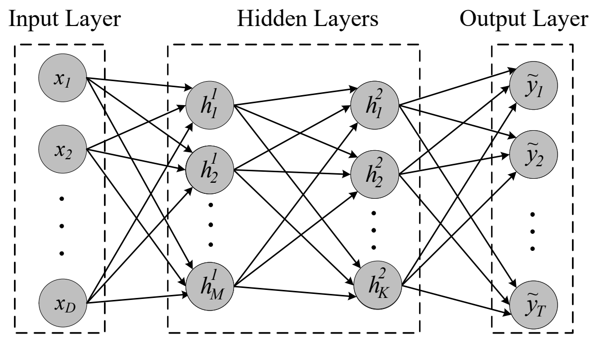


Figure 4.1 Structure of a two-layer feed-forward neural network

constructing M linear equation of the vector of input variables with D features as follows:

$$\Theta_j^1 = \sum_{i=1}^{D} \left(w_{ji}^1 x_i + w_{j0}^1 x_i \right) \tag{4.1}$$

where the subscript $j = 1, \ldots, M$, is the number of units in the first hidden layer and the superscript 1 is the indicator for the first hidden layer. The incoming D arrows represent weights $w_{j0}, w_{j1}, \ldots, w_{jD}$, which are the parameters that are optimized during training of a neural network. w_{j0} is an additional dummy for each neuron i, called bias, which is analogous to the intercept in a regression model, and is used to reposition the linear combination in the N-dimensional space to better represent the distribution of the input. Function Θ_j^1 divides the D-dimensional input's hyperspace into complex regions using a differentiable, nonlinear activation function, called $h(.)$ as follows:

$$z_j = h\left(\Theta_j^1\right) \tag{4.2}$$

where, z_j is the output of hidden unit j in hidden layer 1. These outputs are the inputs of hidden units for the second hidden layer.

$$\Theta_k^2 = \sum_{j=1}^{M} \left(w_{kj}^2 z_j + w_{k0}^2 z_j \right) \tag{4.3}$$

where the subscript $k = 1, \ldots, K$, is the number of units in the second hidden layer and the superscript 2 is the indicator for the second hidden layer. Finally, the output unit activations are transformed using an appropriate activation function $g(.)$ to provide the network outputs y_k.

$$y_k = g\left(\Theta_k^2\right) \tag{4.4}$$

By combining Equation 1 through 4 we can derive the overall network function as:

$$y_k(x,w) = g\left(\sum_{j=0}^{M} w_{kj}^2\, h\left(\sum_{i=0}^{D} w_{ji}^1 x_i\right)\right) \tag{4.5}$$

As shown in Figure 4.1, information propagates only from left to right, which is the reason that, the term "*feed-forward*" is used to describe the computation of a neural network output given an input. As with all supervised learning algorithms, the neural network is trained by minimizing a loss function. The choice of the loss function and the output activation function depends on the nature of the output data. For continuous (regression) outputs, the output activation functions that are often used are "*linear units*" functions, which produce the mean of a conditional Gaussian distribution. The corresponding loss function is the sum-of-squares error (Goodfellow et al., 2016). In this paper we use this structure to predict the time interval between patient's treatments in their clinical pathway.

On the other hand, when the problem is multi-label classification (similar to predicting the next treatment in patient's clinical pathway), the popular activation function is the "*softmax*" function and the corresponding loss function is the "multi-class cross-entropy error" function. The softmax activation function is given by:

$$y_k = softmax\left(\Theta_k^2\right) = \frac{\exp\left(\Theta_k^2\right)}{\sum_k \exp\left(\Theta_k^2\right)} \tag{4.6}$$

and the cross-entropy error function is defined as:

$$E\left(y,\widetilde{y}\right) = -\frac{1}{n}\sum_{x}\sum_{k=1}^{K}\left(\widetilde{y}_k \ln y_k + \left(1-\widetilde{y}_k\right)\ln\left(1-y_k\right)\right) \tag{4.7}$$

As for the choice of hidden units activation functions, the rectified linear units ($h(z) = \max(0,z)$), also known as "*ReLU*", have been shown to dramatically improve the convergence speed of a neural network. However, since the function is not differentiable at 0, for training purposes the derivative at 0 is usually taken from the left side $\left(h^{'}(0) = 1\right)$. Other choices include logistic sigmoid, hyperbolic tangent and radial basis functions. (Goodfellow et al., 2016)

The non-linearity of the activation functions of neural networks cause the loss functions to become non-convex; hence they are usually trained through iterative back propagation with gradient descent. During training, forward propagation continues onward to produce a scalar loss function $E(y)$. The back propagation algorithm refers to the method for computing the gradient by allowing the information from the error function to flow backward through the network (Rumelhart et al., 1986). Given the sum-of-squares error or the cross-entropy error functions, $E : W, X, \widetilde{Y}$,(where W is the matrix of weights that we want to train) the gradient descent requires computing:

$$\nabla_W E(W) = \frac{1}{n}\sum_{i=1}^{n}\Big(\nabla_W\, E(x^i, \widetilde{y}^i, W)\Big) \tag{4.8}$$

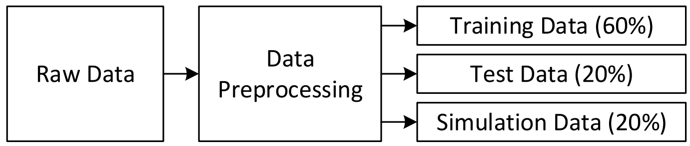
where n is the number of input samples used in training the data. In each step W is updated using the computed gradient and learning rate η. Learning rate η is a hyper-parameter, typically in the range of (0,1), that controls how fast the change of weights affects the actual matrix.

$$W \leftarrow W - \eta\, \nabla_W\, E(W) \tag{4.9}$$

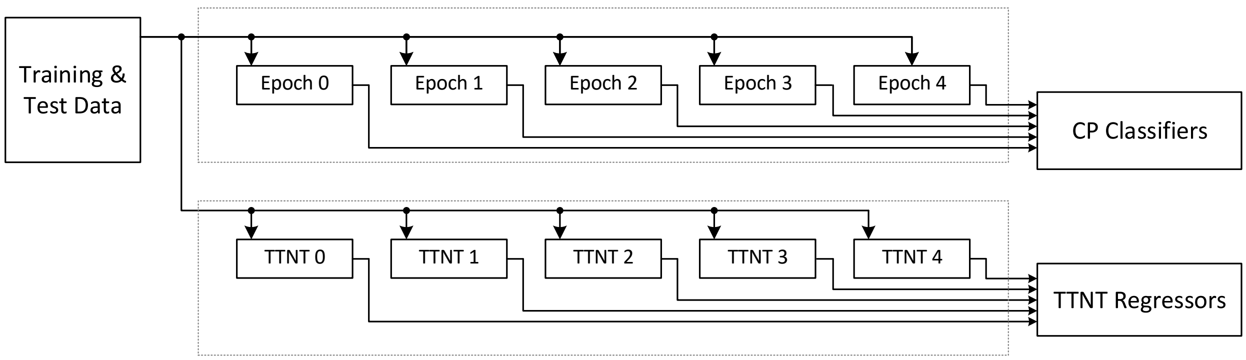
The problem with gradient descent is that its computationally expensive, specifically, when the feature space or the number of samples are large. As the feature space or the number of samples becomes larger, the process time for a single gradient step increases substantially (Bengio et al., 1994). To address this issue, an extension to the gradient descent algorithm called "*Stochastic Gradient Descent (SGD)*" is used. SGD updates the weights after each sample, or a set of samples called "*mini-batch*". During each step of the algorithm a mini-batch of samples are drawn uniformly from the training set $\left(\Psi = \left\{x^1, \ldots, x^{n'}\right\}\right)$. n' is often chosen to include a relatively small number of samples, ranging from one to a few hundred, depending on the size of the training set. (Goodfellow et al., 2016)

4.4 Model

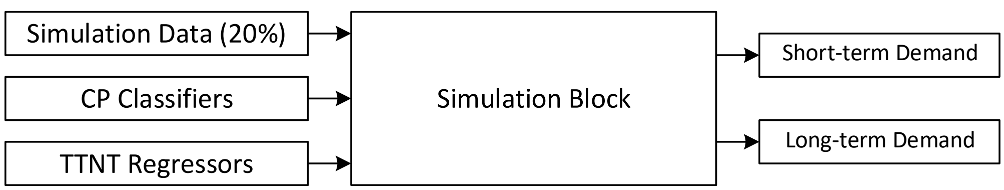
In this section we describe the framework of the proposed hybrid predictive tool of chronic patients' demand for hospital scares resources. First, we develop and train a series of feed-forward neural classifiers to predict treatments at epochs of patient' CP. Second, we develop and train a series of corresponding feed-forward neural regressors to predict the time interval between CP's epochs. Third, we use the trained models on a separate dataset to predict individual patient' demand for scares resources over the course of their pathways. We then use a simulation model to sync the prediction results in respect to planning horizon and aggregate patients' demand and show how the proposed predictive tool can forecast patients demand for hospital's resources in both short and long-term periods with high accuracy. We compare the results of the simulation with the actual observed demand and find that in the mid-term our model can predict demand with 75% percent recall on average, while, in long-term the recall average drops to 63%. Finally, we demonstrate how the proposed hybrid predictive tool can be used by hospital resource planners to devise more effective capacity and resource planning strategies.



(a) Data preprocessing stage



(b) CP classifiers and Time To Next Treatment (TTNT) regressors training stage



(c) Aggregated demand prediction with simulation

Figure 4.2 General framework of the proposed hybrid predictive tool

Figure 4.2 illustrates the high-level architecture of our hybrid model. We first pre-process the hospital's patient database to prepare the data for the learning phase. The preprocessing algorithm includes dealing with missing data, scaling and normalizing for numerical features and one-hot encoding for categorical and label features and cross referencing and grouping of treatment's procedure code. We then divide the processed data to train and test sets by taking 60% of the data as the training set and leaving a sufficient 20% to test the performance of the model. We will preserve the remaining 20% "new patient" to generate the input for our aggregated demand forecasting simulation. The "new patient" samples are used to simulate patients CPs as well as their mid-term and long-term demand for hospital resources based on their predicted treatment label and the time to the patients next treatments (Figure 4.2a).

Patients receive treatments at most at N epochs in their pathways. For example, consider the scenario where, patient X and patient Y start their pathway with undergoing the same surgical procedure. However, patient X leaves the hospital afterward, hence terminating her pathway, while patient Y needs to receive additional chemotherapy treatment and then leave the hospital. Therefore, the total number of epochs in patient' *X* treatment pathway is 1, and for patient Y is 2. The highly variable number of treatment epochs across patients justifies the training of distinct neural networks at each epoch. Moreover, patient treatment history at each epoch provides useful information for predicting future epochs. Note that at each point of time, the hospital faces demand from patients at different stages of their CPs. Hence, to predict the future demand at each point of time, for each patient present in the database, we identify their current epoch number and use the trained neural network models to predict their future epochs (Figure 4.2b).

The trained neural networks are then used in the demand forecasting simulation (Figure 4.2c). For illustration purposes, Figure 4.3 presents the process of demand prediction for an individual patient. As illustrated in this figure, patient Z has already received two treatments in the initial time period of the simulation and is awaiting her third treatment. To predict patient Z's remaining CP, we first use the epoch-2 neural network to predict the third treatment. Next, to predict when this treatment is required, we use the "time to next treatment" neural network. Subsequently, the predicted third treatment becomes an input to the epoch-3 neural network as a means to predict the fourth treatment and so on. Once a neural network at any epoch predicts "No Treatment" for a patient, we terminate the patient's pathway at that point. Once the future pathway of all individual patients are predicted, we then aggregate demand across all patients. In order to predict demand on a monthly basis, which is the more practical resource planning agenda utilized in hospitals, we translate the time-to-next-treatment outputs from "days" to "months" by simply dividing the time-to-next-treatment output by 30. Finally, we provide a list of anticipated demand on a monthly basis for each resource in the hospital.

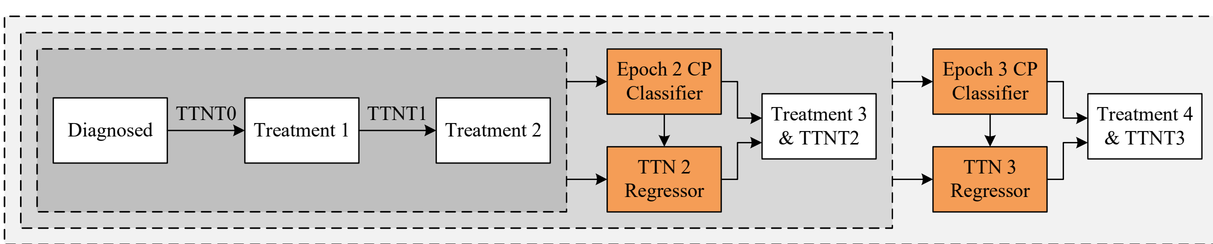


Figure 4.3 Process of demand prediction for an individual patient

4.4.1 Colorectal Cancer

The target population of this study is the Colorectal Cancer (CRC) patients at Montreal's Jewish General Hospital (JGH). JGH is one of Quebec's largest and busiest acute-care teaching hospitals, with 637 beds, more than 23,000 patients' admission on an annual basis. CRC care program in JGH consists of an inter-disciplinary team with expertise in medical and radiation oncology, gastroenterology, nursing, cancer genetics, cancer prevention, psychosocial support and palliative care medicine. The dataset includes 3,082 observations for 667 colorectal cancer patients that were under care from 2012 to 2014. Although we apply our framework to CRC patients, it can easily be applied to other chronic diseases.

CRC is the third most commonly diagnosed cancer in Canada and is the second leading cause of cancer-related death in men and the third among women. In 2017, an estimated 26,800 Canadian were diagnosed with CRC and 9,400 died from this disease (Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2017). The estimated five-year survival rate is 92 % in individuals diagnosed with Stage I of CRC, 69% in individuals diagnosed with stages II and III and only 11% in individuals with stage IV. (Siegel et al., 2017)

4.4.2 Clinical Pathways of CRC Patients

Once patients are initially diagnosed with CRC, the decision in regards with the appropriate CPs is made by an interdisciplinary team within the hospital, based on the state of cancer, the general health of patients, risks and adverse effect associated with treatments, anticipated quantity (life-years) and quality of life, etc. It is important to note that the initial predicted CPs could drastically change during the course of treatments of patients, based on change of status of state of patients, patients response to treatments, etc.

A critical review of CRC clinical guidelines developed by National Comprehensive Cancer Network (NCCN) and Cancer Ontario illustrates the great variation across CRC CPs across patients. For example, NCCN guidelines for patients in stage II colon cancer (Stage IIA: T3, N0, M0, Stage IIB, T4a, N0, M0, Stage IIC: T4b, N0, M0) (refer to Appendix for more information) are categorized as either low-risk (T3 lesion) or high-risk (T4 lesion). After reviewing the pathology report and staging the cancer, surgeon determines if the patient is operable. If operable, the patient is scheduled for colectomy. After the surgery, the patient's images are send to pathology, if surgery was successful, low-risk stage II patients proceed to the cancer follow-up care pathway, however the high-risk stage II patients with complete resection of colon cancer are referred to a medical oncologist and are considered for sessions of adjuvant chemotherapy and then proceed to cancer follow-up care pathway. The recommended chemotherapy protocols by Cancer Ontario and NCCN for Stage II high-risk patients remains controversial. However, NCCN recommends FOLFOX for high-risk Stage II. 5-FU + Leucovorin protocol (also known as Mayo clinic protocol) is also used at this stage. The medically inoperable patients are referred to medical oncologist for appropriate palliative chemotherapy or radiation therapy.

Meanwhile, patients in Stage IV (IVA, IVB) are first considered for colon resection only if there is an imminent risk of destruction or significant bleeding. If liver and/or lung metastases exist and are resectable, the patient is either directly considered for staged resection of metastatic and colon cancer or is first referred to a medical oncologist for neoadjuvant chemotherapy and then is scheduled for surgery. After the surgery, the patient is considered for adjuvant chemotherapy. If the liver and/or lung metastases are potentially resectable, first the patient proceeds to chemotherapy. After the chemotherapy treatment patient is re-evaluated for resectability; if resectable, the patient is scheduled for staged resection of metastatic and colon cancer. Subsequently, patient is scheduled for adjuvant chemotherapy. However, if not resectable the patient is scheduled for palliative chemotherapy and radiation therapy. Reviewing patients profile at JGH revealed various pathways. Pathways such as Radiotherapy+ Protocol Mayo and then resection, just resection, chemotherapy protocols such as MAYO, FOLFOX6, FILFRI, before and after resection and in some cases combinations of radiotherapy and chemotherapy protocols were also observed.

For rectal cancer patients in stage II and III, different CPs are prescribed by NCCN. After reviewing the pathology report and staging the cancer, the surgeon needs to decide whether the cancer is resectable or not. If resectable, the patient is referred to radiation oncologist and medical oncologist for preoperative therapy which includes preoperative chemo-radiotherapy or preoperative hypo fractionated radiotherapy alone. After preoperative therapy, the patient is scheduled for resection surgery. After the surgery, the patient is referred to pathology for further tests and thereafter to adjuvant chemotherapy, if necessary. However, if not resectable, the possibility for down-staging the cancer with chemo-radiotherapy is assessed. If possible, patient is referred for chemo-radiotherapy while being re-evaluated for resectability by allowing adequate time for down-staging. If down-staged the patient is scheduled for surgery, however, if down-staging is not successful,the patient is instructed palliative chemotherapy. If there is no possibility for down-staging the cancer with chemo-radiotherapy, the patient is instructed for palliative radiation with or without chemotherapy.

The review of clinical guidelines for CRC CPs, shows that patients treatments and CPs are highly complex; specifically, for developing predictive support tools of hospitals' resources demand for chronic patients in large scales. Hence, we argue that the proposed model here can provide decision makers (without specific medical knowledge of diseases and their treatments) with a tool that can be easily integrated to hospital systems and provide support for hospital's resource planning.

4.4.3 Data

CRC patients interact with various resources within the hospital in their trajectories, such as medical and radiation oncology, gastroenterology, nursing, cancer genetics, cancer prevention, psychosocial support and palliative care medicine. After extensive conversations with the medical experts in the hospital, we select the three most common departments with scarce resources that CRC patients have the most interactions with and also have to compete over with other patients in the hospital: surgical unit, medical oncology and radiation oncology department. We have, therefore, limited our analysis to these three resources.

Various factors play parts in the selecting the correct CP for colorectal cancer patients. Inputs include patient's attributes including age, sex, town of residence, location of initial diagnosis and designated physician, surgeon and oncologist (if assigned) and disease specific attributes such as diagnosis codes, designated physician, surgeon and oncologist, date of initial diagnosis, clinical and pathological staging TNM, visit to the hospital and procedures. All codes were timestamped with the patients visit time. (Table 4.1 provides a summary of diagnosis counts for cancer types found in our dataset)

Table 4.1 Colorectal Cancer Diagnosis

Diagnosis	count
Anal	23
Colon	341
Rectosigmoid	44
Rectal	257
Grand Total	665

4.4.4 Data Preprocessing

There are 95 unique surgical procedures, 37 unique chemotherapy protocols and 18 radiotherapy protocols. However, many of these unique procedures are very granular, therefore, to predict diagnosis and CPs, we cross-referenced codes into higher-order categories. For the surgical procedure codes, we use the Canadian International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10-CA), yielding 12 unique codes. For chemotherapy and radiotherapy protocols we grouped the protocols and ignore additive prescribed drugs. There exist 10 unique chemotherapy protocol groups and 3 radiotherapy protocols, yielding 25 total treatment types. This number varies across CP epochs from 14-18. All given treatments were timestamped with the patients visit time. If a patient received multiple treatments in a single visit, those codes were given the same timestamps (Detailed information on colorectal cancer characteristics and its associated treatment procedures are provided in the Appendix).

In order to deal with missing data we experimented with both mean imputation and median imputation. We find that median imputation provides more accurate predictions and hence we replaced all missing values with median imputation in our model.

In addition, for numerical data we experimented with both scaling and normalization of the numerical features and found that normalization yields better results. We use one-hot encoding to create multi-hot label vectors to represent the categorical features. Furthermore, because time to the next treatment in the CP can be highly skewed, we define a new feature as the logarithm of the time duration between treatments and train all regression models to predict the logarithm of the time duration between treatments.

4.4.5 Model Training

For training purposes of our predictive models including the baselines, we used 60% of the patients samples as the training set and 20% as the test set. The performance of all models was evaluated against the test sets to avoid overfitting. The training was performed for hidden units of sizes of 15 to 50, however, since the performance started to saturate around 20, we set the hidden units size to 20 to avoid overfitting of the trained model. We use the k-fold cross-validation method for training purposes. In this method, the dataset is split into k equal-sized parts and one randomly selected part is used as the test set and the remaining $(k1)$ parts are used as the training dataset.

