Predicting waiting times in Radiation Oncology using machine learning

Ackeem Joseph, B.Sc.

Master of Science

Medical Physics Unit

McGill University Montreal, Quebec October 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Medical Radiation Physics

 \bigodot Ackeem Joseph 2016

ABSTRACT

This thesis describes a study of machine learning and its application to waiting times in radiation oncology. Specifically, an evaluation of waiting time estimates for daily radiation treatment appointments at the McGill University Health Centre was conducted and two unique communication tools for conveying these waiting time estimates were developed. To evaluate waiting time estimates, a subset of previously-treated patient records was used for training and modelling by several off-the-shelf machine learning algorithms and tested using treatment records unseen by the algorithms. Furthermore, machine learning was explicitly used to predict treatment durations in order to calculate an overall waiting time estimate. It was found that the random forest regression algorithm provided the best model for daily radiation treatment durations. On average, the model is able to predict treatment durations to within 7.2 minutes. This is a significant improvement on the rough estimates currently given to patients, and also on the estimates one would obtain by conducting simple averages on all patients. One of the tools developed to communicate waiting time estimates to patients made use of the best-trained model by simulating the delivery of waiting time estimates to all patients in real-time. Waiting time estimates from this study have not yet been delivered to patients. However, a separate and ongoing research is discussed in this dissertation, examining the delivery of personal health information, including waiting time estimates, to patients via a mobile phone application.

i

RÉSUMÉ

Ce mémoire décrit une étude de l'apprentissage machine et son application aux temps d'attente en radio-oncologie. Plus précisément, le temps d'attente des rendez-vous quotidiens des traitements externes au Centre universitaire de santé McGill ont été estimé et deux outils de communication uniques pour transmettre ces estimations ont été élaborés. Un sous-ensemble de patients traités précédemment a été utilisé pour la formation et la modélisation à l'aide de plusieurs algorithmes d'apprentissage machine. Ceux-ci ont été testés en utilisant les données exclues par la modélisation. De plus, l'apprentissage machine a été utilisé explicitement pour prévoir les durées de traitement afin de calculer une estimation globale du temps d'attente. L'algorithme de random forest regression fournissait le meilleur modèle pour les durées quotidiennes de traîtement par irradiation. En moyenne, le modèle est capable de prédire la durée des traitements à l'intérieur de 7,2 minutes. Il s'agit d'une amélioration considérable par rapport à l'éstimation du temps d'attente actuellement transmis aux patients et aussi par rapport aux estimations que l'on obtiendrait en effectuant des moyennes simples. Un des outils pour communiquer le temps d'attente simule la livraison des données à tous les patients en temps réel en utilisant le meilleur modèle d'apprentissage machine. À ce jour, les estimations de temps d'attente de cette étude n'ont pas encore été fournies aux patients. Cependant, une recherche distincte examinant la distribution de renseignements personnels sur la santé, includant les estimations de temps d'attente via une application mobile, est présentement en cours. Ceci est discutée dans cette dissertation.

ACKNOWLEDGEMENTS

This dissertation would have not been possible without the help of so many people in so many ways. I wish to express my sincere gratitude to my supervisor, Dr. John Kildea, for his patience, support, and tenacious encouragement throughout this endeavour. I am extremely grateful for the time and effort he spent with me on this work and for the helpful advice and guidance he provided. I would also like to thank the co-leads of the Health Informatics Group, Dr. Tarek Hijal, Dr. John Kildea and Prof. Laurie Hendren, for getting together and creating this amazing research group and for the wonderful opportunity to work on this research project. Of course, my appreciation also extends to the students of the Health Informatics Group for their comprehensive and thoughtful feedback, perspective, and collaboration.

I would like to express my deepest appreciation for the support provided by the McGill Medical Physics Unit, including fellow teachers, fellow classmates, colleagues, and faculty administrators. I would like to thank the McGill University Writing Centre for their guidance and writing expertise throughout the course of my dissertation.

I take this opportunity to thank all members of the department of radiation oncology at the McGill University Health Centre for their help, encouragement, cooperation and collaboration; as well as those who, directly or indirectly, have lent their helping hand in this venture.