4.4.6 Model performance

Our CP classification feed-forward neural network models predict the treatment label that patient will receive at each epoch of their clinical pathway. Our training examples $S = ((x_1, \widetilde{y}_1), \ldots, (x_N, \widetilde{y}_N))$ with instances $x_i \in X$ and a set of treatment labels. We define the label set as follows

$$T = \{t_1, \ldots, t_N\} \tag{4.10}$$

As in previous sections we define the actual treatment vector $\widetilde{y}$ consists of N elements for N test patients:

$$\{\widetilde{y}_1, \ldots, \widetilde{y}_N\} \in T \tag{4.11}$$

and our multi-class feed-forward neural network models generates a prediction vector y of N elements

$$\{y_1, \ldots, y_N\} \in T \tag{4.12}$$

Since our problem consists of a moderate number of classes (14-18) at each treatment pathway epoch, we evaluate the performance of our classifiers using the confusion matrix. Confusion matrix rows represents predicted classes and the columns represent the actual classes. Thus, element α_{ij} of the matrix indicates the number of observations with actual class j which were predicted as class i. If $i = j$ then the prediction is accurate and if $i \neq j$ then α_{ij} indicates the number of misclassified samples. We define the confusion matrix as follows:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots \\ \vdots & \ddots & \\ \alpha_{T1} & & \alpha_{TT} \end{bmatrix}$$

Based on the confusion matrix we evaluate the performance of our classifiers on each class by defining the metric "Class Prediction Accuracy" as the portion of correctly classified observations from class t:

$$CPA_t = \frac{\alpha_{t,t}}{\sum_{i=1}^{T} \alpha_{i,t}}, t = 1, 2, \ldots T. \tag{4.13}$$

We then average the accuracy over all classes to compute the average per-class effectiveness of our classifiers. However, the accuracy metric is sensitive to the prior distribution of the classes (i.e., some treatments are more common than others), and hence does not fully describe the difficulty of the classification problem when faced with highly imbalanced data. Therefore, following the literature, we use the Recall metric to evaluate the performance of our classifiers. The recall for class k (on the test sample) is the fraction of examples from class t_i that are correctly predicted by our classifiers.

$$Recall_{t_i} = \frac{\sum_{n=1}^{N} \mathbf{1}\left(y_n = t_i, \widetilde{y}_n = t_i\right)}{\sum_{n=1}^{N} \mathbf{1}\left(y_n = t_i\right)} \tag{4.14}$$

To evaluate the performance of the time-to-next-treatment neural network we use the *coefficient of determination* (R2), which compares the accuracy of the prediction with respect to the simple prediction by the mean of the target variable.

$$R^2 = 1 - \frac{\sum_{i=1} (y_i - \widehat{y}_i)^2}{(y_i - \overline{y}_i)^2} \tag{4.15}$$

4.5 Results

We limit the training to the first five treatments of the clinical pathway. We do so because only 1.1% of our patients receive more than five treatments in their CP and of those patients all receive their treatments in the second year following their initial diagnosis. Since we have limited the time period of our resource demand prediction to a maximum of one year time span, our decision is justified.

Table 4.2 presents the performance of our feed-forward neural networks: FFNN-1 (using one hidden layer) and FFNN-2 (using two hidden layers) classifiers, compared to their counterpart multinational logistic regression (MLR) classifiers. MLR is a common way to perform classification tasks, however, it does not account for complex relationships among input vectors specifically for high-dimensional categorical input variables. FFNNs outperform the MLR at all epochs both in terms of accuracy and recall. The performance of our FFNNs classifiers is decreasing in epochs, which demonstrates the increased complexity of treatments if the initial treatments does not yield favorable results. However, the performance achieved is still very effective in predicting the patient's treatment pathway with recall ranging from 68% to 79%. The results also confirms that having multiple layers when using FFNN improves its ability to learn more efficient representations of the patterns.

Table 4.2 Performance of Epoch-0 to Epoch-4 Classifiers in forecasting Next Treatment

	Epoch-0		Epoch-1		Epoch-3		Epoch-4		Epoch 5	
	CPA	Recall	CPA	Recall	CPA	Recall	CPA	Recall	CPA	Recall
FFNN-1	64.31	77.92	68.41	74.21	61.41	71.42	59.41	67.67	59.18	64.27
FFNN-2	65.11	78.81	68.93	75.44	62.32	73.21	60.02	69.41	61.32	67.94
MLR	37.12	41.23	35.41	38.17	34.21	36.53	29.21	32.31	24.32	29.41

Table 4.3 compares the performance of time-to-next-treatment FFNN-1 (using one hidden layer) and FFNN-2 (using two hidden layers) to the performance of a Multivariate Linear Regression (MLR) , designed to predict the time duration between two consequent treatments in the CP of patients. The results indicate that our approach significantly improves the accuracy of predicting the time duration until the next visit compared to the counterpart.

Table 4.3 Performance of time-to-next-treatment regressions

	Epoch-0	Epoch-1	Epoch-3	Epoch-4	Epoch 5
	R^2	R^2	R^2	R^2	R^2
FFNN-1	0.3476	0.3554	0.3211	0.2519	0.2653
FFNN-2	0.3641	0.3572	0.3192	0.2581	0.2675
MLR	0.1131	0.1281	0.1866	0.1007	0.1242

4.5.1 Hospital's Resources Agent-based Demand Forecast Simulation

In this section, we discuss the proposed agent-based demand forecast simulation. We use the trained CP FFFN-2 classifiers and the time-to-next-treatment FFNN-2 regressions from the previous section to simulate the CRC patients' demand to access resources in the surgical department, radiation oncology and cancer treatment unit (responsible for administrating chemotherapy) at JGH.

We utilize the remaining 20% of the previously unseen data from the period of January 2013 to December 2013 to demonstrate the efficiency of our proposed tool. The simulation algorithm is depicted in Figure 4.4. As shown, we divide our time period into 12 months and attempt to predict demand for each month since hospitals plan their resources on a monthly basis. In Figure 4.4, each t represents one month of the simulation and $t = 1$ and $t = 12$ represent January and December, respectively. For each $t = k$, our resource-planning tool predicts demand for the months $k, \ldots, 12$ ($t = k$ and onwards).

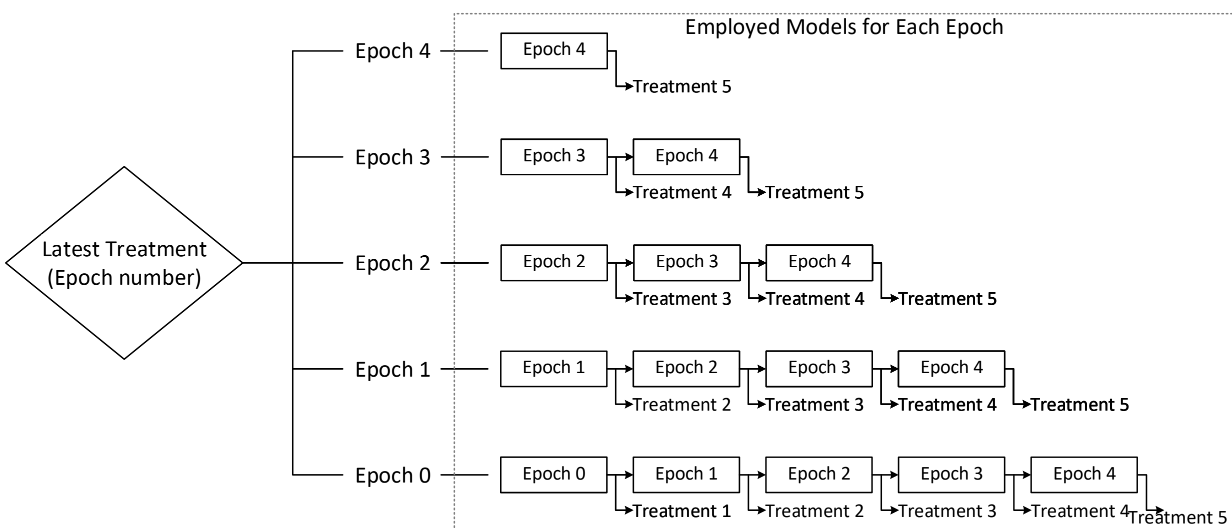


Figure 4.4 Agent-based demand forecast simulation

At the beginning of each month of the simulation, the tool receives an input dataset comprising information on patients undergoing care only at that month in the hospital. In other words, to mimic a real-world resource planning event, for each month we omit subsequent months' data from the analysis. Note that these patients are at different epochs of their CP and therefore, their future demand for resources (remaining treatments in their pathway) depends on their treatment history as well as their disease attributes and personal characteristics. Hence, initially, the algorithm identifies the most recent treatment (epoch) of the patient's clinical pathway (Figure 4.4). Once each patient's most recent epoch is identified as $j\,(\in j = 1, \ldots, 5)$, the algorithm calls the trained CPFFN-2 classifiers of epoch-j, ..., epoch-5 and the corresponding time-to-next-treatment FFNN-2 regressions of $epoch - j, \ldots, epoch - 5$ to predict the patient's remaining pathway. The algorithm is repeated for all patients in the dataset.

Table 4.4 presents the performance of the classifiers in predicting the complete sequence of the CP of patients, averaged over the 12 periods. As expected, the accuracies of the models have declined in comparison with the prediction models of the previous section. The reason is that whereas the treatment prediction models of the previous section were derived using actual patient data, in this section the output of each epoch feeds the input of its proceeding stage. Therefore, any error in any stage propagates through all subsequent stages, and consequently, hurt the prediction accuracy of future treatments. Therefore, the proposed tool performs much better in predicting mid-term demand (less than 6 month in the future) compared to long-term demand (6 to 12 months in the future).

Table 4.4 Performance of classifiers in predicting the sequence of treatment in patient's CP

	1st-Treatment		2nd-Treatment		3rd-Treatment		4th-Treatment		5th-Treatment	
	CPA	Recall	CPA	Recall	CPA	Recall	CPA	Recall	CPA	Recall
Epoch=0	65.43	77.15	61.63	73.32	57.41	68.67	56.41	63.41	55.12	64.91
Epoch=1			66.18	74.65	67.71	73.21	62.14	69.33	59.12	67.71
Epoch=2					64.21	71.22	65.12	67.43	58.12	61.67
Epoch=3							58.42	62.31	60.01	64.13
Epoch=4									59.98	65.34

For illustration purposes Figures 4.5, 4.6 and 4.7 presents the results of our simulation for predicting demand for all three resources, starting from May 2015.

Surgical Unit Demand								
Month	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13
Total Demand	21	17	14	18	17	19	17	15
Forecasted Surgical Procedures	C20.9.20	C19.9.20	C20.9.20	C19.9.20	C18.9.20	C18.9.40	C20.9.20	C18.9.20
	C20.9.20	C19.9.20	C20.9.20	C20.9.20	C18.9.20	C20.9.20	C18.9.20	C18.9.50
	C20.9.20	C20.9.20	C18.9.20	C20.9.20	C20.9.20	C18.9.20	C18.9.40	Hepatectomy
	C20.9.20	C20.9.70	C18.9.20	C20.9.20	C19.9.20	C18.9.20	Hepatectomy	C18.9.40
	C20.9.20	C18.9.20	C18.9.20	C20.9.20	C18.9.40	C19.9.20	C19.9.20	C20.9.70
	C20.9.20	C18.9.20	C18.9.20	C20.9.20	C18.9.40	C18.9.50	C20.9.20	C18.9.40
	C20.9.30	C18.9.20	C18.9.20	C20.9.20	C18.9.20	C20.9.20	C18.9.40	C20.9.20
	C20.9.70	C18.9.40	C18.9.20	C20.9.20	C18.9.20	C18.9.20	C18.9.20	C18.9.40
	C20.9.70	C18.9.40	C18.9.40	C20.9.20	C18.9.20	C19.9.20	C18.9.50	C18.9.50
	C18.9.20	C18.9.40	C18.9.40	C20.9.30	C18.9.40	C18.9.20	C20.9.70	C19.9.20
	C18.9.20	C18.9.40	C18.9.40	C18.9.20	C20.9.20	C20.9.20	C18.9.40	C20.9.20
	C18.9.20	C18.9.40	C18.9.40	C18.9.40	Hepatectomy	C20.9.20	Hepatectomy	C18.9.40
	C18.9.20	C18.9.40	C18.9.40	C18.9.40	C20.9.20	C18.9.50	C20.9.20	C18.9.40
	C18.9.40	C18.9.40	C18.9.40	C18.9.40	C18.9.40	C18.9.20	C20.9.20	C20.9.20
	C18.9.40	C18.9.40		C18.9.40	C18.9.40	C18.9.40	C20.9.20	C18.9.20
	C18.9.40	C18.9.40		C18.9.40	C20.9.20	C20.9.20	C19.9.20	C18.9.70
	C18.9.40	C18.9.40		C18.9.40	C18.9.20	C20.9.70	C20.9.20	C18.9.40
	C18.9.40			C18.9.40		C20.9.20		C18.9.20
	C18.9.40					C18.9.50		C20.9.20
	C18.9.40							C19.9.20
	C18.9.70							

Figure 4.5 CRC patients demand forecast of surgical unit for the period of May-2013 to December-2013

Chemotherapy Unit Demand								
Month	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13
Total Demand	21	12	22	15	24	15	25	13
Forecasted Chemotherapy Protocols	FOLFOX 6 Protocol	Fluorouracil	FOLFOX Protocol	Irinotecan	Fluorouracil	Fluorouracil	FOLFOX Protocol	FOLFOX Protocol
	Fluorouracil	Mayo Protocol	FOLFOX 6 Protocol	FOLFOX Protocol	Irinotecan	Oxaliplatin	SAI	FOLFIRI Protocol
	FOLFOX 6 Protocol	FOLFOX Protocol	Capecitabine	Irinotecan	FOLFOX 6 Protocol	FOLFOX Protocol	FOLFIRI Protocol	Fluorouracil
	FOLFOX Protocol	FOLFOX Protocol	FOLFOX 6 Protocol	FOLFIRI Protocol	FOLFOX Protocol	FOLFOX 6 Protocol	5-FU Protocol	Oxaliplatin
	FOLFOX 6 Protocol	Fluorouracil	Fluorouracil	FOLFOX Protocol	XELOX Protocol	FOLFOX 6 Protocol	FOLFOX 6 Protocol	FOLFOX Protocol
	5-FU Protocol	Oxaliplatin	Oxaliplatin	Fluorouracil	SAI	SAI	FOLFOX Protocol	FOLFOX 6 Protocol
	FOLFOX 6 Protocol	Irinotecan	FOLFOX 6 Protocol	Irinotecan	FOLFOX Protocol	Fluorouracil	Fluorouracil	Fluorouracil
	Fluorouracil	Fluorouracil	FOLFOX Protocol	Irinotecan	Fluorouracil	Irinotecan	Oxaliplatin	Oxaliplatin
	Fluorouracil	Oxaliplatin	Fluorouracil	FOLFIRI Protocol	Oxaliplatin	Fluorouracil	FOLFOX Protocol	FOLFOX Protocol
	5-FU Protocol	Fluorouracil	Oxaliplatin	FOLFOX 6 Protocol	FOLFOX 6 Protocol	SAI	Oxaliplatin	Irinotecan
	Fluorouracil	Oxaliplatin	FOLFOX 6 Protocol	SAI	SAI	FOLFOX 6 Protocol	FOLFIRI Protocol	FOLFOX 6 Protocol
	Oxaliplatin	FOLFOX 6 Protocol	FOLFOX 6 Protocol	FOLFOX 6 Protocol	SAI	FOLFOX 6 Protocol	FOLFIRI Protocol	SAI
	FOLFOX 6 Protocol		Fluorouracil	FOLFOX 6 Protocol	Irinotecan	FOLFOX Protocol	Fluorouracil	FOLFOX 6 Protocol
	Fluorouracil		Irinotecan	SAI	Irinotecan	FOLFOX Protocol	Irinotecan	
	SAI		Capecitabine	FOLFOX 6 Protocol	Fluorouracil	Capecitabine	Irinotecan	
	Mayo Protocol		Fluorouracil		Oxaliplatin		FOLFOX 6 Protocol	
	Fluorouracil		Oxaliplatin		SAI		FOLFOX Protocol	
	FOLFOX 6 Protocol		SAI		FOLFOX Protocol		FOLFOX 6 Protocol	
	Capecitabine		Fluorouracil		Fluorouracil		Fluorouracil	
	FOLFOX 6 Protocol		SAI		Oxaliplatin		SAI	
	FOLFOX Protocol		FOLFOX 6 Protocol		FOLFOX Protocol		FOLFOX 6 Protocol	
			FOLFOX Protocol		FOLFOX 6 Protocol		FOLFOX 6 Protocol	
					FOLFIRI Protocol		FOLFOX 6 Protocol	
					Capecitabine		SAI	
							Irinotecan	

Figure 4.6 CRC patients demand forecast of chemotherapy for the period of May-2013 to December-2013

Radiotherapy Demand								
Month	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13
Total Demand	9	16	11	13	6	15	15	10
Forecasted Chemotherapy Protocols	External pelvic	External pelvic	External pelvic	External pelvic	External pelvic	Endocavitary	to the Spine	Endocavitary
	External pelvic	External pelvic	Endocavitary	External pelvic	Endocavitary	Endocavitary	Endocavitary	External pelvic
	External pelvic	Endocavitary	External pelvic	External pelvic	Endocavitary	to the Spine	External pelvic	Endocavitary
	Endocavitary	Endocavitary	External pelvic	SAI	Endocavitary	SAI	External pelvic	External pelvic
	External pelvic	Endocavitary	Endocavitary	SAI	Endocavitary	External pelvic	SAI	External pelvic
	Endocavitary	SAI	External pelvic	External pelvic	Endocavitary	Endocavitary	SAI	Endocavitary
	Endocavitary	SAI	Endocavitary	External pelvic		Endocavitary	Endocavitary	Endocavitary
	Endocavitary	External pelvic	Endocavitary	Endocavitary		Endocavitary	SAI	External pelvic
	External pelvic	Endocavitary	Endocavitary	External pelvic		Endocavitary	External pelvic	SAI
		Endocavitary	Endocavitary	External pelvic		Endocavitary	to the Spine	External pelvic
		External pelvic	External pelvic	External pelvic		Endocavitary	Endocavitary	
		External pelvic		External pelvic		Endocavitary	External pelvic	
		External pelvic		External pelvic		Endocavitary	Endocavitary	
		External pelvic				External pelvic	External pelvic	
							Endocavitary	

Figure 4.7 CRC patients demand forecast of radiotherapy for the period of May-2013 to December-2013

4.6 Conclusion

With the increase in healthcare demand, it is crucial for hospitals to devise techniques to make better use of their resources by managing them efficiently by anticipating patient needs for these resources in the course of their care, especially for the case of chronic patients who require access to interdisciplinary resources (Harper and Shahani, 2002).

This paper proposes a hybrid machine learning-simulation demand prediction framework that is trained using previous patients' clinical pathways and predicts time-variant aggregated demand for hospitals' scarce resources for new patients. The framework is designed to be integrated with hospitals' resource planning and scheduling tools. Our machine learning method addresses the shortcomings of existing demand prediction methods that utilize crude patient classification schemes by incorporating personal and medical patient data into the prediction model. Not only does our framework improve the prediction accuracy for mid-term and long-term demand, but also, it is robust and practical.

To the best of our knowledge, our framework is the first to use artificial intelligence for demand prediction and resource planning in this context. However, it comes with limitations. Future work may improve this paper on multiple fronts. First, our model was applied to a limited dataset. We anticipate that the prediction results will improve as the model is trained with more data. Second, given the high computation power of our neural network, our framework can be improved if the training data is taken from Electronic Health Records, which provides medical information at a more granular level. Third, our framework has not been tested in a real-world hospital setting; future work may focus on measuring hospital-wide performance after incorporating this framework.

Machine learning and artificial intelligence technology has developed to the point where its potentials are endless. Our work has touched upon one application of machine learning in improving healthcare. We hope to witness its effectiveness in further applications to advance healthcare for the benefit of all.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Lawrence Rosenberg, the Executive Director of Jewish General Hospital, for his insightful comments and providing access to the hospital data collection. We would also like to sincerely thank the administration staff and surgeons at the Montreal Jewish General Hospital for their support. This work was supported by the "Natural Sciences and Engineering Research Council (NSERC) CREATE research" grant and the "Subvention pour projets de développement stratégique innovants 2012-2013 of the Fonds de recherche du Québec-Santé (FRQS)” grant.

CHAPTER 5 ARTICLE 2: AN INTEGRATIVE FRAMEWORK FOR SURGERY DURATION PREDICTION: A SUPERVISED LEARNING APPROACH

Elnaz Karimi[1], *Jean Marc Frayret*[1], *Michel Gendreau*[1], *Vedat Verter*[2]

Submitted to journal of *"Production and Operations Management"*

Abstract: This paper proposes an integrative predictive model for surgical procedures duration. The framework is the first of its kind to incorporate scheduling-related, operational and temporal attributes in addition to patient specific, procedure specific and surgeon specific attributes to predict surgical procedures duration. Furthermore, the framework illustrates the effectiveness of machine learning algorithms such as Random Forest and Support Vector Machine to capture the complex relationships among the predictors of surgical procedures duration. We applied the proposed framework to real hospital data and found an improvement of 31% in the accuracy of our predictive model compared to its practice benchmark. Furthermore, the results show that scheduling-related decisions such as procedure sequencing and block assignment have a significant impact on surgical procedures duration. This result has significant implications for operating room planning and scheduling literature at both tactical and operational levels. Namely, it indicates that optimal operating room planning is achieved only through joint optimization of surgical procedures duration and operating room scheduling and sequencing.

Keywords: Operating Room Scheduling; Operating Room Sequencing; Predictive Modeling; Machine Learning; Supervised Learning

[1]Department of Mathematics and Industrial Engineering, Polytechnique Montreal University, Montreal, Canada

[2]Desautels Faculty of Management, McGill University, Montreal, Canada

5.1 Introduction

Operating rooms are the costliest and largest profit-generating units in hospitals. Hence, it is of utmost importance for hospitals to minimize their operating room costs. One factor that hospitals consider in this respect is how to optimize the scheduling of the operating rooms. Operating room scheduling is concerned with where patients of various departments in a hospital must be placed in the rooms' daily schedule to not only provide timely service to patients with heterogeneous needs, but also operate the room at its maximum possible utilization. While an underutilized operating room will result in obvious revenue loss, an over-utilized operating room may lead to canceled surgeries or hospitals having to pay their medical staff for overtime work (Cardoen et al., 2010). In addition, a suboptimal operating room schedule may lead to congestion in the supporting units, such as the PACU, ICU, and wards (Gupta, 2007), or high waiting times for patients.

Operating room scheduling has received much attention in the field of operations management on the decision-making or objective-alignment fronts. On the decision-making front, many advanced surgical scheduling methods have been developed to address the problems arising from operating room scheduling in the hospital. Depending on whether patient waiting times are typically short or long, the scheduling decision may be dynamic or static, respectively. In the dynamic case, a scheduler such as the surgeon assistant assigns patients a surgery date at the time of their consultation according to hospital policies. In the static case, patients are put on a waiting list and scheduled simultaneously on, for example, a weekly basis. On the objective-alignment front, researchers have attempted to address the issues with conflicting objectives among the stakeholders: the hospital whose goal is to maximize its profits through maximum operating room utilization, its medical staff who strive to avoid working overtime, and patients who wish for minimum wait times (Gupta, 2007). Samudra et al. (2016) provide a comprehensive review of the state-of-the-art and the recent trends on operating room scheduling.

Predicting surgical durations is the key to optimizing operating room schedules. Nonetheless, accurate estimation of surgical durations is not a trivial task due to the inherent uncertainty in surgical settings. As of the date of this paper, the stochastic nature in the arrival of non-elective patients and surgery durations has been incorporated into the surgical scheduling models specifically using simulation techniques to evaluate the cost-benefits of interventions such as surgery cancellation. This stochasticity has been usually addressed by fitting a probabilistic distribution based on analyzing historical data (most notably a normal or log-normal probability distribution) with no attention to the underlying heterogeneity within the population (May et al., 2000). Subsequently, despite the efforts made in academic studies to incorporate variability in the surgical scheduling models, the practicality of their proposed methods is questionable because surgery durations are highly patient-specific and case-specific and do not possess a purely random nature (Eijkemans et al., 2010). Therefore, the conventional approach of fitting distribution functions to historical data is not sophisticated enough to predict the surgical durations accurately, rendering the methods based on distribution-fitting inapplicable to real hospital settings. Currently, hospitals employ a deterministic approach to estimate the duration of a particular surgery by simply using the historical average duration of that surgery, excluding any notion of uncertainty or case-specific prediction of the surgery duration. Subsequently, as stated in Cardoen et al. (2010), it is necessary to develop a practical solution which incorporates uncertainty to predict surgical durations.

Developing predictive models based on the drivers of variability can result in developing effective and practical surgical scheduling methods. Hence, in this paper our objective is to develop an integrative framework using advanced machine-learning algorithms that model surgical procedure durations. We do so by relying on patient, surgeon, scheduling and operational attributes to train our model to capture much of the variance in our data. Compared to the current practice of using linear regression, our machine-learning algorithm enhances the accuracy of the surgery prediction by an average of 31%, corresponding to a decrease of ≈30 minutes in the prediction error. The rest of the paper is organized as follows. Section5.2 provides a literature review which covers the state-of-the-art methods for predictive modeling of surgery durations. Section 5.3 explains the fundamentals of the applied supervised learning algorithms. Section 5.4 presents the framework for the integrative predictive model of surgery duration and illustrates its application on real-world data. Section 6.5 discusses the results of this application. Finally, section 6.6 provides our concluding remarks.

5.2 Literature Review

The stochastic approaches in the surgical scheduling literature that incorporate the uncertainty regarding surgery duration, do so by assuming a particular probabilistic distribution for either the duration of the surgery (i.e., process time), or the number of patients operated in each surgical block (i.e., flow rate); note that this number can serve as a proxy for surgical duration variability based on the assumption that if a scheduled surgery takes longer than expected the subsequent surgeries will be canceled and if it is shorter than expected more patients will be scheduled (Beliën and Demeulemeester, 2007; Guda et al., 2016). Most studies have used statistical modeling to capture the uncertainty of surgery duration. Since surgery durations are positive with a non-zero start time and are often heavily tailed, the general consensus in the field is that the log-normal distribution is the most suitable distribution for predicting surgery durations. Therefore, most related papers in the field of management science, have focused on estimating the optimal parameters of the log-normal distribution (May et al., 2000; Spangler et al., 2004). Although these papers provide an analytical and statistical guideline for incorporating surgery duration variability in scheduling methods and show its benefits in optimization, they do not specifically predict surgery durations. In other words, in these conventional approaches, a random number is assigned to every given surgery based on its assumed distribution as opposed to predicting its duration based on the characteristics of that surgery. Subsequently, we switch focus to papers that employ predictive modeling as opposed to distribution fitting.

The first step in developing predictive models is to identify the factors that explain the variability in surgical durations and sources of heterogeneity in surgical cases. Several papers in medical journals have attempted to identify these predictors. The most important and obvious predictor in surgery duration is the type of procedure and actions taken during the surgery. Dexter et al. (2008) found that the types of performed procedures are the most important factor when predicting general thoracic surgery durations. Other contributing factors included the surgical team and the type of anesthesia. Similarly, Shukla et al. (1990) found that predictive models based on the type of procedure, the surgeon and the level of case complexity generally performed better in estimating the surgical duration than case-specific estimates of surgeons. Other sources of variability considered in the literature are: patient attributes including age, gender, weight, health risk factor (Hsu et al., 2007; Strum et al., 2000), and operational and scheduling attributes such as time of day, whether a surgeon had the block to herself or followed another surgeon, surgical team composition, and type and duration of preceding cases (Kayis et al., 2012; Pandit and Carey, 2006; Wachtel and Dexter, 2009). We contribute to this stream of research by studying the impact of scheduling attributes such as type of block assignment (i.e., whether it is assigned to a specific specialty or shared between surgeons from different specialties, and whether the block is assigned to a single surgeon or multiple surgeons), block-mix (type of surgeries performed in the block) and sequencing (type and duration of the prior and subsequent surgeries) on surgery durations.

In recent years, several papers developed predictive models for surgery duration. Stepaniak et al. (2009) use a log-linear regression to predict surgery durations by considering surgeon-related predictors such as age, experience, gender, surgical team composition and time of the day. They find that including surgeon factors can improve the accuracy of the predictive models by 15% compared to the traditional estimation method of averaging the "last ten similar surgical cases". Kayis et al. (2012) use factors such as temporal attributes (i.e., year, month, week and time of the day) and operational attributes (i.e., number of surgeons, number of anesthesiologists, number of nurses, number of scheduled surgeries and the individual and joint experience of the surgery team) to predict the surgery duration. Using linear regression, they find that incorporating these factors in the regression model can highly reduce the prediction error. Zhou et al. (2016) develop a classification method to detect when an expert prediction (surgeon estimation) of pediatric surgical durations is an overestimation or underestimation of the actual surgical duration.

Generally, it can be inferred that developing a predictive model is a difficult task due to the high variability of surgery times as discussed above and the high number of predictors (most of which are interdependent categorical variables) that can impact surgery durations. Hence, as we show in the subsequent sections, a simple linear regression is not sophisticated enough to capture the relationships between the explanatory variables. We present a sophisticated predictive model for surgery duration based on supervised learning techniques.

5.3 Methods

This section describes the fundamentals of machine learning methods used for predicting surgery procedure durations. Machine learning is a form of applied statistics that exploits the advances in computational statistics to derive meaningful and complex patterns among data. Machine learning algorithms are able to learn from examples and are classified into three categories: supervised, unsupervised and reinforcement learning. In a supervised learning setting, the algorithm uses a set of known inputs and outputs called the "training set" to find patterns between the data and use them to map unseen inputs to their respective outputs. The outputs may be categorical (i.e. classification problems) or continuous (i.e., regression problems). In an unsupervised learning setting, the algorithm's task is to discover the similarities between inputs and cluster them based on similar traits. This technique is useful in image classification applications or, generally, grouping/sorting data with many features. Finally, in a reinforcement learning setting, the learning algorithm must make decisions based on its training inputs to reach a certain output to maximize its "reward" (Goodfellow et al., 2016). A recent application of this was used by Silver et al. (2017) to develop the ground-breaking chess engine, Alphazero. An in-depth description of each technique and their pros and cons are provided in (Bishop, 2006).

The tremendous advances in machine learning techniques in the recent years has led many scholars such as Kleinberg et al. (2015) and Athey (2017) to believe that the prediction problems which are otherwise unpredictable or inaccurate using traditional regression approaches, are solvable using learning algorithms, which, in turn, leads to better data-driven decision-making and policy making. Recent papers in the field of healthcare operations management have used machine learning techniques to predict emergency department wait time (Ang et al., 2015b), reduce hospital readmissions (Helm et al., 2016b), classify hospital inpatient admission (Krämer et al., 2017), and solve nurse staffing problems (Ban and Rudin, 2014).

In this work, our goal is to predict surgery durations based on historical data. A supervised learning algorithm is the most suitable for this application because it is convenient to use past examples to train our algorithm and use it to predict future surgery durations. Also, since the target output is continuous, we are interested in supervised learning algorithms that predict a target numerical value (y) given the input (x) (i.e., "regression"). All supervised learning algorithms consist of an optimization algorithm, a cost (loss) function, a hypothesis model, a training dataset with m observations to build the model, and a test dataset to implement the model. In this section, we provide an overview of the commonly used supervised learning algorithms.

5.3.1 Regularized Multivariate Linear Regression

A multivariate linear regression is one of the simplest and most basic supervised learning techniques (Bishop, 2006). The algorithm takes a vector of inputs ($x \in R^n$), and uses a linear combination of them to predict the value of a scalar ($y \in R$). Equation 5.1 describes the linear hypothesis, where each θ_j corresponds to the weight of its corresponding x_j and, specifically, θ_0 is called a *bias* and corresponds to a dummy variable $x_0 = 1$. The bias terminology is derived from the fact that the output (target value) is biased toward being θ_0 in the absence of any input.

$$h_\theta(x) = \sum_{j=0}^{N} \theta_j x_j \tag{5.1}$$

We use the sum-of-squares error (SSE) function as the cost function. Intuitively, SSE is the Euclidean distance between the predicted target values and the actual target values and decreases to 0 when $h_\theta\left(x^{train}\right) = y^{train}$. Therefore, the objective of the minimization is to find the vector θ such that SSE is minimized. We further add the regularization term $\lambda \sum_{j=1}^{n} \theta_j^2$ to the cost function to avoid overfitting by penalizing large θs, where λ is called the *regularization parameter* and indicates how important the regularization term is with respect to the SSE. The m parameter indicates the size of the training set. Finally, the SSE minimization problem is called "regularized least square cost".

$$\min_{\theta_0,\theta_1,\ldots,\theta_n} J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right] \tag{5.2}$$

Since the learning algorithm requires a high amount of numerical computation, we use the gradient descent algorithm to solve the optimization problem as opposed to using the Normal equation. Gradient descent is an iterative process that starts with a guess for θ (usually 0), uses the vector of partial derivatives of $J(\theta)$ with respect to θ and continues until it reaches the minimum. Algorithm 1 describes the gradient descent algorithm, where α is the learning step that determines the size of the step of each iteration and is often set to a small constant.

Algorithm 1 Gradient Descent Algorithm for Regularized Multivariate Linear Regression

1: **procedure** GRADIENT DECENT
2: $\quad Set = 0$
3: $\quad$ **repeat** $\theta_j := \theta_j \left(1 - \alpha\frac{\lambda}{m}\right) - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_\theta\left(x^{(i)}\right)\right)x_j^{(i)}$
4: $\quad$ **until** *Convergence*

In this algorithm, the α and λ parameters need to be pre-determined.

5.3.2 Random Forest Regression

The Random Forest algorithm belongs to the class of "*ensemble learning*" methods, where an aggregation of multiple individual learning algorithms are used to solve a problem (Liaw et al., 2002). The Random Forest algorithm fits numerous regression trees (described next) to a dataset, and aggregates the predictions from all trees in the ensemble to find the optimal prediction.

Regression trees partition the input space to a set of regions; suppose $\tau = 1, \ldots |T|$ are the indexes for regions R_τ in the input space, where, each R_τ includes N_τ observations. The regression tree then implements a piecewise linear regression over the entire input space (all regions). The predicted value of the target output $((h_\tau(x)))$ within each given region is obtained by averaging the values of the actual target values (y) for the data points that fall within that region (equation 5.3).

$$h_\tau = \frac{1}{N_\tau}\sum_{x_m \in R_\tau} y_m \tag{5.3}$$

Similar to the regularized multivariate linear regression, a regression tree's cost function comprises of the sum-of-squares error function and a regularization term. However, whereas in the former case the regularization term was to avoid large θ values, in a regression tree, the regularization term is to contain the number of nodes in the tree to avoid over-fitting (equation 5.4). A joint greedy optimization method is used to find the optimal structure of the tree. In this method, decision variables include the choice of input features for each split node and their corresponding thresholds (Bishop, 2006).

$$J_\tau(T) = \sum_{\tau=1}^{|T|} \sum_{x_m \in R_\tau} (h_\tau - y_m)^2 + \lambda |T| \tag{5.4}$$

The Random Forest method, proposed by Breiman (2001), is one of the most well-known methods of ensemble learning and constructs more powerful prediction models compared to regression trees. While both methods aim to find the optimal node splits which lead to the least SSE, their approaches are different. Whereas in a standard regression tree, the optimal node split is chosen among the full set of input features, the Random Forest method does so by considering only a *random selection* of input features. The Random Forest algorithm is as follows: the algorithm grows a number of regression trees (n_{tree}) on *bootstrapped* observations from the training dataset. At each node the algorithm randomly selects m_{try} number of the features ($m_{try} \leq M$, where M is the total number of features in the dataset) input features while growing each bootstrapped regression tree; m_{try} is often equal to the square root of the total number of the input features. At each bootstrap iteration the error rate of the grown regression tree is obtained by predicting the target values for observations outside of the bootstrap sample; these observations are called "*out-of-bag*" or OOB data. On average, each bootstrapped tree is grown using around two-thirds of the total number of the observations. Therefore, each observation is an OOB around 33% of the times. For each observation i, the final prediction is obtained by averaging all OOB predictions for that observation and the Random Forest algorithm error is similarly estimated using the final prediction of all OOB observations. For more elaboration on this method, refer to Breiman (2001).

For this algorithm, a number of parameters must be predetermined: n_{tree} (total number of trees to be grown), m_{try} (the number of features to be sampled at each split), and node size which sets the minimum number of observations in each node; this, in turn, determines when the tree will stop to grow since nodes with fewer than node size observations will not be further split.

5.3.3 ϵ-Support Vector Machine

The support vector machine (SVM) technique is one of the most influential approaches in supervised learning and is extensively applied to both classification and regression problems in various settings (Cortes and Vapnik, 1995). SVM for regressions or support vector regression (SVR) has shown to be highly generalizable when applied to new data. Furthermore, the technique is consistent when applied to new datasets as it relies solely on a limited number of learning patterns known as "support vectors" (described later in the section). SVM maps data into a high-dimensional feature space using a nonlinear function called the *kernel function* and constructs an optimal separating hyperplane. Kernel functions are the inner product of the images of two data points (x_i, x_j) in the feature space (H): i.e., $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$, where $\Phi : X \to H$ (refer to Drucker et al. (1997)). Kernel functions allow the learning algorithm to explore high dimension spaces looking for meaningful patterns without significant computational cost. The SVR then fits a linear regression function in the newly created high-dimensional feature space. As the most basic example, assuming that the kernel function were linear and equal to 1 (i.e., $\Phi(x_i) = x_i$) and the mapping were performed only once, this would result in the inclusion of "interaction terms" in a conventional regression, where, each two input features are multiplied and included in the regression.

The SVR hypothesis model is given in equation 5.5. Similar to the multivariate linear regression hypothesis, the SVR hypothesis includes a scalar bias term, and is linear in the model's weight vector w^T.

$$f(x) = w^T \Phi(x) + b \tag{5.5}$$

Unlike the simple linear regression quadratic cost function discussed in section 5.3.1, the SVR's cost function is *insensitive* to errors less than ϵ and thus creates a *tube* around the true target output values, which is called the ϵ-*tube*. The model is called ϵ-SVR and the corresponding ϵ-insensitive loss function is defined by:

$$L_\epsilon(f(x) - y) = \begin{cases} 0 & \text{if } |f(x) - y| \prec \epsilon; \\ |f(x) - y| - \epsilon, & \text{otherwise.} \end{cases} \tag{5.6}$$

The ϵ-SVR primal objective function also includes an additional regularization term to penalize large weights and reduce complexity $\left(\frac{1}{2}\|w\|^2\right)$. Ultimately, the ϵ-SVR error function can be written as:

$$E(w) = C \sum_{m=1}^{M} L_\epsilon(f(x) - y) + \frac{1}{2}\|w\|^2 \tag{5.7}$$

where C is the regularization parameter. By introducing slack variables $\xi_m \geq 0$ and $\widehat{\xi}_m \geq 0$ for each observation x_m, where $\xi_m > 0$ corresponds to the observation for which $f(x) > y + \epsilon$, and $\widehat{\xi_m} > 0$ corresponds to the observation for which $f(x) < y - \epsilon$, we can rewrite the primal error function by:

$$\begin{aligned}
\underset{w}{\text{minimize}} \quad & E(w) = C \sum_{m=1}^{M} \left(\xi_m + \widehat{\xi}_m\right) + \frac{1}{2}\left\|w\right\|^2 \\
\text{subject to} \quad & f(x) \leq y + \epsilon + \xi_m, \\
& f(x) \geq y - \epsilon - \widehat{\xi}_m, \\
& \xi_m \geq 0, \\
& \widehat{\xi}_m \geq 0, \ \ (i = 1, \ldots, m).
\end{aligned}$$

This optimization problem can be solved using Lagrangian multipliers and writing the dual maximization problem (see Drucker et al. (1997) for more information). By introducing the kernel function as $k(x, x^{'}) = \phi(x)^T \phi(x^{'})$, the hypothesis function can be derived as:

$$f(x) = \sum_{m=1}^{M} \left(\alpha_m - \widehat{\alpha}_m\right) k(x, x^{'}) + b \tag{5.8}$$

where α_m and $\widehat{\alpha}_m$ are the Lagrange multipliers corresponding to the optimization problem's first and second constraints, respectively, and are the solutions to the following quadratic problem:

$$\begin{aligned}
\underset{\alpha_m, \widehat{\alpha}_m}{\text{maximize}} \quad & -\frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} \left(\alpha_m - \widehat{\alpha}_m\right) \left(\alpha_n - \widehat{\alpha}_n\right) k(x_m, x_n) \\
& - \epsilon \sum_{m=1}^{M} \left(\alpha_m + \widehat{\alpha}_m\right) + \sum_{m=1}^{M} \left(\alpha_m - \widehat{\alpha}_m\right) y_m \\
\text{subject to} \quad & \sum_{m=1}^{M} \left(\alpha_m - \widehat{\alpha}_m\right) = 0 \\
& \alpha_m, \widehat{\alpha}_m \in [0, C]
\end{aligned}$$

A nonzero value of α_m or $\widehat{\alpha}_m$ for each data point implies that it lies outside of the ϵ-tube. Therefore, only the nonzero values of α_m and $\widehat{\alpha}_m$ contribute to the prediction of the target value and hence, are known as the *support vectors* and all data points that lie inside the ϵ-tube (i.e. both α_m and $\widehat{\alpha}$ are zero) do not contribute to the regression function and therefore are excluded from the model.

5.3.4 Parameter Selection

As previously mentioned, the kernel function $k(x, x^{'})$ enables the dot product to be performed in a high-dimensional feature space. Kernels that are commonly used in SVRs are shown in table 2. In our model, we experiment with both polynominal and ANOVA RBF due to their superior performance in regression problems (Karatzoglou et al., 2005). Regarding the ANOVA RBF, two key parameters σ (also known as the *radius of influence* of each support vector) and d (degree) must be preselected. Regarding the polynomial function, three parameters: d (degree), λ (scale) and c (offset) must be preselected. There are two more parameters that must be selected prior to training the SVR learning algorithm: C and ϵ; as previously mentioned, parameter C controls the overfitting of the model, while ϵ determines the training tolerance (boundary of the tube). Following the SVM literature (Karatzoglou et al., 2005), we use multiple combinations of all parameters for both kernel functions and score each combination based on the model's performance in the k-fold cross validation test. The k-fold cross validation test is an iterative algorithm where the number of iterations is predetermined. First, it divides the training set into k equal-sized subsets of data randomly. In each iteration, it holds one subset as the validation set and uses the other $k-1$ subsets to train the SVM model. The SVR learning algorithms were carried using the *kernlab* package in R (Karatzoglou et al., 2004).