Finally, and most importantly, I thank my family and friends for their conscientiousness and endless support. I give my heartfelt appreciation and unending gratitude to each and every one of them for their love and encouragement throughout this process.

I am extremely proud and grateful to be where I am now.

TABLE OF CONTENTS

ABSTE	RACT i				
RÉSUMÉ ii					
ACKN	ACKNOWLEDGEMENTS iii				
LIST C	OF FIGURES				
LIST C	OF TABLES				
1 In	troduction				
1. 1. 1. 1.	1Overview of Thesis12Radiation Therapy13Cancer21.3.1Cancer Outlook21.3.2Cancer Treatment34Radiation Therapy In Practice31.4.1The Radiation Oncology Team45Medical Physics71.5.1The Role of a Medical Physicist7Clinical Service7Research and Development81.5.2Informatics—The Future of Medical Physics9				
2 W	Vaiting Time Uncertainty — A Universal Challenge 10				
2. 2. 2. 2. 2.	1The Effects of Waiting in Healthcare112Waiting For Care in Radiation Oncology112.2.1The Patient Pathway122.2.2Waiting Time Occurrences153Waiting Time Studies154The Significance of Electronic Health Records17				
3 M	Tachine Learning—The Personalized Solution				
3. 3. 3.	1Overview192Learning From Data203.2.1Types of Learning223Learning Algorithms23				

		3.3.1 List of Common Machine Learning Algorithms	23
		Linear Regression	24
		Logistic Regression	24
		Decision Tree	25
		Support Vector Machine	28
		K-Means	31
		Random Forest	31
		3.3.2 Performance Evaluation	33
		Mean Absolute Error	35
		Median Absolute Error	35
		Mean Squared Error	35
		Coefficient of Determination	35
		3.3.3 Sampling	35
	2.4	3.3.4 Feature Selection	37
	3.4	Modelling Waiting Times in Radiation Oncology	39
		3.4.1 Steps to Applying Machine Learning in Clinical	00
		Practice	39
4	Desig	gn and Implementation of a Clinical Machine Learning Solution	41
	4.1	Data Background	41
	4.2	Data Collection	42
		$4.2.1$ Aliasing \ldots	44
	4.3	Waiting Time Calculation	45
	4.4	Learning Process	46
		4.4.1 Data Extraction	47
		4.4.2 Preprocessing	49
		4.4.3 Machine Learning Models	50
		4.4.4 Feature Selection	51
	4.5	Communicating Waiting Time Predictions	52
		4.5.1 Opal — The Oncology Portal and Application	52
		4.5.2 Waiting Time Delivery Simulator	52
5	Resul	lts–Evaluation of Waiting Time Estimates	60
	5 1	Production Models	60
	5.1	Feature Trends	61
	0.2	5.2.1 Imaging-Phase Study	62
	53	Comparison Against Simple Averages	63
	$5.0 \\ 5.4$	Results from Waiting Time Simulator	68
	5.5	Prediction Provision	70
	0.0	5.5.1 Opal	70
6	Discu	ission	72
	0.1		
	6.1	Model Performance	72
	0.2	reature Set	(4

	6.3	Real-Time Delivery of Waiting Time Estimates	75
7	Conclu	sions and Future Work	78

LIST OF FIGURES

Figure		page
1.1	The radiation oncology team is a patient-centred group that consists of various radiation medicine personnel fulfilling the goal of diagnosing and treating a patient with cancer. Prox- imity to the patient in this model represents the closeness of involvement with the patient in practice – Direct (inner ring) and indirect involvement (outer ring). Table 1.1 provides descriptions of radiation oncology team members with their associate responsibilities	5
2.1	An overview of the patient's radiotherapy pathway with colour- coded waiting periods. In radiation oncology, patients expe- rience three types of waiting times. One type of wait occurs when the patient is waiting at home for their treatment plan to be prepared by staff after receiving a CT-scan (purple). Another waiting period occurs in the waiting room for daily fractionated radiotherapy (blue). Similarly, another wait- ing period occurs in the waiting room to see a physician in consultation, intra-treatment examination, or follow-up (green).	16
3.1	An overview of the machine learning setup. Adapted from (Abu-Mostafa et al., 2012)	21
3.2	Machine learning according to the nature of the input train- ing data. In unsupervised learning, the outputs are not defined (i.e. unlabelled). In semi-supervised learning, some of the outputs are defined (i.e. semi-labelled). In supervised learning, all the outputs are defined (i.e. labelled)	22
3.3	An example linear regression model. In this example, a person's weight is a function of one feature; the person's height. With a regression line defined (red), a new person's weight can be predicted from their height. Adapted from (Buitinck et al., 2013).	25

3.4	An example logistic regression model. In this classification example, there are two outcomes – either a person passes an exam or fails. The two outcomes are a function of one feature; the amount of hours spent studying. A logit function (red) models the scenario as a prediction probability of achieving either outcome. For example, if a person studies for three hours, they are around 80% likely to pass. Adapted from (Buitinck et al., 2013).	26
3.5	A decision tree example. In this classification example, the objective is to predict whether a person is male or female. According to the decision nodes, the first split is based on the most significant feature; a person's height. If the person is shorter than 170 cm, then a second decision is made based on the second most significant feature; their weight. This is a crude example of course, but essentially new outcomes are predicted by traversing the tree model and evaluating the specific input features at each node (blue) until an end leaf node (green) is reached.	27
3.6	A support vector machine classification example. In this example, there are two features and two classifications. Each input feature represents one coordinate and each target class is labelled as colours (red and blue). The objective is to predict a class from new data. A SVM model determines the best split on the training data such that the closest training point(s) in each class are furthest away from the line. These closest points are known as <i>support vectors</i> and are defined as margins (dotted lines) in which separation occurs (Buitinck et al., 2013).	29
3.7	A support vector machine regression example. In this example, the target output is plotted as a function of one feature. An SVM model fits the training data to within some margin defined by an epsilon (ϵ) parameter (dotted band). The optimization process ignores training points that are located within the band and attempts to minimize the errors on the training points outside the band. Adapted from (Buitinck et al., 2013)	30

32
32
34
38
39
43

4.2	A flowchart of the AEHRA architecture. Select data from Aria (and potentially other hospital EMRs) are copied across to our centralized database, AEHRA_db. The times at which transfers occur are determined by the cron job scheduler and set using the Cron Control Interface. The exact metadata that are copied is set using the Alias Interface, in which internal clinical codes are translated and grouped into standard taxonomies. AEHRA is a standalone software in which many types of studies can be created (using the Study Interface) for multi-purpose data analysis	44
4.3	Defining waiting in the context of daily radiotherapy appoint- ments. In this example, patient 6 arrives just as patient 2's treatment appointment has finished. Both patient 1's and patient 2's appointments have incurred a combined 10- minute delay at the time of patient 6's arrival. Machine learning is performed to calculate the treatment durations of patients 3, 4, and 5, generating a cumulative duration that may or may not run over into patient 6's appointment. In this example, it has. The total waiting time for patient 6 is thus the time of arrival to the estimated start time inferred from the cumulative preceding patient durations	45
4.4	Preprocessing the dataset using sklearn for learning required constructing a training input matrix of N samples and M features and a separate matrix of N target values	49
4.5	Example of one-hot encoding. In this example, there are four distinct values for one feature of "animals." Therefore, four binary variables are used where a value of "1" is assigned to the appropriate value and zero otherwise. Note that in this example, the feature space increases by a factor of four	50
4.6	Sklearn's "cheat sheet" for choosing the right algorithm(s). This study involved predicting daily treatment appointment durations. In other words, the outputs were real-valued. Thus, various supervised regression models (light blue section) were used in this study. Image taken from http://scikit-learn.org/.	51