Table 5.1 Common Kernel Functions

Name	Function
Linear	$\left\langle x, x^{'} \right\rangle$
Gaussian Radial Basis Function (RBF)	$\exp\left(-\sigma \left\Vert x - x^{'} \right\Vert^{2}\right)$
Polynomial	$\left(\lambda \left\langle x, x^{'} \right\rangle + c\right)^{d}$
ANOVA Radial Basis function (RBF)	$\left(\sum_{k=1}^{m} \exp\left(-\sigma \left(x^{k} - x^{'k}\right)^{2}\right)\right)^{d}$

5.3.5 Performance Evaluation

Several measures can be used to evaluate the success of numerical prediction models, that is how close the model prediction of the target value of observation i ($\tilde{y}$) is to the actual target value for that observation (y) in the test dataset. The choice of the measure is often determined by the type of error that is of importance. If the *absolute* error values are the indicators of the performance of the model, then the most commonly used measure is the mean-squared error (MSE). Note that sometimes root mean-squared error (RMSE) is used so that the error measures have the same dimensions as the predicted value. However, MSE or RMSE measures are biased with regards to the effect of outliers in the model; they penalize large discrepancies much more heavily than small ones. An alternative measure is the mean absolute error (MAE) that averages the magnitude of individual errors, mitigating the effect of the aforementioned bias. On the other hand, if the *relative* error values are of importance then the relative squared error (RSE) is measured. RSE compares the prediction error of the model with the prediction error of a simple predictor such as simply "averaging the observations in the training set to predict the test set (denoted as $\overline{y}$)". Similar to MSE, RSE is biased toward heavily penalizing large discrepancies and similar to MAE, the relative absolute error (RAE) measure is recommended to deal with the bias. (See equations 5.9−5.12)

Absolute Error Measures	Relative Error Measures
$$RMSE = \sqrt{MSE} = \sqrt{\sum_{i=1}^{n} \frac{(\tilde{y}_i - y_i)^2}{n}} \quad (5.9)$$	$$RRSE = \sqrt{RSE} = \sqrt{\frac{\sum_{i=1}^{n} (\tilde{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \quad (5.10)$$
$$MAE = \sum_{i=1}^{n} \frac{\lvert \tilde{y}_i - y_i \rvert}{n} \quad (5.11)$$	$$RAE = \frac{\sum_{i=1}^{n} \lvert \tilde{y}_i - y_i \rvert}{\sum_{i=1}^{n} \lvert y_i - \overline{y} \rvert} \quad (5.12)$$

To evaluate and compare the performance of our surgery duration prediction models we use all four aforementioned measurements (see section 6.5). However, since the objective of the models is to improve the current surgery duration predictor (average of last 10 similar surgeries), we will mostly rely on RRSE and RAE.

5.4 Model

In this section, we describe the proposed integrative framework for surgery duration prediction using supervised learning models. We begin by describing the model setting in section 6.4, followed by a description of the data used to train the model in section 5.4.2. Figure 5.1 illustrates the entire model development process. As shown, the framework begins with a data preprocessing stage, followed by a data split step where the data is divided into a training set and a test set. The training set is used in the subsequent stage to train the model using the three methods described in section 5.3. The three trained models, each using one of three methods, are then applied to the test set in the performance evaluation stage. The trained model with the best performance is chosen as the optimal prediction model. In the following subsections, we first explain the data preprocessing and data split procedures in section 5.4.3 and 5.4.4, respectively, followed by the training procedure in section 5.4.5.

5.4.1 Model Setting

The Montreal Jewish General Hospital (JGH) is one of Canada's largest and busiest acute-care teaching hospitals, with 637 beds and more than 23,000 patient admissions per year. The

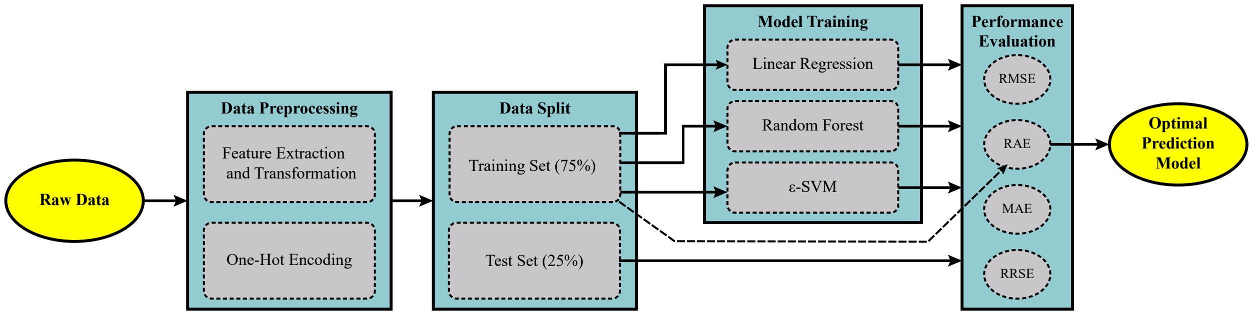


Figure 5.1 General framework of the prediction model

surgical pavilion in JGH consists of 13 operating rooms, a 40-bed ward and ICU; and performs operation procedures from 14 specialties: breast oncology, plastics, otolaryngology (E.N.T.), ophthalmology, cardiac, thoracic, gynecology, orthopedics, colorectal, urology, neurosurgery, general surgery, dental, and vascular. At JGH, surgeries are scheduled using a state-of-the-art operating room information system software (OPERA) that provides decision-makers with supportive information regarding the allocation of operating room blocks to the surgical specialties and individual surgeons. The criteria for block scheduling at JGH include patients' wait time, hospital costs, expected operating room overtime, expected operating room idle time, etc. OPERA keeps records of all surgeries performed at JGH and uses the average of the last 10 similar procedures to predict the surgery duration and required OR time for a specific procedure. Individual surgeons then develop their block schedule that includes the surgical block-mix (set of surgical procedures), and sequence their patients according to the predicted surgery durations.

5.4.2 Data

OPERA collects administrative data on each surgery performed at JGH. The data includes patient attributes (patient ID, gender, wait-time, type of visit, type of anesthesia, primary diagnosis and secondary diagnosis), procedure attributes (specialty, primary surgeon, date of surgery, time stamps of patient entry to the operating room, start of the procedure, end of the procedure, patient departure from the operating room, etc.) and operating room schedule attributes (operating room number). We use administrative data from OPERA for the years 2013-2014 to train our model. It must be noted that real-time data from OPERA can always be used to improve the accuracy of the model. We exclude all non-elective (emergency) patients from the data set for the following reasons: first, our model is designed to provide support for operating room scheduling decisions and therefore, cannot be used on emergency patients who need surgery immediately. Second, due to the immediate nature of the operations, non-elective procedures may introduce unwanted noise to our model. With these exclusions, our full dataset comprises 14,103 observations.

The time a patient spends in the operating room can be divided into three segments: 1) this pre-operation represents the time from when the patient enters the operating room to when the surgery begins and includes preparing procedures such as performing anesthesia on the patient; 2) this operation represents the surgical procedure length; 3) this post-operation represents the time from when the surgery ends to when the patient is removed from the operating room (this segment is insignificant since it usually takes less than a few minutes). Compared to the duration of the pre-operation stage, the duration of the operation stage is not only longer on average, but is much more variable. For example, our analysis on the 18 most common surgeries across all specialties shows that the average standard deviation across all 18 surgeries for the pre-operation duration is 8.4 minutes compared to 35.2 minutes for that of the operation stage. Therefore, we believe that using the average predictor performs well enough for estimating the pre-operation duration and there is little to be gained by applying machine learning algorithms to predict their duration. Therefore, our model focuses on predicting the duration of the operation stage. In our model we denote output y as the operation (surgery) duration. Table 5.2 provides a summary of the surgery durations and pre-operation durations of the 10 most commonly performed surgeries at JGH.

Table 5.2 summary of operation and pre-operation durations

	Surgery Duration		Pre-operation Duration	
Intervention	Mean	Standard Deviation	Mean	Standard Deviation
Extraction cataract with IOL (intraocular lens)	17	8	9	4
Arthroplasty hip total	73	26	37	8
Vitrectomy pars plana 25 gauge	46	25	21	8
Arthroplasty knee total	80	30	40	14
Mastectomy segmental and excision lymph node sentinal	49	18	24	7
Cholecystectomy	51	31	20	6
Mastectomy segmental	32	13	19	6
TURBT (Transurethral resection bladder tumor)	33	18	23	7
Thyroidectomy total with unilateral central neck dissection	75	32	18	5
Extraction cataract (previous vitrectomy) with IOL	16	7	15	8

5.4.3 Data Preprocessing

Feature Extraction and Transformation

Feature extraction refers to actions which involve manual searching within the data to discover implicit patterns/features which may impact the outcome of the model prediction. Once discovered, feature transformation refers to the actions taken to transform the patterns/features into meaningful numbers or categories. In this work, we performed feature extraction and transformation on scheduling, sequencing and block-mix, and pooling, as described next.

Scheduling: At JGH each operating room day is divided into a morning block and an afternoon block. Operating room blocks are assigned to each surgical specialty unit one month in advance. In some cases, one unit holds the operating room for both blocks of the same day (i.e., for the entire duration of the day) which we define as a "specialty-day". Conversely, in other cases the morning block and afternoon block are assigned to two different units which we define as a "specialty-block". Each surgical specialty unit will then allocate its block times among its surgeons. Furthermore, each surgeon can share her block with another surgeon (defined as "shared-surgeons"), or hold the entire block to herself (defined as "one-surgeon"). Therefore, scheduling can take one of the four following forms: 1) Specialty-Day One-Surgeon (SDOS), 2) Specialty-Day Shared-Surgeons (SDSS), 3) Specialty-Block One-Surgeon (SBOS), 4) Specialty-Block Shared-Surgeons (SBSS). We hypothesize that block scheduling may affect the surgery durations. For example, if a block is shared between surgeons A and B, surgeon A may speed up her operations to avoid spillovers to surgeon B's time within the block. Although OPERA does not specifically track the type of block scheduling, such information can easily be extracted from the existing data. For each block in the data, we subset unique values for specialty and surgeons to obtain the corresponding scheduling. Table 5.3 provides a summary of the type of scheduled blocks at JGH.

Sequencing and Block-Mix: Once surgeons realize their block assignments, they plan their block schedule, which includes determining the surgical block-mix (set of unique surgical procedures performed in the block) and the sequence of patients. We hypothesize that both block-mix and the sequence of surgeries may affect the duration of surgery. To investigate this, we create two additional features for each observation, the first to store the surgery prior to the observation, and the second to store the surgery after the observation; we denote these as "prior operation" and "subsequent operation", respectively, to capture the sequences and mixes of surgeries.

Surgeon Pooling: In many surgical units, surgeons may be assigned to operate in multiple

Table 5.3 Block Scheduling Type in JGH

Specialty	SDOS	SDSS	SBOS	SBSS	Grand Total
Breast Oncology	204	205	128	6	543
Cardiac	322	54	4	1	381
Colo-rectal	504	27	39	2	572
Dental	66	24	1	0	91
E.N.T.	1247	87	43	4	1381
General surgery	1224	84	204	11	1523
Gynecology	910	35	42	2	989
Neurosurgery	326	10	5	1	342
Ophthalmology	4523	133	0	0	4656
Orthopedics	1667	43	16	0	1726
Plastics	196	4	39	2	241
Thoracic	124	0	4	1	129
Urology	971	40	11	1	1023
Vascular	465	26	13	2	506
Grand Total	12749	772	549	33	14103

operating room blocks simultaneously; for example, surgeon A may be assigned to operating rooms 1 and 2 on a Monday morning block. This is called *operating room pooling*. Operating room pooling allows a surgeon to carry out multiple surgeries in parallel and is usually employed for short operations. The idea is that the main surgeon needs to be present only to perform the *critical* part of the surgery and *trivial* procedures such as stitching or initial cutting can be performed by training individuals such as residents. Therefore, the main surgeon can alternate between patients while their operations are performed simultaneously, or move from from one patient to another before the first patient's operation is complete.

Table 5.4 provides a summary of the inputs to the model.

Table 5.4 Summary of inputs

Input Feature	Type	Levels	Abbreviation
Surgery	Categorical	40	surgery
Sex	Binary	2	sex
Diagnosis	Categorical	3-12	diagnosis
Visit type	Categorical	3	visit
Pre-Operation Duration	Numeric	NA	preops
Surgeon	Categorical	2-5	surgeon
Day of Week	Categorical	5	day
Hour of Day	Categorical	8	hour
Room	Categorical	3-7	room
Block type	Categorical	4	block
Prior operation	Categorical	6-12	prior
Subsequent operation	Categorical	6-12	Subsequent
Number of remaining scheduled patients	Categorical	0-9	remain
Surgeon pooling	Binary	2	pool

One-Hot Encoding

Many of the input features in our model are categorical (type of visit, type of anesthesia, diagnosis, surgeon, block-type, block-mix, pre-operation and post-operation) with multiple label levels, and, in most cases, these labels do not have a natural ordering relationship with respect to each other. For example, patients that undergo the "cholecystectomy" surgery may have had one of the following 6 diagnoses prior to the operation: "cholelithiasis", "cholecystitis", "mass gallbladder", "colic biliary", "cholecystitis acute" or "pancreatitis", which by no means are ordinal (or interrelated). However, most machine learning algorithms, including Random Forest regression and SVR, cannot operate on categorical data directly and require all input and output variables to be numerical. Categorical features can be encoded into numeric features using the so-called "one-hot" method. This method encodes a categorical feature with n distinct labels by generating n dummy features which are 0 by default. Next, for each observation, depending on what label its categorical feature takes, the encoder replaces a 1 in the dummy feature corresponding to that label level. For instance, in the previous diagnosis example, as shown in table 5.5, the one-hot encoder transforms the diagnosis feature to 6 new input features. The "*d*" stands for the primary feature "*diagnosis*" and assuming the diagnosis on observation 1 is "colic biliary" and "cholecystitis acute" for observation 2, the newly created features "d-cholecystitis acute" and "d-colic biliary" are 1 for the first and second observations, respectively.

Table 5.5 One-hot encoder example

Observations	d-Cholelithiasis	d-Cholecystitis	d-Mass gallbladder	d-Colic biliary	d-Cholecystitis acute	d-Pancreatitis
Observation 1	0	0	0	1	0	0
Observation 2	0	0	0	0	1	0

5.4.4 Data Split

To evaluate and compare the prediction accuracy of our models, we divide the data into two parts: training set and test set. The training set is used to train the model and the test set is used to evaluate it. Typically, the more a model is trained, the better will be its prediction accuracy. In this work, due to the limited size of our total dataset and following the literature on supervised learning, we select 75% of the data as the training set and leave a "sufficient" 25% to test the performance of the model. In our work, the decision on which set to allocate each observation is on a purely random basis.

5.4.5 Model Training and Cross-Validation

As discussed in section 5.3, the three discussed methods, multi-variate linear regression, Random Forest, and ϵ-SVR, require a few key parameters to be pre-selected. In summary, the multivariate linear regression requires the α (learning rate) and λ (regularization parameter) parameters to be predetermined; for the Random Forest algorithm, n_{tree} (number of the trees), m_{try} (number of features to be sampled at each split), and the minimum node size are predetermined; meanwhile, the ϵ-SVR algorithm requires parameters such as: Kernel-specific hyper-parameters, the C constant of the regularization term in the Lagrange formulation, and the ϵ (boundary of error insensitivity) parameter in the insensitive-loss function.

We use the cross-validation technique for the optimal selection of all parameters with one exception; we rely on the built-in heuristic algorithm developed by Karatzoglou et al. (2004) to optimally select the kernel-specific hyper-parameters. For each potential value of each parameter, we evaluate its respective model's performance, using the 10-fold cross-validation with 4 iterations.

5.5 Results

We implement the three machine learning models described above to predict surgery durations. We then compare the performance of these models with the current practice at JGH (average of the last 10 similar surgeries) to evaluate the value of using machine learning techniques to predict surgery durations. Table 5.6 summarizes the average RMSE, MAE, RRSE, and RAE of each of the machine learning models across 40 different types of surgical procedures. Among the studied supervised learning models, while all models are superior in predicting surgery durations compared to the current practice at JGH, on average, the Random Forest model outperforms ϵ-SVR and multivariate linear regression in terms of its forecasting accuracy. Note that the tabulated error values are averages across all 40 surgical interventions studied. For 19% of the surgeries (7 out of 40), the improvements in accuracy reached was in the range of $40\% - 45\%$, which is significantly higher than that of reported predicted models to date. Figure 5.2 specifically compares the measure of success: $(1 - RAE) \times 100\%$ for the Random Forest and ϵ-SVR models across 20 different type of surgical procedures. A closer look at the individual results reveals that in 7 out of the 20 surgical procedures shown the ϵ-SVR model performed better than the Random Forest model and in 2 cases they had similar performance. We further discuss the process of picking the optimal algorithm in the next section.

Table 5.6 Comparison of leaning algorithms performances

Intervention	RMSE	MAE	RRAE	RAE
Current Method	44.67	37.18	1	1
Regularized Multivariate Linear Regression	35.56	25.83	0.89	0.84
Random Forest	15.41	9.96	0.74	0.69
ϵ- Support Vector Regression	16.62	10.12	0.73	0.71

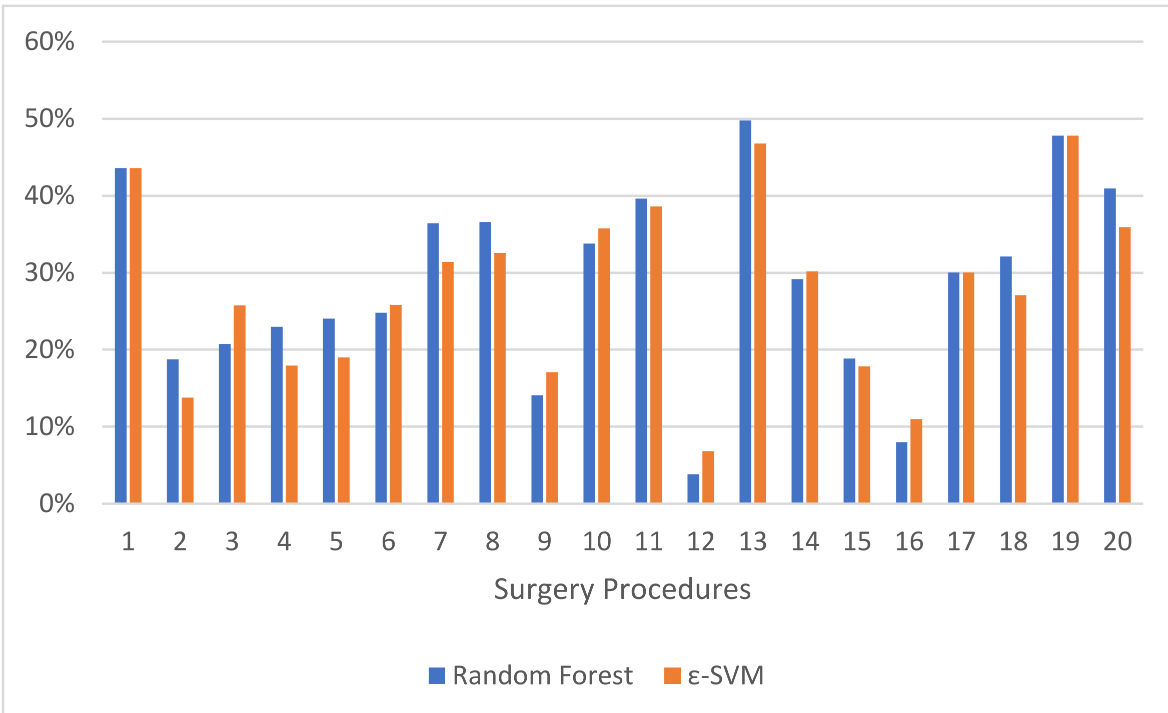


Figure 5.2 Comparison of the measure of success of the Random Forest and ϵ-SVR methods across 20 surgery procedures

5.5.1 Importance of Model Features

The Random Forest algorithm allows us to investigate the importance/significance of the model features in the regression. To this end, the Random Forest model uses a metric called "*mean decrease in accuracy*" to rank the importance of the features in developing the model. The idea is to understand which features contribute the most to the development of the model (in our case, the model for prediction of surgery durations). Intuitively, the mean decrease in accuracy evaluates how much accuracy is lost if each feature is permuted. Mathematically, it is computed through calculating the increase in error caused by permutation of features in the OOB data. For each feature, specifically, the procedure for calculating the mean decrease in accuracy is as follows: 1) For each tree, the MSE of its OOB is calculated; 2) the algorithm permutes the feature from all trees; 3) For each newly created tree, the MSE of its OOB is recalculated; 4) the mean decrease in accuracy is calculated by subtracting the original MSE in step 1 from the MSE in step 3, averaged over all trees.

For the Random Forest regression algorithms developed for each of the 40 surgical procedures, table 5.7 tabulates the ranking of the features in terms of their importance (top six) and the frequency at which they are placed at each rank. The results show that the patient's diagnosis prior to surgery and the surgeon are the two main predictors of surgery duration (ranked either first or second most of the time), which is somewhat intuitive and previously shown (Stepaniak et al., 2009). However, these two predictors, alone, cannot explain the high variability of surgery durations. Other features, such as surgeon pooling, number of remaining surgeries, pre-operation duration, type of block assignment, prior surgery, room and hour of day, are all indicated as important predictors. Furthermore, sequencing of the surgeries ("prior" predictor), number of remaining scheduled patients, and surgeon pooling which we extracted, as previously explained in section 5.4.3, are indicated as significant features in predicting surgery durations that have not been considered in previous predictive models. Another interesting observation is that the set of significant predictors are quite diverse across different types of surgeries. For example, while "surgeon" is the most important feature for some types of surgical procedures, it does not play any role in predicting other type of surgeries. This observation further highlights the necessity of using supervised algorithms with the ability of *feature selection* for training the surgery procedure duration prediction model.