xi

- An overview of Opal's communication architecture. Select data 4.7 from Aria and other hospital EMRs are transferred to a centralized database known as *OpalDB*. Data transfer occurs using a "cron job" Auto Update scheduler, that copies across approved appointments, tasks, documents, and lab test results. Approval and timing of data transfer are determined by a list of "publishing rules" that are set using a *Publish* Manager web interface. Secure serving of data through the hospital's firewall to Opal is facilitated by *Firebase*, a cloud database operated by Google Inc. (firebase.com). Firebase it designed such that all applications connected to it are served data in real time. Data is encrypted once before sending to Firebase and encrypted again by Firebase's own encryption protocols. Data ultimately reaches out to the patient in the outside world on their mobile phone or web browser. A demo of Opal can be found online at depdocs.com/opal....
- A magnified and detailed version of figure 4.3 describing the 4.8 waiting time delivery simulation process. In this example, patient 3 has started treatment. Patients 4, 5, and 6 have vet to be treated on the same treatment machine. For each pending patient, the simulator prints out a live time block estimate (light blue) based on the sum of predicted durations (yellow) of preceding patients, along with an indication on the amount of patients in the queue and the status of clinical overruns by the time their respective treatments start. For instance, patient 4 is after patient 3. The waiting time delivery simulator prints a waiting time estimate range of "10-20 mins" since the actual wait is $15 mins \pm error$. Patient 5 is second in the queue. Therefore, the simulator prints a waiting time estimate range of "30-45 mins" since the actual wait is $(15+15) = 30 \ mins \pm error$, etc.

53

4.9	A screenshot of a timeline log for the number of scheduled (black) and actual (blue) patients in the queue for a selected patient. The beginning of the timeline log represents the time the patient checked-in. In this example, when the pa- tient checked-in there was one patient already in the queue. Over time, the scheduled patients arrived, checked-in, and the number of actual patients in the queue grew to match the scheduled number of patients. As the preceding patients underwent treatment, the number of pending patients de- creased. The end of the timeline log represents the time the patient was called in for treatment. In this screenshot, the patient was called in before a preceding patient finished treatment. The vertical line represents the scheduled start time for this particular patient	56
4.1	0 A screenshot of an estimated start timeline log for a selected patient. The beginning of the timeline log represents the time the patient checked-in and the end of the timeline log represents the time the patient was called for treatment. Red represents behind schedule and green represents ahead of schedule. In this screenshot, preceding patients were taking longer than their allocated scheduled time. Thus, the patient's estimated start was pushed over time. The information displayed in the blue box (upper right) represents details about the data point at the end of the timeline. When the patient was called in for treatment, a waiting time estimate of "< 10 minutes" was logged	57
4.1	1 A screenshot of the clinical schedule timeline log for a selected patient. The beginning of the timeline log represents the time the patient checked-in and the end of the timeline log represents the time the patient was called for treatment. Red represents behind schedule and green represents ahead of schedule. Over time, preceding patients were taking longer than their allocated scheduled time. Thus, by the time the patient was called in for treatment, the clinic was 25 minutes behind schedule	58
5.1	A residual plot histogram for the performance of the random forest regression model on the testing set.	62
5.2	Plot of average treatment duration versus diagnosis for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion)	63

5.3	Plot of average treatment duration versus treatment appoint- ment day of the week for all retrospective data ($1 = $ Sunday, 7 = Saturday). Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion)	64
5.4	Plot of average treatment duration versus treatment appoint- ment hour for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to single value data points	65
5.5	Plot of average treatment duration versus fraction number for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to a single value data points	65
5.6	Plot of average treatment duration versus allocated treatment appointment time for all retrospective data. Error bars rep- resent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to a single value data points	66
5.7	Plot of treatment duration versus median past duration for all retrospective data. The red line represents a linear regression of all data points.	67
5.8	Histograms of imaging occurrence as a function of fraction number per diagnosis (pre-2015). The y-axis represents the frequency of imaging and the x-axis represents the fraction number. Note that every patient has a first fraction (i.e. treated for the first time) and imaging is always done on the first fraction. The frequency at which imaging occurs in later fractions decreases due to the decreasing number of patients that are treated at higher fractions	67
5.9	Histograms of imaging occurrence as a function of fraction number per diagnosis (post-2015). The y-axis represents the frequency of imaging and the x-axis represents the fraction number. Note that every patient has a first fraction (i.e. treated for the first time) and imaging is always done on the first fraction. The frequency at which imaging occurs in later fractions decreases due to the decreasing number of patients that are treated at higher fractions	68
5.10	Pie chart showing the estimate details that were delivered to checked-in patients immediately prior to treatment	70