5.5.2 Optimal Learning Model

The objective of any learning algorithm is to minimize the expected error of its predictions. However, there exist a bias-variance trade-off inherent in any learning algorithm (Bishop, 2006). For example, if we assume that the surgery duration is a function of model features and the random variable (ϵ) that is normally distributed with mean 0 and variance σ_ϵ ($y = h(x) + \epsilon$) and $\widehat{h}(x)$ is our estimated model, then the expected squared prediction error at a point x is:

$$E\left[\left(y-\widehat{h}(x)\right)^2\right] = \left(E[\widehat{h}(x)] - h(x)\right)^2 + E\left[\left(\widehat{h}(x) - E[\widehat{h}(x)]\right)^2\right] + \sigma_\epsilon^2 \tag{5.13}$$

Table 5.7 Frequency and ranks of features across 40 surgery procedures

	Rank					
Feature name	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
sex	0	0	0	0	2	0
diagnosis	20	5	3	1	0	0
visit	0	0	0	3	0	0
preops	0	7	8	12	6	7
surgeon	16	15	4	1	0	0
day	0	0	2	0	0	1
hour	0	1	0	6	0	8
room	1	0	5	0	5	3
block	0	0	3	2	5	4
prior	0	0	6	6	11	8
Subsequent	0	0	0	0	1	2
remain	1	2	4	3	6	2
pool	2	3	3	4	5	5

The error due to bias (first term on the left side of equation 5.13) is the difference between the expected prediction of the learning algorithm and the real target value; for example, if $h(x)$ is non-linear and we choose a linear model for $\widehat{h}(x)$, our model will suffer from high bias. The error due to variance (second term on the left side of equation 5.13) is the extent to which the predictions of x vary between different training datasets (i.e., how much variability exists across models trained with different datasets). As discussed, the bias-variance trade-off implies that a lower variance results in higher bias and vice versa. Underfitting often results in *high* errors in both training and testing, and is an indication of a large *bias* in the model. On the other hand, overfitting often results in a *low* training error and a *high* testing error and is an indication of a large *variance* in the model. A good fit is often accompanied by low training and testing errors with the testing error being slightly greater than the training error. Therefore, following the literature (Bishop, 2006; Friedman et al., 2001), we will use both training and testing errors to select the optimal model for each surgery type. Table 5.8 presents a comparison of the training and testing errors among the supervised learning algorithms, averaged across the 40 surgery types. On average, the Random Forest model has the best performance on both on the training and test datasets with an improvement of 31% $(1 - 0.69) \times 100\%$, while the ϵ-SVR model closely follows. Following the indications described above, the regularized multivariate linear regression contains the highest testing errors and, therefore, suffers from a serious overfitting problem. The main reason for the poor performance of this model is that the the surgery duration problem has a high-dimensional input vector and, generally speaking, linear regressions do not perform well when faced with high-dimensional input vectors. Indeed, the results of table 5.8 justify our selection of supervised learning techniques, such as Random Forest and ϵ-SVR, that can effectively handle high-dimensional data through *feature selection*.

Table 5.8 Comparison of the performance of supervised leaning algorithms.

Intervention	Training RAE(%)	Testing RAE(%)
Regularized Multivariate Linear Regression	72.24	84.11
Random Forest	68.21	69.16
ϵ- SVR	67.14	71.39

5.6 Conclusion

In this paper, we developed an integrative framework for surgical duration prediction. Improving the accuracy of prediction of surgical durations can lead to substantial benefits for hospitals by enabling more efficient surgical schedules. Due to their deterministic approach (using the average surgery durations) or incorporating uncertainty disregarding the case-specific features of the surgery, conventional prediction methods of surgical durations lead to inaccurate predictions of surgery durations, which, in turn, leads to the under-utilization or over-utilization of operating rooms. Consequently, underestimated surgery duration predictions lead to cancellation of surgeries due to insufficient remaining time for the remainder of the surgeries scheduled for the day, and overestimated surgery durations lead to congestion in down-stream resources such as PACU and ICU, since they cannot provide beds for the incoming operating room patients in a timely manner.

Additionally, the proposed framework contributes to the surgical scheduling literature, by developing a practical model to incorporate into surgery scheduling tools such as OPERA. To test our method, we applied our framework to predict the surgery durations at Montreal Jewish General Hospital (JGH). The results show that, on average, our predictive model can improve the accuracy of the currently used surgery duration prediction practice at JGH (which is to take the average of the last 10 similar surgeries) by 31%. For 19% of the type of surgeries we predicted, the improvement in the accuracy reached up to 45%. These results show a significant improvement over other predictive models discussed in the state-of-the-art literature to date; for example, Stepaniak et al. (2009) reported a 15% improvement in the accuracy of the prediction using their distribution-fitting technique, and in Kayis et al. (2012), while prediction accuracy was somewhat improved by incorporating operational and temporal factors, their predictions largely vary, rendering distribution-fitting methods somewhat inapplicable to practical operating room scheduling problems.

Furthermore, we studied the impact of surgical scheduling-related features such as type of block assignment, sequencing of surgeries, and the number of surgeries scheduled in the block, on surgery durations. Interestingly, not only considering these attributes can result in better prediction of surgery durations, but also the results have important implications for surgical scheduling because these attributes are often the prominent decision-making variables in surgical scheduling problems.

The primary goal of this paper was to develop an integrative framework for the prediction of surgery durations. For future work, we suggest that researchers incorporate the relationships discovered in our prediction models into analytical models or simulations that are developed for surgical scheduling. Furthermore, due to data limitations, we were unable to incorporate other features that have been reported to impact surgery durations in our model; attributes such as patient comorbidities or risk indexes, surgeon attributes such as age and experience, and operational attributes such as the surgical team composition, can be easily incorporated to our model to improve the accuracy of the predictive model.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Lawrence Rosenberg, the Executive Director of Jewish General Hospital, for his insightful comments and providing access to the hospital data collection. We would also like to sincerely thank the administration staff and surgeons at the Montreal Jewish General Hospital for their support. This work was supported by the "Natural Sciences and Engineering Research Council (NSERC) CREATE research" grant and the "Subvention pour projets de développement stratégique innovants 2012-2013 of the Fonds de recherche du Québec-Santé (FRQS)” grant.

CHAPTER 6 ARTICLE 3: RISK OF OPERATING ROOM OVERTIME: A PROBABILISTIC LEARNING APPROACH

Elnaz Karimi[1], *Jean Marc Frayret*[1], *Michel Gendreau*[1], *Vedat Verter*[2]

Submitted to journal of *"Health Care Management Science"*

Abstract: Operational failures in hospitals such as operating room overtime result in significant adverse consequences such as additional costs, lower quality of care and patient and staff dissatisfaction. Therefore, at the operational level it is important to predict and avoid decisions that have a high risk of failure. In this paper, we apply probabilistic machine learning techniques to the operating room overtime problem. We show that the proposed algorithms are capable of classifying operating room schedules that result in overtime with an accuracy of 88% when applied to real hospital data. The predictive performance is further improved through the use of calibration techniques applied to the output of machine learning algorithms. The proposed risk model has significant implications for practice and operational level resource scheduling literature. First, the proposed risk model can easily be integrated into operating room scheduling systems at the hospital which ultimately assist decision makers in avoiding risky schedules. Second, the proposed risk model may be used in conjunction with existing operating room scheduling models to improve the operational performance of commonplace solutions.

Keywords: Operating Room Overtime; Operating Room Scheduling; Predictive Modeling; Probabilistic Supervised Learning; Calibration Techniques

[1]Department of Mathematics and Industrial Engineering, Polytechnique Montreal University, Montreal, Canada

[2]Desautels Faculty of Management, McGill University, Montreal, Canada

6.1 Introduction

Operating rooms (ORs) are among the most scarce and expensive resources in hospitals and their operation has significant impact on hospitals' performance metrics such as revenue, patient throughput, utilization of ORs, and its upstream and downstream resources and staff overtime costs (Fixler and Wright, 2013). Delays in OR scheduling results in overtime shifts for OR staffs, causing a high labor cost for the compensation of staff, nurses and anesthesiologists. Moreover, OR overtime amplifies and propagates inefficiency to downstream resources such as the PACU, wards and ICUs. In this paper we use machine learning techniques to develop a decision support system in the form of a risk model that identifies the OR schedules with high risk of overtime.

The significant costs and burdens associated with the over-utilization of ORs have been empirically investigated and documented in operations management and medical literature (Marcon and Dexter, 2006; McIntosh et al., 2006; Tyler et al., 2003; Venkataraman et al., 2018). For example, McIntosh et al. (2006) find that the costs associated with overtime are 1.5 to 2 times higher than the costs associated with under-utilization of the OR. Meanwhile, Lamiri et al. (2008b) sets the overtime cost to 500 euro per hour in their stochastic model for OR schedule planning model with the objective to minimize the expected overtime costs. Interestingly, OR overtime implications are not only in terms of labor costs, but also the intangible costs sustained by hospital staff and surgeons through job dissatisfaction from uncertain work schedules (Olivares et al., 2008; van Veen-Berkx et al., 2016b; Venkataraman et al., 2018). In addition, overtime causes stress in surgeons and other staff due to working longer than their regular shift which, in turn, reduces operating room efficiency (Guerriero and Guido, 2011). To make matters worse, from the patients' perspective, cancellation or rescheduling of surgical cases due to risk of overtime increases patient complaints and reduces overall patient satisfaction (Olivares et al., 2008). Remarkably, a recent article in CBS News documented the heavy frustration of a woman with spinal stenosis, a painful chronic condition, whose surgery was canceled by the hospital even though the patient was on a gurney on her way to the operating room, only to avoid OR overtime (CBCNews, 2018).

Despite its importance, due to the inherent uncertainty and variability of surgical procedure durations, efficient and effective OR scheduling is very challenging (Cardoen et al., 2010). Hence, a large body of operations research literature is dedicated to optimizing OR scheduling and sequencing, with the objective of minimizing OR overtime (Cardoen et al., 2010). The critical feature of current optimization methods is the incorporation of the inherent uncertainty of surgical procedure durations through developing stochastic and robust optimization models (Hosseini and Taaffe, 2015; Jebali and Diabat, 2015). In other words, current OR scheduling methods assume certain distribution functions to incorporate all variances and uncertainties. However, since these scheduling methods bundle all features that cause OR variances into a single probability distribution, they yield suboptimal results. Therefore, even though current OR scheduling methods are designed to minimize the expected overtime, they often still face the risk of overtime (Hans et al., 2008).

In this paper, we develop an online decision support system that identifies the OR schedules with high risk of overtime. We leverage the advances in machine learning and existing massive data collected by hospitals to propose a system that learns from previous OR schedules to identify schedules that run a high risk of overtime. The proposed support system is designed for integration with hospitals' master operating room scheduling systems to identify and flag schedules with high risks of overtime. The information can help decision makers understand the overtime risk associated with any OR schedule they devise, and make the necessary adjustments to lower the risk of overtime and lead to better OR operational performances. If deemed necessary, the adjustments to the OR schedule may be performed well in advance of the surgery dates, and may range from a simple reduction in the number of scheduled surgical procedures to more sophisticated rescheduling schemes.

In this study, we first use several binary classification machine learning algorithms, including Random Forest, Support Vector Machine, and Feedforward Neural Network to train effective classifiers that predict whether an OR schedule will result in overtime or not (Bishop, 2006). For this, we leverage features derived from a hospital's operational information, including operational and scheduling parameters, to obtain a trained model with a higher prediction accuracy. Next, we use the isotonic regression calibration technique to map the classification results of developed algorithms to the posterior probabilities of risk of overtime (Zadrozny and Elkan, 2002). The reader may refer to Bishop (2006) and Zadrozny and Elkan (2002) for a detailed analysis of these methods.

By applying these steps to a dataset from a major Canadian hospital, we empirically show that an uncalibrated Feedforward Neural Network is able of predicting with an AUC [3] of 91.25% and that the isotonic regression calibration method improves the performance of Random Forest and Support Vector Machine classifiers by a margin of 6.5% to 7% in predicting the risk of overtime of different OR schedules. In summary, our risk assessment model predicts the probability of overtime for an OR schedule with an excellent AUC of 97.85%.

The rest of the paper is organized as follows. Section 6.2 discusses the current state of the art related to our research. In Section 6.3, we describe the machine learning algorithms and the

[3] AUC is an estimate of the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

calibration technique used to develop the overtime risk assessment system. In Section 6.4, we describe the research setting of the study, including the dataset used, preprocessing of the dataset and the performance evaluation metrics. Section 6.5 presents the training details of the machine learning algorithms, the results of the classifiers derived from the hospital data, the results of applied calibration technique, and provides a comparison of the performance of proposed methods with a Logistic Regression benchmark model. Section 6.6 concludes the paper with a discussion on future work that addresses limitations of the current method and data and potential avenues for future research.

6.2 Literature Review

We develop an OR schedule overtime risk assessment model which predicts the risk of overtime using machine learning. Therefore, our work is closely related to the literature on developing risk models using machine learning in healthcare as well as OR scheduling optimization models. In this section, we provide a review of these streams of literature. The key drivers of OR overtime, and the main challenges in developing efficient OR schedules are related to the uncertainties associated with the operations at each stage of the patient flow. These uncertainties include, but are not limited to, the duration of surgical procedures, patient recovery period, congestion in downstream resources and arrival of emergency patients (Epstein and Dexter, 2002; Samudra et al., 2016; Wullink et al., 2007). Therefore, operations research scholars have developed stochastic models that address the problem of overtime in an uncertain environment through either setting the objective of the model to minimize the expected costs of OR overtime (Adan et al., 2011; Choi and Wilhelm, 2014; Meskens et al., 2013) or considering soft overtime capacity constraints in their models (Guinet and Chaabane, 2003; Jebali et al., 2006). In fact, overtime minimization is the most frequently used performance measure of operating room scheduling optimization models in the literature (Samudra et al., 2016). However, due to the uncertainty and the highly variable nature of surgical procedure durations, surgical schedules developed by proposed stochastic models still run the risk of overtime. This is because most proposed stochastic models incorporate uncertainty of surgical procedure durations by assuming a probabilistic distribution based on historical data (May et al., 2000; Spangler et al., 2004). Nonetheless, various studies have shown that the surgery durations often depend on complex factors such as patient characteristics, scheduling characteristics, operating surgeon, surgical team, etc. (Eijkemans et al., 2010; Kayis et al., 2012; Strum et al., 2000). To address the issues and complexities with predicting surgical procedure durations on an individual basis, a few studies have dealt with the risk of overtime based on a simple normal approximation of the sum of surgical procedure durations in an OR block, consisting of multiple surgeries (Shylo et al., 2012). The idea is that contrary to predicting surgery durations on an individual basis, predicting the durations for the entire OR block benefits from additive variances which, ultimately, leads to a better overall prediction. However, a more thorough and accurate approach to improve the effectiveness of proposed stochastic models is to take advantage of all characteristics that may influence surgery durations to develop prediction models. Our work leverages advances in machine learning to develop an integrative predictive risk assessment model which incorporates many OR and patient characteristics to minimize the risks associated with overtime in OR schedules.

Risk models are becoming increasingly popular in the field of healthcare operations management and are applied to a wide range of applications such as predicting the risk of readmission for discharged patients (Helm et al., 2016a; Kansagara et al., 2011; Kulkarni et al., 2016; Shulan et al., 2013), predicting the risk of adverse clinical outcome of patients (Jen et al., 2011), identifying patients at higher risk of diseases (Chen et al., 2017), etc. Traditional risk models were often developed using Logistic Regression; however, advances in machine learning algorithms have provided an opportunity to develop risk models that predict well-calibrated probabilities that achieve more accurate discrimination among classes of variables. Such risk models often rely on applying post-processing calibration techniques to the results of machine learning classification algorithms (Gaudoin et al., 2015; Naeini et al., 2015). Calibration techniques map the outputs of classification methods to their posterior class probabilities. Most well-known calibration techniques are Platt Scaling and Isotonic Regression. Our model employs the Isotonic Regression calibration technique (Zadrozny and Elkan, 2002) to map the classification results of developed algorithms to the posterior probabilities of risk of overtime.

Isotonic Regression is the most commonly used non-parametric calibration methods applied to supervised learning algorithms; it has been shown to perform very well when applied to binary classification algorithms such as ours (Caruana and Niculescu-Mizil, 2006). In general, machine learning algorithms have been shown to perform very well when applied to classification problems (Caruana and Niculescu-Mizil, 2006). Nevertheless, there are many prediction and decision-making applications where predicting well-calibrated probabilities to develop accurate risk models is even more critical than the discriminative capabilities of these algorithms. The isotonic regression is one of many post-processing algorithms that has been developed to produce well-calibrated probabilistic predictions, when applied to existing machine learning algorithms' outputs (Naeini et al., 2015).

This is the first paper to address the problem of predicting the risk of OR overtime in operating room scheduling. We add to the rich operating room scheduling literature by proposing the first method that predicts well-calibrated probabilities of overtime using machine learning techniques, which can be integrated into stochastic optimization models. Employing our model minimizes the expected costs associated with OR overtime and develops more efficient OR schedules. Furthermore, we show that by applying the isotonic regression, a post-processing method, to supervised learning algorithms, a well-calibrated classification model is attainable. Specifically, we also show the effectiveness of using the isotonic regression calibration method in boosting the prediction accuracy of Random Forest and Support Vector Machine when applied to predicting the risk of overtime in different OR schedules.

6.3 Methodology

Binary classification supervised problems focus on differentiation between two classes of data. Namely, to differentiate (or group) data based on their common features. For example, an algorithm may classify patients based on whether they are candidates for hospital readmission or not, or mortgage owners based on whether they will default or they will payback their loans. In problems that prediction tools are used for decision making regarding high risk contexts such as credit card fraud, patient readmission and OR overtime, predicting the probability of belonging in each of the binary classes accurately is especially important. In fact, the probabilities provide a level of confidence on the prediction. These probabilities are known as risk models. For example, in the case of an OR overtime, accurately predicting the probability of an OR running overtime for a particular schedule (i.e. belonging in the "overtime" class or not) is crucial for effective scheduling.

One way to achieve this goal is to post-process the output of machine learning classification models to obtain more accurate probabilities. Such post-processing algorithms are refereed to as *"calibration methods"*. In this section, we describe the different methods that will be used for predicting the probability (risk) of overtime of OR schedules . First, we will describe the binary classification methods used to classify the OR schedules to "over-time" or not. Second, we will describe the method used to calibrate the classification methods to derive the probability (risk) of overtime for each schedule. Logistic Regression and Feedforward Neural Network classifiers can predict probabilities on their own and do not require post-processing calibration. In fact, empirical evidence has shown that FFNN is capable of predicting well-calibrated probabilities on its own using sigmoid output activation functions and does not need post-training calibration (Niculescu-Mizil and Caruana, 2005). However, for calibrating the probabilities of output produced by Random Forest and Support Vector Machine, we use the well-known Isotonic Regression method.

6.3.1 Supervised Learning Methods

This section briefly describes the supervised learning methods used for the binary classification problem of surgical schedule overtime (Additional details can be found in the cited references). With only two labels ("overtime" or "no overtime"), the problem of predicting whether a surgical schedule results in overtime or not is the classic binary classification problem. Supervised machine learning methods learn complex and often non-linear prediction patterns from large training datasets and apply these patterns to new data to make outcome predictions. We consider the following well-known methods for the classification problem: Logistic Regression, Random Forests, kernel-based Support Vector Machine (SVM) and Feedforward Neural Network (FFNN). We chose these methods because they are known to perform well in similar binary classification problems.

Logistic Regression

Logistic Regression is a popular class of generalized linear regression models that is commonly applied to classification problems that aim to predict the posterior probability of the occurrence of an *"event"* as a function of a number of predictor variables (Bishop, 2006). For example to predict whether or not a specific OR schedule runs a risk of "overtime", the Logistic Regression is employed to estimate the probability of *"overtime"* (positive class) versus *" no overtime"*(negative class). The probability that the OR schedule belongs to the positive class is then estimated by:

$$Prob(overtime|X = x) = \frac{1}{1 + e^{\beta_0 + \beta x}} \tag{6.1}$$

where, x is the input vector, β is the weight vector and β_0 is the bias. Beta's are estimated using the gradient descent algorithm.

Random Forest

Random Forest is a popular machine learning ensemble method for solving a wide range of classification problems. Ensemble learning methods operate by generating many classifiers and then aggregating the prediction results (Liaw et al., 2002). Random Forest constructs each tree by drawing n_{tree} bootstrap samples from the original data and splits each node using the best option among a subset of predictors (m_{try}) randomly chosen at that node. Finally, the algorithm classifies new data by aggregating the results of classification of all the trees, called *"majority vote"*. At each bootstrap iteration, the error rate of the grown tree is calculated by predicting the data not in the bootstrap sample, called ("*out-of-bag*", or OOB data). Finally, the OOB predictions are aggregated and the total error rate, called the OOB estimate of error rate, is calculated.

Support Vector Machine

Support Vector Machines (SVM) are binary classifiers that construct an optimal separating hyperplane between two classes of data (Boser et al., 1992). The main idea behind binary linear SVM classifier is that the best linear classifier is the one that maximizes the margin between a separating line and the nearest data points. SVM use an implicit mapping Φ of the input data into a high-dimensional feature space using what is called a kernel function. A kernel function returns the inner product $\langle \Phi(x_i), \Phi(x_j) \rangle$ of the images of two data points x_i, x_j in the feature space. The SVM then solves a constrained quadratic programming problem that maximizes the margin of separation between the two classes. The hyperplane of SVMs is denoted as:

$$\langle w, K(x_i, x_j) \rangle + b = 0 \tag{6.2}$$

which corresponds to the decision function:

$$h(x) = \mathrm{sign}\left(\langle w, K(x_i, x_j) \rangle + b\right) \tag{6.3}$$

where sign denotes the signum function, which extracts the sign of a real number.