LIST OF TABLES

Table		page
1.1	Radiation oncology team members and a brief description of their responsibilities	6
3.1	List of common machine learning algorithms along with the model type (classification or regression) and learning method (supervised or unsupervised)	24
4.1	List of features extracted for each input sample set. Features were either categorical as string variables or real valued (continuous)	47
5.1	Performance evaluation of each machine learning regression model used for this study. The standard deviation error is the standard deviation of the calculated residuals	61
5.2	Top five features as defined by sklearn's <i>feature_importance</i> method using the random forest regression model. Other proposed features corresponded to a total remainder importance of 8%.	61
5.3	Evaluation and comparison of the random forest regression (RFR) model against simple averages. For each predictor listed here, the average of each distinct category was used to predict treatment appointment durations. The standard deviation error is the standard deviation of the residual (predicted minus actual).	69
5.4	Performance evaluation of the random forest regression model on prospective data. Performance metrics were collected on a daily basis, between June 2016 and October 2016. The average and standard deviation were taken for each metric.	69
	0	

CHAPTER 1 Introduction

1.1 Overview of Thesis

This thesis builds upon the subject of machine learning in radiation oncology by adding a new application area – waiting times. This research presents an effort to describe and evaluate waiting time estimates, and to communicate these estimates to patients in radiation oncology at the McGill University Health Centre (MUHC) in Montreal, Canada.

This chapter introduces the study of radiation therapy and cancer treatment, and clinical services in radiation oncology, including the field of medical physics and informatics. Chapter 2 describes the effects of waiting in healthcare and outlines waiting in the context of radiation oncology. Chapter 3 details machine learning and its application to waiting times in clinical practice. Chapter 4 focusses on the rigorous methodology undertaken in this project. Chapters 5 and 6 present the results and discussion, respectively. The conclusions and future work for this research are described in chapter 7.

1.2 Radiation Therapy

Radiation therapy is a branch of medicine that is concerned with the use of ionizing radiation to treat cancer and some benign diseases. Ionizing radiation damages cells by destroying the genetic material (DNA) that controls how cells grow and divide. While both healthy and cancerous cells are damaged by radiation, the goal of radiation therapy is to deliver the maximum amount of radiation to the cancerous cells, while minimizing the amount of radiation to the surrounding healthy cells (Washington and Leaver, 2015). To achieve this goal, radiation therapy relies heavily on knowledge gained from the physical and biological sciences, as well as computer science and engineering. For many patients with cancer, radiation therapy plays a vital role in their overall treatment.

1.3 Cancer

In normal tissue cell growth, there is a regulated balance between cellular division and self-limitation such that proliferation only occurs when necessary (Macdonald et al., 2004). As a cell grows, it undergoes mitosis and divides into daughter cells. The cell division process repeats in an ordered manner until a mature cell with a specific function results. In tumours, however, cells may continue to divide and grow with no resulting mature cell to halt the process, causing abnormal cellular proliferation (Washington and Leaver, 2015).

Cancer is a class of many diseases in which cells undergo mutation towards uncontrolled cell division, resulting in intrusions on surrounding tissues or sometimes spread to different areas of the body to generate further growth (metastasis). Most cancers form neoplastic tumours (layers of cancer cells), but some, such as leukaemia, do not. Benign tumours are self-limiting tumours that do not metastasize and are rarely life threatening, unless vital structures, such as blood vessels or nerves, are constricted.

1.3.1 Cancer Outlook

The Canadian Cancer Society (2015) estimated that in 2015, 45% of Canadian males and 42% of Canadian females have a lifetime probability of developing some form of cancer, and approximately one in four Canadians will die of cancer. According to the Canadian Cancer Society, 196,900 new cases of cancer were expected to be diagnosed during 2015. Lung, breast, colorectal and prostate cancer, were expected to represent more than half (51%) of these

new cases. Cancer is the leading cause of death in Canada and is responsible for 30% of all deaths.

1.3.2 Cancer Treatment

Cancer care demands a multidisciplinary approach. Cancer specialists including radiation oncologists, medical oncologists, surgeons, medical physicists, radiation therapists, nurses, radiologists, and other medical personnel, often work together to create a patient's overall treatment plan that effectively treats the cancer while attempting to allow the patient to maintain a high quality of life. Most cancer types can be treated using one or more of three treatment modalities: (1) Surgery (localized resection of the tumour), (2) chemotherapy (use of cytotoxic drugs that kill primary tumour cells and those that may be circulating through the body), and (3) radiation therapy (use of ionizing radiation to kill tumour cells).

Some patients with cancer may receive only one treatment modality (radical treatment) but most patients undergo a combination of treatments (adjuvant treatment). Radiation therapy may be used radically or adjuvantly depending on the tumour type, location, stage (how advanced it is), grade (how aggressive it is), patient preference, as well as the general health of the patient (Washington and Leaver, 2015). More than half of all patients with cancer receive radiation therapy as part of their treatment (El Naqa et al., 2015).

1.4 Radiation Therapy In Practice

As stated above, the goal of radiation therapy is to deliver a prescribed dose of radiation to a target (typically a tumour) while minimizing radiation damage to the surrounding healthy tissue. Modern radiation therapy involves highly conformal techniques such as intensity-modulated radiation therapy (IMRT), stereotactic radiosurgery, volumetric modulated arc therapy and proton therapy that almost "paint" the target three-dimensionally with radiation, which greatly spares normal tissues (Washington and Leaver, 2015). Following diagnosis, and a decision to treat with radiation therapy, a radiation oncologist prescribes a dose of radiation, and works with the other members of the radiation oncology team to determine a suitable treatment plan and treatment schedule.

1.4.1 The Radiation Oncology Team

The effectiveness of patient care and treatment relies on unity and synergy amongst the patient and members of the entire radiation therapy department (Figure 1.1). From the receptionist to the physician, each team member has a responsibility in achieving the goal of treating a patient with cancer. The radiation oncologist has the overall responsibility for the patient's care and treatment. Other members of the radiation therapy team work under the direction of the oncologist. The outlining of simple organs on radiographic images is the responsibility of a dosimetrist. These images and outlines, set up by the dosimetrists, are used by a radiation oncologist to conduct detailed planning such as outlining the tumour and the organs at risk of radiation. The treatment itself is administered by a radiation therapist. Responsibility for the overall quality of the treatment (physical accuracy and reliable delivery of the radiation dose) falls onto the medical physicist (See section 1.5). Table 1.1 provides an overview of the various radiation oncology team members with their associated roles and responsibilities.

The patient is also an essential member of their radiation oncology team. The patient works in collaboration with other members of the team by complying with treatment requirements and reporting back the outcome of their



Figure 1.1: The radiation oncology team is a patient-centred group that consists of various radiation medicine personnel fulfilling the goal of diagnosing and treating a patient with cancer. Proximity to the patient in this model represents the closeness of involvement with the patient in practice – Direct (inner ring) and indirect involvement (outer ring). Table 1.1 provides descriptions of radiation oncology team members with their associate responsibilities.

Profession	Responsabilities
Radiation Oncologist	A specialized physician responsible for prescribing ionizing radiation to treat cancer. Works with other members of the radiation therapy team to ensure each treatment is delivered accurately and safely. Contours the tumour and organs at risk (of radiation) on radiographic images.
Clinical Medical Physicist	Ensures accurate delivery of radiation through calibration and maintenance of treatment machines and other radiation related instrumentation. Maintains proper radiation safety programs to prevent excessive radiation exposure to staff, patients and the public. Collaborates with other members of the radiation therapy team to solve various physical, mathematical and computational problems.
Radiation Therapist	Administers radiation treatments by operating radiation therapy equipment. Monitors the condition of patients and assesses changes in the treatment plan, if required, such as re-positioning or dose re-calculation.
Engineer	Ensures safe and secure operation of treatment machines and other radiation related instrumentation. Works with a radiation safety officer to ensure proper installation, maintenance, and decommissioning of radiation machines.
Nurse	Monitors the patient's radiation treatment experience including assessing symp- toms, therapeutic communication, information, and overall care, psychosocial support and referring to other radiation medicine professions to manage identi- fied problems.
Receptionist	Schedules appointments and check-ins for patients.
Social Worker	Helps patients navigate through the hospital system so that they can receive efficient and streamlined care. Provides counselling and supportive services.
Medical Dosimetrist	Contours simple organs on radiographic images for radiation oncologists to use. Employs treatment planning techniques to calculate where and how to distribute radiation. Ensures the physician's prescription is met while minimizing damage to surrounding tissues.
Other Allied Health Groups & Professionals	A group of health care providers who apply their expertise and deliver support to patients and other members of the radiation oncology team. Providers such as: dentists (for Head and Neck patients), patient support groups, and nutritionists.