The hyperplane is constructed by solving a constrained quadratic optimization problem whose solution is $w = \sum_1^m \alpha_i y_i \Phi(x_i)$, where m is the size of of training input and $y_i = \pm 1$ represents the corresponding two classes. The optimization problem is as follows:

$$\begin{aligned} &\text{minimize} && t(w, \xi) = \frac{1}{2} \|w\|^2 + \frac{C}{m} \sum_{i=1}^{m} \xi_i \\ &\text{subject to} && y_i \left(\langle \Phi(x_i), w \rangle + b\right) \geq 1 - \xi_i \quad (i = 1, \ldots m) \\ &&& \xi_i \geq 0 \quad (i = 1, \ldots m) \end{aligned} \tag{6.4}$$

The data points that lay on the boundaries between classes are called *support vectors* and carry the relevant knowledge regarding the classification problem. Furthermore, the cost parameter C denotes the penalty of misclassifying a training point and thus the complexity of the prediction function. A high value of C results in a complex prediction function that misclassifies as few training points as possible, while a low value of C results in a simpler prediction function. Therefore, SVMs that are built solving Equation 6.4 are called C-SVM (Karatzoglou et al., 2005). Finally, commonly used kernel functions include the linear kernel, the polynomial kernel, and the Gaussian radial basis function (RBF) kernel (Suykens and Vandewalle, 1999).

Feedforward Neural Network

Feedforward Neural Networks (FFNNs) are popular machine learning methods applied to a wide range of classification problems. As a type of supervised learning, FFNN performs the classification task by learning non-linear complex patterns between a set of predictors which are guided by their corresponding "*class label*" target (Goodfellow et al., 2016). FFNNs architecture consists of a hierarchical layered network, each layer consists of number of *"neurons"* that contain the basic computations of the neural network. The main body is composed of an input layer, a number of hidden layers and an output layer. The information propagates forward from the input layer through the network to the -*"neurons"*- in the hidden layers and output layer, and finally is converted by a transfer function to obtain the"*class label*" outputs.

FFNN learns using the *"back propagation"* algorithm which is based on the gradient descent algorithm. In this algorithm the prediction error is traced back through the network in order to appropriately update neuron's weight matrices through computing the gradient incrementally by "propagating" backwards through the network. Thorough back propagation, training samples are passed through the network and the outputs are compared to the actual outputs. The weight vectors connecting adjacent neurons are updated at each step such that the prediction error is minimized to produce a network that results in the best predictive performance.

6.3.2 Calibration with Isotonic Regression

A few approaches have been proposed to calibrate machine learning classification algorithms. These methods map the outputs of classification methods to their posterior class probabilities. The most notable calibration methods are Platt Scaling and Isotonic Regression. Platt Scaling was first introduced by Platt et al. (1999) to transfer the outputs of a binary Support Vector Machine classifier to their posterior probabilities by passing them through a Sigmoid function. Other researchers have successfully applied this method to calibrate other classification techniques including Boosted Trees and Random Forest (Caruana and Niculescu-Mizil, 2006; Gaudoin et al., 2015). However, Platt scaling is most effective when the predicted probabilities are sigmoid shaped (Niculescu-Mizil and Caruana, 2005). Therefore, in this paper, we apply the more general and commonly used approach of Isotonic Regression. Isotonic Regression is a nonparametric regression method that calibrates predictions of binary classification algorithms such as SVM, Naive Bayes and decision trees using an isotonic (monotonically increasing) mapping function to map the outputs of classifiers to their posterior probabilities (Zadrozny and Elkan, 2002).

This calibration method uses the predictions h_i from a classifier and the true targets y_i such that:

$$y_i = m(h_i) + \epsilon_i \tag{6.5}$$

where m is a non-increasing function. The objective of the regression then is to find an $\widehat{m}$ that minimizes the mean squared error:

$$\widehat{m} = argmin_g \sum (y_i - g(h_i))^2 \tag{6.6}$$

Following the literature, we use the "Pool Adjacent Violators Algorithm" (PAVA) to find a stepwise constant solution to the Isotonic Regression applied to the aforementioned supervised learning methods (Barlow, 1972; Niculescu-Mizil and Caruana, 2005). PAVA finds $\widehat{m}$s that are monotonically increasing by minimizing the mean square error. It achieves calibration by successively "pooling adjacent violators", i.e., probability estimates that violate the monotonicity criterion. The algorithm first sorts the training set (h_i,y_i) according to h_i. Note that h_i of SVM model is the classification score of each instance, where the score is the signed distance from the input to the decision boundary such that if the score is negative then the instance is labeled as "no overtime" and if it is positive then the instance is labeled as "overtime". While for Random Forest h_i is the average of the unweighted class votes of all the trees. If h_i is an "overtime" instance then the initial estimation of the $\widehat{m}(h_i)$ function is: $\widehat{m}(h_i) = 1$, otherwise if h_i is an "no overtime" instance then the initial estimation of $\widehat{m}(h_i)$ function is: $\widehat{m}(h_i) = 0$. If the initial $\widehat{m}$ is isotonic (strictly increasing), then the function is learned and these values are considered to be the final estimates of the probabilities of the two classes. Otherwise, the algorithm looks for two adjacent instances (h_{i-1}, h_i) that violate the isotonic assumption. Both probabilities $\widehat{m}(h_{i-1})$ and $\widehat{m}(h_i)$ are then replaced by the average of the two, such that $\widehat{m}(h_{i-1})$ and $\widehat{m}(h_i)$ no longer violate the isotonic assumption. The algorithm is then repeated until all estimated probabilities are isotonic and there remains no instances that violate the assumption. The final vector of $\widehat{m}$ renders the probabilities of overtime.

Furthermore, to avoid bias we use a separate validation set to train the Isotonic Regression function. It is important to note that Isotonic Regression has greater flexibility compared to other calibration techniques such as Platt Scaling. However, when data is scarce, Isotonic Regression is prone to overfitting (Menon et al., 2012).

6.4 Research Setting

To test the effectiveness of the proposed machine learning algorithms in predicting the risk of overtime for OR schedules, in this section we apply these methods to a real-life setting.

6.4.1 Learning Dataset

Our dataset consists of 14,104 surgery records that took place between April 2013 to December 2014, at Montreal's Jewish General Hospital. These surgeries span over 14 surgical specialties where 95 distinct surgeons performed more than 700 distinct surgical procedures in 17 operating rooms. The hospital's OR scheduling system is designed based on the demand of each surgical specialty 6 weeks in advance of each month. The OR schedule at JGH distributes the OR blocks (AM and PM) among various specialties. The heads of surgeries for each specialty distribute the time among their own surgeons. Each surgeon is then responsible for sequencing his/her patients such that all scheduled surgeries are completed within their assigned OR block.

The original dataset provided the following information for each surgical procedure record: date requested, date of surgery, surgeon, specialty, procedure, room, patient entrance to OR time-stamp, patient departure from OR time-stamp, operation start time, operation finish time. Note that patient entrance and departure to and from the OR time-stamps are different from the operation start and finish time. Patient entrance and departure marks the entire time the patient is utilizing the operating room, while operation start and finish time is the time the surgeon is performing the surgery on the patient.

6.4.2 Data Preprocessing

For the purpose of this study, we have removed: 1) the surgical blocks that where assigned to emergency cases, 2) the elective surgical blocks that were interrupted by the arrival of emergency cases, 3) the surgical blocks in which only a single surgical procedure was performed. This preprocessing left us with roughly 81% of our original dataset. Furthermore, for each surgical block schedule, we extracted a number of operative features from the corresponding scheduled surgical procedures.

Overtime Tagging

Figure 6.1 provides the distribution of the OR closing time (i.e., time-stamp of last patient's departure) at the hospital during the period of the study. As explained in Section 6.4.1, for each OR it is possible that its AM and PM blocks are assigned to different specialties or both AM and PM blocks are assigned to a single specialty. In the latter case, we combine the AM and PM blocks into a single training sample block for the following reason: if the OR is assigned to a single specialty for the entire day, then it is unknown (and unimportant) at which point in the day the delays occurred. This is because the total overtime is calculated at the end of the day since the staff is compensated for working the entire day. Conversely, when the AM and PM blocks are assigned to different specialties, then any delay in the AM block propagates to the PM block, which increases the risk of overtime in the PM block. Also, since the staff is changed when the AM and PM blocks are assigned to different specialties, and is separately compensated for the overtime they worked in their shift, the risk of overtime is more pronounced.

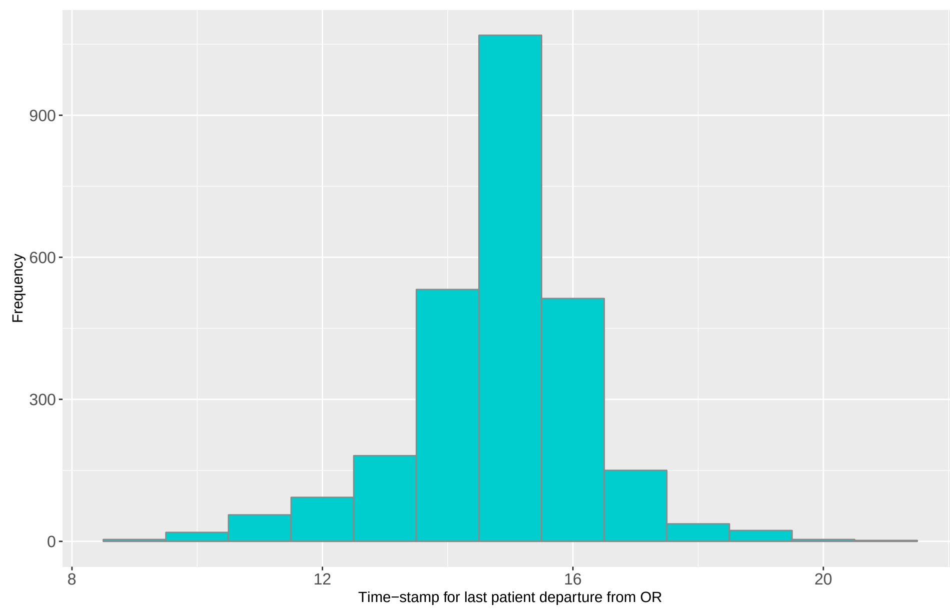


Figure 6.1 Distribution of OR closure time

We label the surgical blocks where the finish time of the last surgical procedure exceeds the normal OR operating time as “overtime”. Table 6.1 provides a summary of the instances of overtime of operating rooms at the hospital.

Table 6.1 Summary of OR block overtimes

Specialty	Overtime Blocks	Total Blocks	Total Surgical Procedures
Breast Oncology	10	109	517
Cardiac	32	34	67
Colo-rectal	35	134	342
Dental	5	29	80
E.N.T.	52	336	1244
General surgery	43	326	1286
Gynecology	44	228	708
Neurosurgery	20	40	96
Ophthalmology	37	529	4431
Orthopedics	124	484	1491
Plastics	6	73	215
Thoracic	9	34	87
Urology	36	211	584
Vascular	36	116	320
Grand Total	**489**	**2683**	**11468**

Features Extraction

One crucial step is to extract information from the data that contributes to the accurate classification of schedules which pose risk of overtime from schedules with no such risk. We extract these information from the dataset by defining the following input variables from the corresponding surgical database of the hospital. The input variables include:

- Block Type: whether a block is an all day, AM or PM block.
- Surgeon Mix: Whether a block is assigned to a single surgeon or multiple surgeons.
- Surgeon Sequence: The sequence of surgeons that are scheduled to perform surgery in each block.
- Procedure Count: Number of surgical procedures in each block.
- Block-Mix: The mix of surgical procedures assigned to each block.

- Surgical Sequences: The sequence of surgical procedures in a block.

Other features such as specialty, room and pool of patient visit types have also been considered.

6.4.3 Model Performance Evaluation

Following the literature, we base the evaluation of the performance of our four machine learning algorithms on metrics driven from the confusion matrix (Caruana and Niculescu-Mizil, 2006). A confusion matrix is built based on the classification results. Each predicted class is compared with its actual class for each instance to calculate four metrics:

- *True Positives (TP)*: the number of positive instances that are correctly classified as positive classes.
- *False Positives (FP)* : the number of negative instances that are incorrectly classified as positive classes.
- *True Negatives (TN)*: the number of negative instances that are correctly classified as negative classes.
- *False Negatives (FN)*: the number of positive instances that are incorrectly classified as negative classes.

Using the confusion matrix, we can drive the performance evaluation metrics as follows:

$$\mathit{Error} = \frac{FN + FP}{TN + TP + FN + FP} \tag{6.7}$$

$$\mathit{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{6.8}$$

$$\mathit{Sensitivity} = \frac{TP}{TP + FN} \tag{6.9}$$

$$\mathit{Specificity} = \frac{TN}{FP + TN} \tag{6.10}$$

Moreover, the confusion matrix measures can be used to construct a plot (Receiver Operating Characteristic or ROC curve), to show the trade-off of FN and FP rates to model the classification errors. Therefore, in this paper, we also use the area under ROC curve (AUC) to evaluate how well the methods discriminate the positive and negative instances in the feature space (Niculescu-Mizil and Caruana, 2005).

To evaluate the performance of the Isotonic Regression calibration on Random Forest, SVM and FFNN, we utilized three performance measure: root mean square error (RMSE), expected calibration error (ECE) and the maximum calibration error (MCE) (Niculescu-Mizil and Caruana, 2005). ECE and MCE are calculated by partitioning the output space of the binary classifier, which is the interval [0, 1], into K fixed number of bins (we consider K = 10 in our models). The estimated probability for each instance will be located in one of the bins. For each bin we can define the associated calibration error as the absolute difference between the expected value of predictions and the actual observed frequency of positive instances. ECE and MCE are calculated as follows:

$$ECE = \sum_{k=1}^{K} P(k).\,|o_k - e_k| \tag{6.11}$$

$$MCE = \max\left(|o_k - e_k|\right) \tag{6.12}$$

where $P(k)$ is the empirical probability or the fraction of all instances that fall into bin k, e_k is the mean of the estimated probabilities for the instances in bin k, and o_k is the observed fraction of positive instances in bin k. The lower the value of ECE and MCE, the better is the calibration of a model.

6.5 Training Details and Results

To train the Random Forest models, we need to specify the total number of classification trees (n_{tree}) and the total number of features to try at each node split (m_{try}). Choosing small values of n_{tree} and m_{try} may results in poor model performance on the training dataset, whereas large values may result in overfitting of the model on test data. overfitting results in poor performance of model when applied to new data and, hence, cause the model to lack generalizability. Therefore, we develop Random Forest models using 4, 8, and 12 features at each split and 500, 1000 and 2000 trees to tune both parameters.

The training of the algorithm involves tuning the following hyperparameters including the C cost value and kernel function's parameters. A high value of C results in a more complex prediction function with lower misclassification error on the training dataset, however, it may result in overfitting of the model and hence, high error rate on the test dataset. Alternatively, a low value of cost C will lead to a simpler prediction function, and hence a higher error rate on the training dataset. Therefore, for the purpose of tuning the cost value C, we perform a grid search over the interval [1; 1000]. Moreover, we use the radial kernels in this study as they have been proven to deliver excellent results when applied to high dimensional problems (Karatzoglou et al., 2005).

We develop the FFNN models using the ReLU activation functions for hidden units and sigmoid function as the output activation function (Guo et al., 2017). We then optimize the network's parameters using the stochastic gradient descent, halving the learning rate every 50 epochs and setting the initial learning rate at 0.1. Moreover, to find the optimal architecture of the network, we performed a grid search over the number of hidden layers (in the range 1-3) and over the number of hidden nodes, choosing between 4, 8 and 12.

For our experiments, we split the data to 75% and 25% for training and testing, respectively. Each method was implemented in R using publicly available libraries. We build the models using the training data, then we optimize the supervised learning models using 10-fold cross validation (Bishop, 2006). The 10-fold cross-validation helps us to tune parameters, and therefore, to prevent overfitting of the algorithms. This methodology partitions the data randomly into 10 subsets with the same distribution of both classes ("overtime" and not). At each iteration, 9 subsets (folds) are used for training the algorithm and the remaining fold is used for testing the algorithm. The fold that was used for testing in the first iteration is replaced by another fold at the next iteration, and so on until all the folds are used for testing the algorithm.

Table 6.2 presents the comparison of performances of developed Random Forest models corresponding to combination of tuning parameters m_{try} (number of features considered at each split) and n_{tree} (number of trees). The performances appear very robust across the tuning parameters. The combination with the best performance is shown in bold font and is selected as the optimal Random Forest model to be compared with other methods. The final model yields a sensitivity of 85.17% and a specificity of 90.41%. Thus, the final Random Forest classifier detects surgical schedules with risk of overtime very well. However, slightly more "overtime " schedules will be misclassified as "no overtime" than there will be "no overtime" misclassified as "overtime" schedules.

Table 6.2 Performance evaluation of Random Forest models

n_{tree}	m_{try}	Error Rate	Sensitivity	Specificity	AUC
500	4	0.1284	0.8356	0.9004	0.9109
500	8	0.1329	0.8413	0.9015	0.913
500	12	0.1299	0.8438	0.8997	0.9275
1000	4	0.1282	0.8494	0.9068	0.9147
1000	**8**	**0.1275**	**0.8517**	**0.9041**	**0.9277**
1000	12	0.1286	0.8498	0.9026	0.9183
2000	4	0.1327	0.8406	0.8988	0.9139
2000	8	0.1283	0.8428	0.8971	0.9145
2000	12	0.1335	0.8356	0.9018	0.9255

Table 6.3 compares the performances of FFNN models developed for combinations of 1:3 hidden layers and 4, 8 or 12 number of hidden nodes. Similar to Random Forest models, FFNN performance is quite robust across different network structures. The combination with the best performance is shown in bold font and is selected as the optimal FFNN model. The optimal FFNN model yields a sensitivity of 84.03% and a specificity of 92.31%.

Table 6.3 Performance evaluation of Feedforward Neural Network Models

Hidden Layers	Hidden Nodes	Error Rate	Sensitivity	Specificity	AUC
1	4	0.1327	0.8222	0.9044	0.9358
1	8	0.1297	0.8370	0.9141	0.9271
1	12	0.1228	0.8292	0.9042	0.9335
2	4	0.1282	0.8375	0.8980	0.9311
2	**8**	**0.1185**	**0.8403**	**0.9231**	**0.9361**
2	12	0.1213	0.83	0.9095	0.9263
3	4	0.1292	0.8229	0.8987	0.9256
3	8	0.1195	0.8367	0.9164	0.9258
3	12	0.1287	0.8372	0.9011	0.9329

We compared the performance of the optimal Logistic Regression, Random Forest, SVM and FFNN classifiers on both the training and the test datasets. Table 6.4 presents the comparison of these classifiers. As shown, Random Forest, SVM and FFNN all outperform the Logistic Regression model by a large margin. While FFNN outperforms the Random Forest and SVM classifiers in terms of error rate, specificity and AUC, Random Forest has a better performance than FFNN in terms of sensitivity (correctly classifying the overtime schedules).

Table 6.4 Performance comparison of classifiers in predicting overtime schedules

	Error Rate		Sensitivity		Specificity		AUC	
	Training	Test	Training	Test	Training	Test	Training	Test
LR	0.3634	0.4321	0.6469	0.5438	0.6906	0.6119	0.6581	0.6063
RF	0.1275	0.1387	0.8517	0.8311	0.9041	0.8873	0.9277	0.9024
SVM	0.1391	0.1303	0.8183	0.7964	0.859	0.8137	0.9052	0.8898
FFNN	0.1185	0.129	0.8403	0.8258	0.9231	0.8997	0.9361	0.9125

Figure 6.2 shows the performance of classifiers when both sensitivity and specificity are considered together. Logistic Regression is dominated by all three other classifiers. While both FFNN and Random Forest dominate SVM in terms of sensitivity and specificity, Random Forest dominates FFNN in terms of sensitivity.

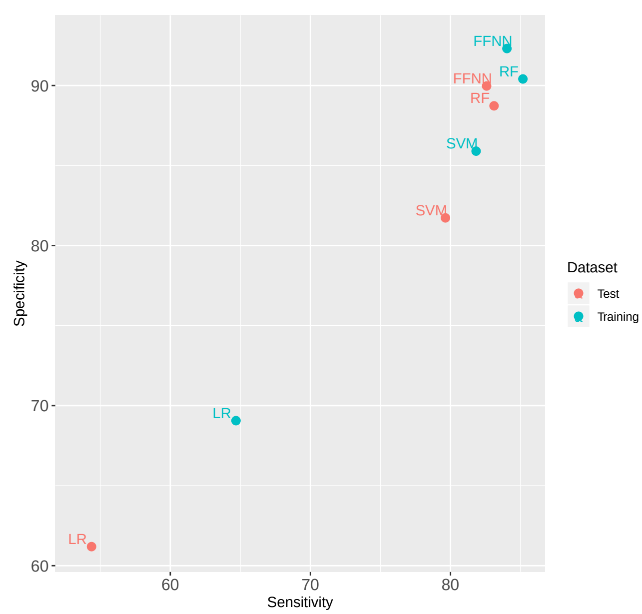


Figure 6.2 Performance comparison of classifiers in terms of sensitivity and specificity

Table 6.5 compares the performance of LR and FFNN with calibrated RF and SVM models as measured by AUC, RMSE, ECE and MCE. As shown, the calibration of RF and SVM models improves their performance of (AUC) by a margin of 6.5% and 7%, respectively. In fact, after calibration, both RF and SVM models performs better than an uncalibrated FFNN in predicting the risk of overtime. In terms of classification performance, i.e., discrimination of "overtime" and "no overtime" schedules, measured by error rate and AUC, the results show that FFNN either outperforms SVM and RF or it has a performance that is not statistically significantly different. However, as far as predicting the risk of overtime is concerned, a calibrated RF performs much better than an uncalibrated FFNN in terms of AUC and has a slightly better performance in terms of RMSE, ECE and MCE. Overall, the results show that calibrated Random Forest and FFNN can predict the risk of overtime with great accuracy.

Table 6.5 Comparison of the performance of LR and FFNN with calibrated RF and SVM in predicting the risk of schedule overtime

Model	AUC	RMSE	ECE	MCE
Logistic Regression	0.6063	0.4511	0.2444	0.2902
Random Forest: Calibrated with Isotonic Regression	0.9651	0.2647	0.0973	0.1566
Support Vector Machine : Calibrated with Isotonic Regression	0.9785	0.3167	0.1457	0.1832
Feedforward Neural Network	0.9125	0.2789	0.101	0.1619

To test whether calibration can help the LR and FFNN models to predict better probabilities, we apply the PAVA Isotonic Regression to both models. Figure 6.3 compares the MSE for each classification method before and after calibration applied to the test data. We find that calibration does not help LR and FFNN in predicting probabilities and actually it slightly hurts the performance of FFNN. This result is in line with the literature. Specifically, FFNN performs very well in predicting the risk of overtime on its own (Niculescu-Mizil and Caruana, 2005).

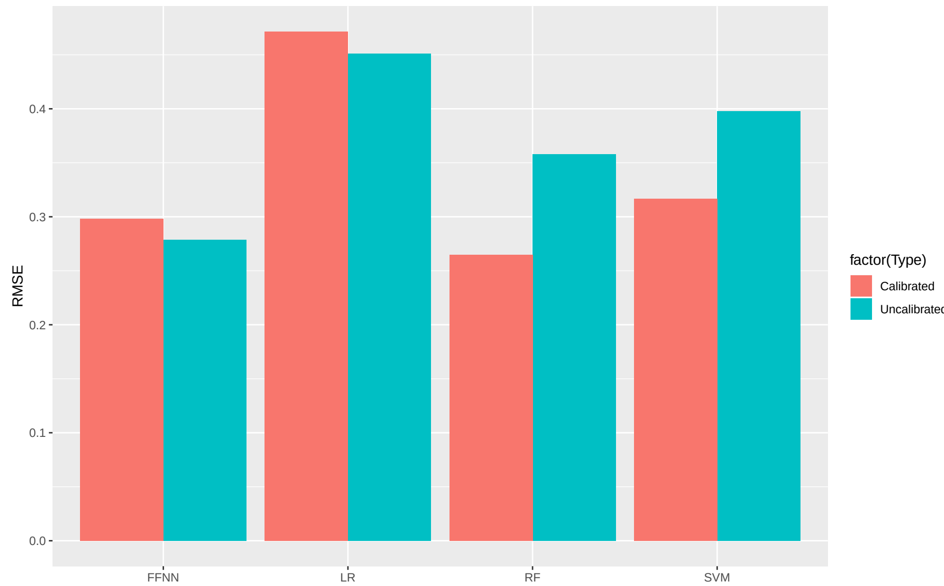


Figure 6.3 Comparison of performance of calibrated vs uncalibrated models in root mean squared error

6.6 Conclusion

OR overtime is a burden to multiple stakeholders such that it 1) brings financial burden to hospitals due to overtime labor costs; 2) complicates hospital operations due to delays at upstream and downstream resources and cancellation or rescheduling of remaining surgical operations; 3) results in job dissatisfaction of surgeons and OR staffs due to uncertain schedule; and 4) causes patients dissatisfaction due to long waits or cancellation and rescheduling of their surgeries. Therefore, many scholars in the field of operation management have attempted to develop schedules that minimizes costs associated with overtime. In this paper, we applied different learning methods to the problem of predicting the risk of overtime for developed OR schedules. Specifically, we applied machine learning techniques which learns from previous OR schedules at hospitals to predict the overtime risks associated with every OR schedule. Our model helps decision-makers understand the overtime risk, and make considerations to reduce this risk for the OR schedules they devise, well in advance of the scheduled surgeries.

To improve the performance of our learning algorithms, we applied calibration techniques to the learning techniques and examined their effectiveness in improving the performance of learning algorithms. We found that the Feedforward Neural Network model was initially the most effective in predicting the risk of overtime with no need for calibration, and was superior to uncalibrated models such as Logistic Regression, Random Forest and Support Vector Machine. Nonetheless, we found that employing a post-processing calibration technique, namely, the Isotonic Regression, to the results of our Random Forest model was able to improve its performance to the point where it surpassed the performance of the Feedforward Neural Network model. Overall, we showed the efficacy of using learning models to predict the risks associated with proposed OR schedules.

The proposed learning algorithms in this paper can learn from previous schedules and predict the risk of overtime for new OR schedules. The algorithms consider scheduling attributes such as block-mix, surgeon-mix and sequence-mix, all of which are determined by OR scheduling systems and decision makers and can easily be adjusted to reduce the risk of overtime. The algorithms are designed to be integrated into OR scheduling systems at hospitals and operate as a decision support system that warns of risky schedules. This paper is the first to address the problem of risk prediction of OR overtime. Future work may focus on the development of stochastic OR schedule optimization and simulation models in which the proposed probabilistic models of schedule overtime would replace the conventional incorporation of surgical procedure uncertainty. The significance of our proposed models will be evident for such work.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Lawrence Rosenberg, the Executive Director of Jewish General Hospital, for his insightful comments and providing access to the hospital data collection. We would also like to sincerely thank the administration staff and surgeons at the Montreal Jewish General Hospital for their support. This work was supported by the "Natural Sciences and Engineering Research Council (NSERC) CREATE research" grant and the "Subvention pour projets de développement stratégique innovants 2012-2013 of the Fonds de recherche du Québec-Santé (FRQS)” grant.

CHAPTER 7 GENERAL DISCUSSION

The hospital resource planning and scheduling problems are comprised of two components 1) accurate prediction of uncertain and variable elements in the system 2) designing decisions systems that most efficiently utilized available resources in the system in the presence of such variability and uncertainty. The operations research literature has mostly focused on the second component and used simple statistical modeling and patient grouping methods to capture the variability and uncertainty in the system; rendering the application of developed decision support systems not generalizable or scalable in real-life settings and the proposed decisions suboptimal. The complexity of predictive modeling for hospital processes arise from 1) heterogeneity of patients' needs and 2) complex interactions of resources in the hospital.

This thesis addresses the problem of predictive modeling for hospital resource planning and scheduling systems by proposing integrative predictive support systems that address variability and uncertainty of processes for strategical, tactical and operational level decision making. The proposed models can seamlessly be integrated with existing hospital resource planning and scheduling systems to provide accurate information and lead to optimal decision-making such that it increases the efficiency of utilization of scarce resources, improves quality of care and reduces operational costs.

We rely on machine learning and artificial algorithms to carefully analyze the predictors of variability and uncertainty at strategical, tactical and operational level. Most importantly we address the heterogeneity of patients' needs and complex interactions of resources in the hospital by incorporating patients' demographic, medical, temporal and operational and scheduling-related factors into our analysis. The improvement in prediction accuracy achieved by our proposed integrative frameworks has not been previously achieved by any other methods in the literature. Furthermore, our framework is the first of its kind to model the complex interactions between patients and resources and between different type of resources in the entire hospital setting.

The frameworks that have been proposed in this thesis aims to guide practitioners in their decision making processes by providing them accurate prediction of demand, estimation of service times and risk assessment of their decision making. The proposed frameworks are scalable and can be adopted by any type of hospitals that aim to improving their decision making processes to 1) govern their costs and revenue by improving resource utilization, 2) improve the quality of care by providing timely access to resources, and 3) increase satisfaction of stakeholders by mitigating the risks associated with adverse consequences of unplanned events.

At the strategic and tactical level, we propose an integrative personalized pathway prediction model that is capable of predicting demand of chronic patients for hospital resources. The framework is the first of its kind to capture the long-term relationship of hospitals with chronic patients, which span over a long-term horizon. We first develop and train a series of sequentially connected deep feedforward neural to predict the transition of a patient between treatments during his/her pathway.The proposed models are capable of predicting patient's next treatment with accuracy (measured by "recall") ranging from 68% to 79%. In addition, we develop a second series of temporal deep feedforward neural network models that predict the expected receiving time for the *next treatment*. The trained pathway and temporal predictive models are combined and incorporated into an agent-based simulation which predicts personalized clinical pathways of patients and their demand for hospitals' scarce resources during the course of their treatment.

At the tactical and operational level, we utilize supervised learning algorithms such as Random Forest and Support Vector Machine to predict surgical procedures duration. We propose a framework that extract useful information from the surgical procedure databases in the hospital including scheduling-related, operational, temporal attributes and in combination with patient specific, procedure specific and surgeon specific attributes, predicts the surgical procedures duration. The proposed framework was applied to real hospital data and the results showed an improvement of 31% in the accuracy of predictive model compared to its practice benchmark. Furthermore, the results show that scheduling-related decisions such as procedure sequencing, surgeon pooling and block-mix have significant impact on the surgical procedure duration. This result has significant implications for operating room planning and scheduling literature at both tactical and operational level; indicating that the optimal operating room planning is achieved only through joint optimization of surgical procedures durations and operating room scheduling.

At the operational level, we propose an integrative predictive model for operational failure risk assessment. Even the most accurate predictive tools still fall short in predicting variability in hospital processes with 100% accuracy. Therefore, it is important at the operational level, to avoid decisions that have a high risk of failure resulting in significant adverse consequences which can inquire additional costs, lower the quality of care and cause patient and staff dissatisfaction. In this thesis, we apply probabilistic machine learning techniques to operating room overtime problem. The framework was applied to real-life data and the result showed an 88% success rate for proposed classification techniques to identify *high-risk* schedules. The predictive performance is further improved through use of calibration techniques applied to the classification algorithms outputs. The managerial and theoretical implications include: first, the proposed risk model can easily be integrated in to operating room scheduling systems at the hospital to assist decision makers in avoiding risky schedules; second, the proposed risk model can be jointly used in existing operating room scheduling models to improve the performance of solutions.

CHAPTER 8 CONCLUSION AND RECOMMENDATIONS

This thesis presented a collection of integrative predictive support systems for hospitals that can be used to design more efficient and robust resource planning and scheduling systems. Decision making in highly volatile and unpredictable environments such as hospitals come with the inherent challenge of dealing with uncertainties and variability. heterogeneity in patient needs, service-time variability, unplanned events such as cancellation and emergency visits, and etc. are among the many examples of uncertain events. The methods employed in this thesis are based on machine learning algorithms that are able to process, analyze and extract complex relationship among personal, medical, operational and temporal data. In terms of prediction accuracy, we showed that the prediction models developed based on these algorithms decidedly outperform conventional models that employ simplistic methods.

This work can assist resource planning and scheduling tools by providing them with more accurate information, and hence improves decision-making on these fronts which in turn, can translate to improvements in operational performance for the hospital in terms of revenues and/or cost, quality, and access to care. With the massive amount of data available, and advances in the field of data analytics, the operations research community is moving towards data-driven research (Green, 2012; Simchi-Levi, 2013). Following this trend, we contribute to the literature by illustrating the application of machine learning methods to improve predictive modeling of hospital processes. Specifically, this work explores resource planning and scheduling on three fronts: strategic level (long-term), tactical level (mid-term), and operational level (short-term). At the strategic and tactical levels, we proposed a novel algorithm based on deep feedforward neural network models to predict personalized clinical pathways for patients. Additionally, we integrated the predictive algorithm in an agent-based simulation to predict long-term and mid-term demand for a hospital's scarce resources. At the tactical and operational levels, we employed supervised learning algorithms to accurately predict personalized surgical procedure durations and showed that scheduling decisions as well as temporal, operational and patient attributes are important predictors of duration of surgical procedures. Finally, on the operational level, we employed probabilistic supervised learning algorithms to develop a risk model for operating room schedule overtime.

This thesis may be seen as a step towards shedding light on novel methods for increasing the accuracy of predictive models. To the best of our knowledge, it is the first to display the efficacy of machine learning methods in resource planning and scheduling on strategic, tactical, and operational levels. Albeit, there are multiple avenues for future research. First and foremost, although we simulated the effectiveness of our tools in resource planning and scheduling, future work may attempt to incorporate the proposed predictive models in real-world scenarios at the strategic, tactical and operational level, and work towards empirical validation of such methods in improving the quality of decision-making. Another avenue of research is the application of the proposed personalized clinical pathway predictive algorithm to EHR data to assess the effectiveness of proposed algorithm in predicting patients' future clinical visits compared to current benchmarks (Choi et al., 2016a,b). Finally, current resource planning and scheduling models may be updated to include the possibility of predictive modelling using our proposed methods. Namely, developing formal models which mathematically evaluate the impact of predictive models on hospital performance on several operational measures such as of revenue/costs, utilization of resources, patient wait times, and etc. may be an interesting step forward.

With the rising demand for chronic care in the Canadian population, it is crucial for hospitals and healthcare delivery services to develop novel methods in improving their operational efficiency to remain financially viable and to serve the persons in need. We hope that this thesis contributes to achieving this ultimate goal.

BIBLIOGRAPHY

Abraham, G., Byrnes, G. B., and Bain, C. A. (2009). Short-term forecasting of emergency inpatient flow. *IEEE Transactions on Information Technology in Biomedicine*, 13(3):380–388.

Adan, I., Bekkers, J., Dellaert, N., Jeunet, J., and Vissers, J. (2011). Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. *European Journal of Operational Research*, 213(1):290–308.

Adan, I., Bekkers, J., Dellaert, N., Vissers, J., and Yu, X. (2009). Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. *Health care management science*, 12(2):129.

Alaeddini, A., Yang, K., Reddy, C., and Yu, S. (2011). A probabilistic model for predicting the probability of no-show in hospital appointments. *Health care management science*, 14(2):146–157.

Ang, E., Kwasnick, S., Bayati, M., Plambeck, E. L., and Aratow, M. (2015a). Accurate emergency department wait time prediction. *Manufacturing & Service Operations Management*, 18(1):141–156.

Ang, E., Kwasnick, S., Bayati, M., Plambeck, E. L., and Aratow, M. (2015b). Accurate emergency department wait time prediction. *Manufacturing & Service Operations Management*, 18(1):141–156.

Anjomshoa, H., Dumitrescu, I., Lustig, I., and Smith, O. J. (2018). An exact approach for tactical planning and patient selection for elective surgeries. *European Journal of Operational Research*, 268(2):728–739.

Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., and Yom-Tov, G. B. (2015). On patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems*, 5(1):146–194.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324):483–485.

Baker, A. (2001). Crossing the quality chasm: a new health system for the 21st century. *BMJ: British Medical Journal*, 323(7322):1192.

Ban, G.-Y. and Rudin, C. (2014). The big data newsvendor: Practical insights from machine learning. Available at SSRN: https://ssrn.com/abstract=2559116 or http://dx.doi.org/10.2139/ssrn.2559116.

Barlow, R. E. (1972). Statistical inference under order restrictions; the theory and application of isotonic regression. Technical report.

Barnes, S., Hamrock, E., Toerper, M., Siddiqui, S., and Levin, S. (2015). Real-time prediction of inpatient length of stay for discharge prioritization. *Journal of the American Medical Informatics Association*, 23(e1):e2–e10.

Barua, B. and Ren, F. (2016). *The Private Cost of Pubic Queues for Medically Necessary Care, 2016 Edition*. Fraser Institute.

Bekker, R., Koole, G., and Roubos, D. (2017). Flexible bed allocations for hospital wards. *Health care management science*, 20(4):453–466.

Beliën, J. and Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2):1185–1204.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Butler, T. W., Leong, G. K., and Everett, L. N. (1996). The operations management role in hospital strategic planning. *Journal of Operations Management*, 14(2):137–156.

Campbell, H., Hotchkiss, R., Bradshaw, N., and Porteous, M. (1998). Integrated care pathways. *BMJ: British Medical Journal*, 316(7125):133.

Canadian Cancer Society's Advisory Committee on Cancer Statistics (2017). Canadian Cancer Statistics. `cancer.ca/Canadian-CancerStatistics-2017-EN.pdf`. [Accessed: 05.05.2018].

Canadian Institute for Health Information (2017). National Health ExpenditureTrends, 1975 to 2017. `https://www.cihi.ca/sites/default/files/document/nhex2017-trends-report-en.pdf`. [Accessed: 09.13.2018].

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European journal of operational research*, 201(3):921–932.

Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168. ACM.

Casanova, R., Hsu, F.-C., Sink, K. M., Rapp, S. R., Williamson, J. D., Resnick, S. M., Espeland, M. A., Initiative, A. D. N., et al. (2013). Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One*, 8(11):e77949.

CBCNews (2018). Patient angry after back surgery cancelled last minute to avoid overtime.

Chen, M., Hao, Y., Hwang, K., Wang, L., and Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 5:8869–8879.

Chen, Z., King, W., Pearcey, R., Kerba, M., and Mackillop, W. J. (2008). The relationship between waiting time for radiotherapy and clinical outcomes: a systematic review of the literature. *Radiotherapy and Oncology*, 87(1):3–16.

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., and Sun, J. (2016a). Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference*, pages 301–318.

Choi, E., Schuetz, A., Stewart, W. F., and Sun, J. (2016b). Medical concept representation learning from electronic health records and its application on heart failure prediction. *arXiv preprint arXiv:1602.03686*.

Choi, S. and Wilhelm, W. E. (2014). An approach to optimize block surgical schedules. *European Journal of Operational Research*, 235(1):138–148.

Chow, V. S., Puterman, M. L., Salehirad, N., Huang, W., and Atkins, D. (2011). Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management*, 20(3):418–430.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Dexter, F., Dexter, E. U., Masursky, D., and Nussmeier, N. A. (2008). Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. *Anesthesia & Analgesia*, 106(4):1232–1241.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., and Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161.

Drupsteen, J., van der Vaart, T., and Pieter van Donk, D. (2013). Integrative practices in hospitals and their impact on patient flow. *International Journal of Operations & Production Management*, 33(7):912–933.

Earnest, A., Chen, M. I., Ng, D., and Sin, L. Y. (2005). Using autoregressive integrated moving average (arima) models to predict and monitor the number of beds occupied during a sars outbreak in a tertiary hospital in singapore. *BMC Health Services Research*, 5(1):36.

Eijkemans, M. J., van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable a new prediction model for operating room times using individual characteristics and the surgeon's estimate. *The Journal of the American Society of Anesthesiologists*, 112(1):41–49.

Epstein, R. H. and Dexter, F. (2002). Uncertainty in knowing the operating rooms in which cases were performed has little effect on operating room allocations or efficiency. *Anesthesia & Analgesia*, 95(6):1726–1730.

Esteban, C., Staeck, O., Baier, S., Yang, Y., and Tresp, V. (2016a). Predicting clinical events by combining static and dynamic information using recurrent neural networks. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*, pages 93–101. Ieee.

Esteban, C., Staeck, O., Baier, S., Yang, Y., and Tresp, V. (2016b). Predicting clinical events by combining static and dynamic information using recurrent neural networks. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*, pages 93–101. IEEE.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115.

Ferrand, Y. B., Magazine, M. J., and Rao, U. S. (2014). Managing operating room efficiency and responsiveness for emergency and elective surgeries—a literature survey. *IIE Transactions on Healthcare Systems Engineering*, 4(1):49–64.

Fetter, R. B. and Freeman, J. L. (1986). Diagnosis related groups: product line management within hospitals. *Academy of management Review*, 11(1):41–54.

Fixler, T. and Wright, J. G. (2013). Identification and use of operating room efficiency indicators: the problem of definition. *Canadian Journal of Surgery*, 56(4):224.

Fraser Institute (2017). Waiting Your Turn Wait Times for Health Care in Canada. `https://www.fraserinstitute.org/sites/default/files/waiting-your-turn-2017.pdf`. [Accessed: 09.13.2018].

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Funkner, A. A., Yakovlev, A. N., and Kovalchuk, S. V. (2017). Data-driven modeling of clinical pathways using electronic health records. *Procedia Computer Science*, 121:835–842.

Gallivan, S., Utley, M., Treasure, T., and Valencia, O. (2002). Booked inpatient admissions and hospital capacity: mathematical modelling study. *Bmj*, 324(7332):280–282.

Gartner, D. and Kolisch, R. (2014). Scheduling the hospital-wide flow of elective patients. *European Journal of Operational Research*, 233(3):689–699.

Gartner, D., Kolisch, R., Neill, D. B., and Padman, R. (2015). Machine learning approaches for early drg classification and resource allocation. *INFORMS Journal on Computing*, 27(4):718–734.

Gaudoin, R., Montana, G., Jones, S., Aylin, P., and Bottle, A. (2015). Classifier calibration using splined empirical probabilities in clinical risk prediction. *Health care management science*, 18(2):156–165.

González-Arévalo, A., Gómez-Arnau, J., DelaCruz, F., Marzal, J., Ramírez, S., Corral, E., and García-del Valle, S. (2009). Causes for cancellation of elective surgical procedures in a spanish general hospital. *Anaesthesia*, 64(5):487–493.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Green, L. (2006). Queueing analysis in healthcare. In *Patient flow: reducing delay in healthcare delivery*, pages 281–307. Springer.

Green, L. V. (2005). Capacity planning and management in hospitals. In *Operations research and health care*, pages 15–41. Springer.

Green, L. V. (2012). Om forum—the vital role of operations analysis in improving healthcare delivery. *Manufacturing & Service Operations Management*, 14(4):488–494.

Gu, W., Fan, N., and Liao, H. (2018). Evaluating readmission rates and discharge planning by analyzing the length-of-stay of patients. *Annals of Operations Research*, pages 1–20.

Guda, H., Dawande, M., Janakiraman, G., and Jung, K. S. (2016). Optimal policy for a stochastic scheduling problem with applications to surgical scheduling. *Production and Operations Management*, 25(7):1194–1202.

Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health care management science*, 14(1):89–114.

Guinet, A. and Chaabane, S. (2003). Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. *arXiv preprint arXiv:1706.04599*.

Gupta, D. (2007). Surgical suites' operations management. *Production and Operations Management*, 16(6):689–700.

Gupta, D. and Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE transactions*, 40(9):800–819.

Hall, R. (2013). Patient flow. *AMC*, 10:12.

Hans, E., Wullink, G., Van Houdenhoven, M., and Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050.

Hans, E. W., Van Houdenhoven, M., and Hulshof, P. J. (2012). A framework for healthcare planning and control. In *Handbook of healthcare system scheduling*, pages 303–320. Springer.

Harper, P. R. (2002). A framework for operational modelling of hospital resources. *Health care management science*, 5(3):165–173.

Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making. *Health Policy*, 71(3):315–331.

Harper, P. R. and Shahani, A. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational research Society*, 53(1):11–18.

Helm, J. E., Alaeddini, A., Stauffer, J. M., Bretthauer, K. M., and Skolarus, T. A. (2016a). Reducing hospital readmissions by integrating empirical prediction with resource optimization. *Production and Operations Management*, 25(2):233–257.

Helm, J. E., Alaeddini, A., Stauffer, J. M., Bretthauer, K. M., and Skolarus, T. A. (2016b). Reducing hospital readmissions by integrating empirical prediction with resource optimization. *Production and Operations Management*, 25(2):233–257.

Helm, J. E. and Van Oyen, M. P. (2014). Design and optimization methods for elective hospital admissions. *Operations Research*, 62(6):1265–1282.

Hof, S., Fügener, A., Schoenfelder, J., and Brunner, J. O. (2017). Case mix planning in hospitals: a review and future agenda. *Health care management science*, 20(2):207–220.

Hosseini, N. and Taaffe, K. M. (2015). Allocating operating room block time using historical caseload variability. *Health care management science*, 18(4):419–430.

Hsu, W.-C., Hwang, J.-S., Chang, W.-C., Huang, S.-C., Sheu, B.-C., and Torng, P.-L. (2007). Prediction of operation time for laparoscopic myomectomy by ultrasound measurements. *Surgical endoscopy*, 21(9):1600–1606.

Hulshof, P. J., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. (2012a). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175.

Hulshof, P. J., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. (2012b). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175.

Isken, M. W. and Rajagopalan, B. (2002). Data mining to support simulation modeling of patient flow in hospitals. *Journal of medical systems*, 26(2):179–197.

Jebali, A., Alouane, A. B. H., and Ladet, P. (2006). Operating rooms scheduling. *International Journal of Production Economics*, 99(1-2):52–62.

Jebali, A. and Diabat, A. (2015). A stochastic model for operating room planning under capacity constraints. *International Journal of Production Research*, 53(24):7252–7270.

Jen, M.-H., Bottle, A., Kirkwood, G., Johnston, R., and Aylin, P. (2011). The performance of automated case-mix adjustment regression model building methods in a health outcome prediction setting. *Health care management science*, 14(3):267–278.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., and Kripalani, S. (2011). Risk prediction models for hospital readmission: a systematic review. *Jama*, 306(15):1688–1698.

Karatzoglou, A., Meyer, D., and Hornik, K. (2005). Support vector machines in r.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab-an s4 package for kernel methods in r. *Journal of statistical software*, 11(9):1–20.

Kayis, E., Wang, H., Patel, M., Gonzalez, T., Jain, S., Ramamurthi, R., Santos, C., Singhal, S., Suermondt, J., and Sylvester, K. (2012). Improving prediction of surgery duration using operational and temporal factors. In *AMIA Annual Symposium Proceedings*, volume 2012, page 456. American Medical Informatics Association.

Kheirkhah, P., Feng, Q., Travis, L. M., Tavakoli-Tabasi, S., and Sharafkhaneh, A. (2015). Prevalence, predictors and economic consequences of no-shows. *BMC health services research*, 16(1):13.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction policy problems. *The American economic review*, 105(5):491–495.

Kong, G., Xu, D.-L., Body, R., Yang, J.-B., Mackway-Jones, K., and Carley, S. (2012). A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research*, 219(3):564–573.

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

Krämer, J., Schreyögg, J., and Busse, R. (2017). Classification of hospital admissions into emergency and elective care: a machine learning approach. *Health care management science*, pages 1–21.

Kulkarni, P., Smith, L. D., and Woeltje, K. F. (2016). Assessing risk of hospital readmissions for improving medical practice. *Health care management science*, 19(3):291–299.

Lamiri, M., Xie, X., Dolgui, A., and Grimaud, F. (2008a). A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185(3):1026–1037.

Lamiri, M., Xie, X., Dolgui, A., and Grimaud, F. (2008b). A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185(3):1026–1037.

Leeftink, A., Bikker, I., Vliegen, I., and Boucherie, R. (2018). Multi-disciplinary planning in health care: a review. *Health Systems*, pages 1–24.

Li, S.-T., Chiu, N.-C., Kung, W.-C., and Chen, J.-C. (2013). Factors affecting length of stay in the pediatric emergency department. *Pediatrics & Neonatology*, 54(3):179–187.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, 2(3):18–22.

Littig, S. J. and Isken, M. W. (2007). Short term hospital occupancy prediction. *Health care management science*, 10(1):47–66.

Liu, N. and Ziya, S. (2014). Panel size and overbooking decisions for appointment-based services under patient no-shows. *Production and Operations Management*, 23(12):2209–2223.

Liu, N., Ziya, S., and Kulkarni, V. G. (2010). Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management*, 12(2):347–364.

Ma, G. and Demeulemeester, E. (2013). A multilevel integrative approach to hospital case mix and capacity planning. *Computers & Operations Research*, 40(9):2198–2207.

Mallor, F. and Azcárate, C. (2014). Combining optimization with simulation to obtain credible models for intensive care units. *Annals of Operations Research*, 221(1):255–271.

Marcon, E. and Dexter, F. (2006). Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, 9(1):87–98.

Marshall, A. and McClean, S. (2003). Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research*, 10(6):565–576.

Marshall, A., Vasilakis, C., and El-Darzi, E. (2005). Length of stay-based patient flow models: recent developments and future directions. *Health Care Management Science*, 8(3):213–220.

Marynissen, J. and Demeulemeester, E. (2016). Literature review on integrated hospital scheduling problems.

May, J. H., Spangler, W. E., Strum, D. P., and Vargas, L. G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20(3):392–405.

May, J. H., Strum, D. P., and Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, 31(1):129–148.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McIntosh, C., Dexter, F., and Epstein, R. H. (2006). The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an australian hospital. *Anesthesia & Analgesia*, 103(6):1499–1516.

Menon, A. K., Jiang, X. J., Vembu, S., Elkan, C., and Ohno-Machado, L. (2012). Predicting accurate probabilities with a ranking loss. In *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, volume 2012, page 703. NIH Public Access.

Meskens, N., Duvivier, D., and Hanset, A. (2013). Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems*, 55(2):650–659.

Moore, C. G., Wilson-Witherspoon, P., and Probst, J. C. (2001). Time and money: effects of no-shows at a family practice residency clinic. *Family Medicine-Kansas City-*, 33(7):522–527.

Naeini, M. P., Cooper, G. F., and Hauskrecht, M. (2015). Binary classifier calibration using a bayesian non-parametric approach. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 208–216. SIAM.

Niculescu-Mizil, A. and Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 625–632. ACM.

Norris, J. B., Kumar, C., Chand, S., Moskowitz, H., Shade, S. A., and Willis, D. R. (2014). An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. *Decision Support Systems*, 57:428–443.

Olivares, M., Terwiesch, C., and Cassorla, L. (2008). Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science*, 54(1):41–55.

Ozcan, Y. A., Tànfani, E., and Testi, A. (2017). Improving the performance of surgery-based clinical pathways: a simulation-optimization approach. *Health care management science*, 20(1):1–15.

Oztekin, A., Al-Ebbini, L., Sevkli, Z., and Delen, D. (2018). A decision analytic approach to predicting quality of life for lung transplant recipients: A hybrid genetic algorithms-based methodology. *European Journal of Operational Research*, 266(2):639–651.

Pandit, J. and Carey, A. (2006). Estimating the duration of common elective operations: implications for operating list management. *Anaesthesia*, 61(8):768–776.

Parizi, M. S. and Ghate, A. (2016). Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking. *Computers & Operations Research*, 67:90–101.

Pham, T., Tran, T., Phung, D., and Venkatesh, S. (2017a). Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of biomedical informatics*, 69:218–229.

Pham, T., Tran, T., Phung, D., and Venkatesh, S. (2017b). Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of biomedical informatics*, 69:218–229.

Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74.

Public Health Agency of Canada (2013). Preventing Chronic Disease Strategic Plan 2013-2016. `http://publications.gc.ca/collections/collection_2014/aspc-phac/HP35-39-2013-eng.pdf`. [Accessed: 05.05.2018].

Public Health Agency of Canada (2015). Improving Health Outcomes: A Paradigm Shift. `http://publications.gc.ca/collections/collection_2016/aspc-phac/HP35-39-2015-eng.pdf`. [Accessed: 05.05.2018].

Rais, A. and Viana, A. (2011). Operations research in healthcare: a survey. *International transactions in operational research*, 18(1):1–31.

Ranjan, C., Paynabar, K., Helm, J. E., and Pan, J. (2017). The impact of estimation: A new method for clustering and trajectory estimation in patient flow modeling. *Production and Operations Management*, 26(10):1893–1914.

Roth, A. V. and Van Dierdonck, R. (1995). Hospital resource planning: concepts, feasibility, and framework. *Production and operations management*, 4(1):2–29.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533.

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., and Rademakers, F. E. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling*, 19(5):493–525.

Shukla, R. K., Ketcham, J. S., and Ozcan, Y. A. (1990). Comparison of subjective versus data base approaches for improving efficiency of operating room scheduling. *Health Services Management Research*, 3(2):74–81.

Shulan, M., Gao, K., and Moore, C. D. (2013). Predicting 30-day all-cause hospital readmissions. *Health care management science*, 16(2):167–175.

Shylo, O. V., Prokopyev, O. A., and Schaefer, A. J. (2012). Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing*, 25(4):682–692.

Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G., Barzi, A., and Jemal, A. (2017). Colorectal cancer statistics, 2017. *CA: a cancer journal for clinicians*, 67(3):177–193.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.

Simchi-Levi, D. (2013). Om forum—om research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management*, 16(1):2–10.

Spangler, W. E., Strum, D. P., Vargas, L. G., and May, J. H. (2004). Estimating procedure times for surgeries by determining location parameters for the lognormal model. *Health care management science*, 7(2):97–104.

Stepaniak, P. S., Heij, C., Mannaerts, G. H., de Quelerij, M., and de Vries, G. (2009). Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesthesia & Analgesia*, 109(4):1232–1245.

Strum, D. P., Sampson, A. R., May, J. H., and Vargas, L. G. (2000). Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 92(5):1454–1466.

Suykens, J. A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300.

Tyler, D. C., Pasquariello, C. A., and Chen, C.-H. (2003). Determining optimum operating room utilization. *Anesthesia & Analgesia*, 96(4):1114–1121.

van Veen-Berkx, E., Elkhuizen, S. G., Kuijper, B., Kazemier, G., and Collaborative, D. O. R. B. (2016a). Dedicated operating room for emergency surgery generates more utilization, less overtime, and less cancellations. *The American Journal of Surgery*, 211(1):122–128.

van Veen-Berkx, E., Elkhuizen, S. G., Kuijper, B., Kazemier, G., and Collaborative, D. O. R. B. (2016b). Dedicated operating room for emergency surgery generates more utilization, less overtime, and less cancellations. *The American Journal of Surgery*, 211(1):122–128.

VanBerkel, P. T. and Blake, J. T. (2007). A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health care management Science*, 10(4):373–385.

Venkataraman, S., Fredendall, L. D., Taaffe, K. M., Huynh, N., and Ritchie, G. (2018). An empirical examination of surgeon experience, surgeon rating, and costs in perioperative services. *Journal of Operations Management*.

Vissers, J. M., Bertrand, J., and De Vries, G. (2001). A framework for production control in health care organizations. *Production Planning & Control*, 12(6):591–604.

Wachtel, R. E. and Dexter, F. (2009). Influence of the operating room schedule on tardiness from scheduled start times. *Anesthesia & Analgesia*, 108(6):1889–1901.

Wang, H., Zheng, B., Yoon, S. W., and Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2):687–699.

Warlick, C., Trock, B. J., Landis, P., Epstein, J. I., and Carter, H. B. (2006). Delayed versus immediate surgical intervention and prostate cancer outcome. *Journal of the National Cancer Institute*, 98(5):355–357.

Weiss, E. N., Cohen, M. A., and Hershey, J. C. (1982). An iterative estimation and validation procedure for specification of semi-markov models with application to hospital patient flow. *Operations Research*, 30(6):1082–1104.

White, D. L., Froehle, C. M., and Klassen, K. J. (2011). The effect of integrated scheduling and capacity policies on clinical efficiency. *Production and Operations Management*, 20(3):442–455.

Wullink, G., Van Houdenhoven, M., Hans, E. W., van Oostrum, J. M., van der Lans, M., and Kazemier, G. (2007). Closing emergency operating rooms improves efficiency. *Journal of Medical Systems*, 31(6):543–546.

Zacharias, C. and Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, 23(5):788–801.

Zadrozny, B. and Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 694–699. ACM.

Zeltyn, S., Marmor, Y. N., Mandelbaum, A., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Shtub, A., Lauterman, T., et al. (2011). Simulation-based models of emergency departments:: Operational, tactical, and strategic staffing. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 21(4):24.

Zhang, B., Murali, P., Dessouky, M., and Belson, D. (2009). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, 60(5):663–673.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598.

Zhou, Z., Miller, D., Master, N., Scheinker, D., Bambos, N., and Glynn, P. (2016). Detecting inaccurate predictions of pediatric surgical durations. In *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*, pages 452–457. IEEE.

Zijm, W. H. (2000). Towards intelligent manufacturing planning and control systems. *OR-Spektrum*, 22(3):313–345.

APPENDIX A COMPLEMENTARY INFORMATION ON COLORECTAL CANCER DIAGNOSIS AND PROCEDURE CODES

Table A.1 Colorectal Cancer CD-O-3 SITE CODES

ICD-O-3 SITE CODES	Diagnosis
C21.1	*Anal*
C18.9	*Colon*
C19.9	*Rectosigmoid*
C20.9	*Rectal*

TNM staging uses three factors for categorizing the stage of cancer:

- T: the size of the primary tumor
- Regional lymph nodes (N): the number and location of any regional lymph nodes that have cancer cells in them
- Distant Metastasis (M) whether the cancer has spread or metastasized to another part of the body

Table A.2 CRC TNM Staging-Primary Tumor(T)

Primary Tumor (T)	Description
TX	*Primary tumor cannot be assessed because of lack of information*
T0	*No evidence of tumor*
Tis	*Carcinoma in situ – tumor is confined to the mucosa of the colon or rectum*
T1	*Tumor extends into submucosa*
T2	*Tumor invades the muscularis propria*
T3	*Tumor invades the subserosa*
T4	*Tumor grows beyond the wall of colon or rectum and invades nearby* • *T4a – tumor perforates visceral peritoneum* • *T4b – tumor directly invades other organs or structures, including other segments of the colon or rectum by way of the serosa*

Table A.3 CRC TNM Staging-Regional lymph nodes(N)

Regional lymph nodes (N)	Description
NX	*Regional lymph nodes cannot be assessed*
N0	*No regional lymph node metastasis*
N1	*Metastasis in 1-3 regional lymph nodes* • *N1a: metastasis in 1 regional lymph node* • *N1b: metastasis in 2-3 regional lymph nodes* • *N1c: nests of cancer cells (satellites) in the lymph drainage areas of the subserosa or the surrounding tissue of the colon or rectum without regional lymph node metastasis*
N2	*Metastasis in 4 or more regional lymph nodes* • *N2a: metastasis in 4-6 regional lymph nodes* • *N2b: metastasis in 7 or more regional lymph nodes*

Table A.4 Colon Cancer Site-Specific Surgery Codes

Surgery Code	Surgery Description
C18.9.20	*Local tumor excision*
C18.9.30	*Partial colectomy, segmental resection*
C18.9.40	*Subtotal colectomy/hemicolectomy*
C18.9.50	*Total colectomy*
C18.9.60	*Total proctocolectomy*
C18.9.70	*Colectomy or coloproctotectomy with resection of contiguous organ*

Table A.5 Rectosigmoid Cancer Site-Specific Surgery Codes

Surgery Code	Surgery Description
C19.9.20	*Local tumor excision*
C19.9.30	*Wedge or segmental resection; partial proctosigmoidectomy*
C19.9.40	*Pull through WITH sphincter preservation (colo-anal anastomosis)*
C19.9.50	*Total proctectomy*
C19.9.60	*Total proctocolectomy*
C19.9.70	*Colectomy or proctocolectomy resection in continuity with other organs; pelvic exenteration*

Table A.6 Rectum Cancer Site-Specific Surgery Codes

Surgery Code	Surgery Description
C20.9.20	*Local tumor excision*
C20.9.30	*Wedge or segmental resection; partial proctectomy*
C20.9.40	*Pull through WITH sphincter preservation (coloanal anastomosis)*
C20.9.50	*Total proctectomy*
C20.9.60	*Total proctocolectomy*
C20.9.70	*Proctectomy or proctocolectomy with resection in continuity with other organs; pelvic exenteration*

Table A.7 Rectum Cancer Site-Specific Surgery Codes

Surgery Code	Surgery Description
C20.9.20	*Local tumor excision*
C20.9.30	*Wedge or segmental resection; partial proctectomy*
C20.9.40	*Pull through WITH sphincter preservation (coloanal anastomosis)*
C20.9.50	*Total proctectomy*
C20.9.60	*Total proctocolectomy*
C20.9.70	*Proctectomy or proctocolectomy with resection in continuity with other organs; pelvic exenteration*

Table A.8 – Conversion of Surgical procedures to their respective codes

Surgery Description	Surgery Code
Anus and anal canal-Polypectomy of the anus	C18.9.20
Anus and anal canal -Excision of anus tumor	C18.9.20
Anus and anal canal-Endo-anal resection	C18.9.20
Appendix-Polypectomy of the colon	C18.9.20
Appendix-Appendectomy	C18.9.20
Appendix-Right Hemicolectomy	C18.9.40
Appendix-colectomy	C18.9.40
Colon-Right Hemicolectomy	C18.9.40
Colon-Anterior Resection	C18.9.20
Colon-Left Hemicolectomy	C18.9.40
Colon-Total Colectomy + Ileostomy	C18.9.50
Colon-hémicolectomie	C18.9.40
Colon-Left Hemicolectomy enlarged	C18.9.40
Colon-Partial colectomy + colostomy	C18.9.40
Colon-Polypectomy of the colon	C18.9.20
Colon-bilateral oophorectomy	C18.9.70
Colon-excision of a bile duct + partial hepatectomy	C18.9.70
Colon-Excisional Biopsy of Retoperitoneum and Peritoneum Tumor	C18.9.20
Colon-Ileum-Caecal Resection	C18.9.20
Colon-Excisional biopsy of a ganglion	C18.9.20
Colon-Polypectomy of the stomach	C18.9.20
Colon-colectomy	C18.9.40
Colon-Total Colectomy	C18.9.50
Colon-subtotal colectomy	C18.9.40
Colon-Right Hemicolectomy enlarged	C18.9.40
Colon-Partial Hepatectomy	Hepatectomy
Colon-Total Colectomy + Iterectal reconstruction	C18.9.40
Colon-hepatectomy	Hepatectomy
Colon-Excision of Retroperitoneum and Peritoneum Tumor	C18.9.20
Colon-Polypectomy of the rectum	C20.9.20
Colon-Colon Surgery	C18.9.20
Rectosigmoid-anterior resection	C19.9.20

Table A.8 – Conversion of Surgical procedures to their respective codes

Surgery Description	Surgery Code
Rectosigmoid-Polypectomy of the rectosigmoid	C19.9.20
Rectosigmoid-Polypectomy of the rectum	C20.9.20
Rectosigmoid-Tumor excision of rectosigmoid	C19.9.20
Rectosigmoid-Polypectomy of the colon	C19.9.20
Rectosigmoid-Colon Surgery	C18.9.20
Rectosigmoid-Excisional biopsy of retroperitoneum and peritoneal tumor	C19.9.20
Rectum-Polypectomy of the rectum	C20.9.20
Rectum-anterior resection	C20.9.20
Rectum-Polypectomy of the colon	C20.9.20
Rectum-excision of rectum tumor	C20.9.20
Rectum-Left hemicolectomy	C18.9.40
Rectum-colectomy	C20.9.30
Hysterectomy-rectum	C20.9.70
Rectum-Abdominoperineal Resection (APR)	C20.9.70
Rectum-hepatectomy	Hepatectomy
Rectum-Total proctocolectomy + ileostomy	C20.9.60
Rectum-Cysteduction of the digestive system	C20.9.30
Rectum-subtotal colectomy	C20.9.30

Table A.9 R publicly available packages used in the thesis

	Package	Description
Chapter 4	MXNetR	The R interface to the MXNet deep learning library
	deepr v 0.1	Dirichlet-multinomial Evolutionary Event Profile Randomization (DEEPR) test
Chapter 5	rpart	Recursive partitioning for classification, regression and survival trees
	randomForest	Breiman and Cutler's Random Forests for Classification and Regression
	e1071	Support vector machine
	fitdistrplus	Help to Fit of a Parametric Distribution to Non-Censored or Censored Data
	logspline	Routines for Logspline Density Estimation
	CDFt	tatistical downscaling through CDF-transform
	MASS	Support Functions and Datasets for Venables and Ripley's MASS
	kernlab	Kernel-Based Machine Learning Lab
Chapter 6	randomForest	Breiman and Cutler's Random Forests for Classification and Regression
	kernlab	Kernel-Based Machine Learning Lab
	MXNetR	The R interface to the MXNet deep learning library
	twostageTE v1.3	Pool adjacent violators algorithm