 Table 1.1: Radiation oncology team members and a brief description of their responsibilities

treatment, including their concerns and worries on an emotional, social, psychological, and physical level. The patient also brings in their family and support system that can have an additional effect on their treatment outcome. The relationship between the patient and radiation therapy staff is as important as any other task accomplished by staff members during the course of treatment. The complete radiation therapy process is detailed in section 2.2.1.

1.5 Medical Physics

Medical physics is a discipline of physics that applies the science of physics to medicine, namely in the diagnosis and treatment of disease using radiation. Radiation includes both ionizing and non-ionizing radiation such as particulate radiation, electromagnetic radiation, and ultrasound. In radiation oncology, ionizing radiation is used to treat a wide variety of cancers through external beam radiation therapy (in which radiation is delivered from outside the patient's body) or brachytherapy (in which a sealed radioactive source is temporarily or permanently placed directly into the region of the patient's body to be treated). Medical physics research and a medical physicist's clinical services are essential to maintaining and improving the success of radiation therapy.

1.5.1 The Role of a Medical Physicist

Medical physicists are concerned with three main areas of practice: clinical services, research and development, and teaching.

Clinical Service

The American Association of Physicists in Medicine (AAPM) states that: "The essential responsibility of the Qualified Medical Physicist's clinical practice is to assure the safe and effective delivery of radiation to achieve a diagnostic or therapeutic result as prescribed in patient care" (aapm.org). In other words, a clinical medical physicist is responsible for quality assurance of all radiation therapy and diagnostic radiology equipment and associated processes. This includes: (1) overseeing all treatment planning (for example establishing adequate protocols to ensure accurate patient dosimetry); (2) incorporating proper radiation safety programs to prevent excessive radiation exposure to staff and the public (from the design of radiation installations to the control of radioactive substances); (3) quality control and operation for accurate measurement and reproducibility of radiation output employed in imaging and therapy; and (4) assisting and collaborating with other health care professionals by providing scientific advice and resources to solve various physical problems that may arise in specialized medical operations.

Research and Development

Medical physicists play a crucial role in medical research. Their activities are well established in the main areas of active research and development such as academia (for example, investigating how radiation affects tissue) and industry (for example, building new imaging and therapy systems). Medical physicists are also concerned with the application of digital computers in medicine (see section 1.5.2) including processing, storing, and retrieving medical images and medical data. Indeed, the role of a medical physicist in research is to continually develop and improve instrumentation and technology for use in diagnostic and therapeutic care.

Teaching

Many medical physicists are affiliated with universities, where they help teach and train future medical physicists, radiation oncology residents, and radiation therapists. They conduct courses in radiation physics, diagnostic imaging, computers in medical physics, applied dosimetry, radiation biology nuclear medicine and radiation protection. Many universities and clinics across Canada provide training and teaching programs in medical physics. The Commission on Accreditation of Medical Physics Education Programs (campep.org), accredits most graduate programs and assures high educational standards in medical physics.

8

1.5.2 Informatics—The Future of Medical Physics

The current challenge of medical physics lies in a shift towards personalized medicine (i.e. tailoring care and treatment on an individual-based level using predictive outcomes). As radiation treatments become more precise, physicists are becoming increasingly responsible for developing and monitoring treatments using sophisticated instrumentation and computer-based platforms. The delivery of patient services in radiation oncology is technologically demanding and requires intense involvement of medical physicists at all levels. Since the advent of computers, medical institutions increasingly process information digitally. This growth of digital information has the potential to move the practice of medical physics into a data-driven science known as informatics, in which data can be used to inform clinical practice and promote knowledge-sharing between radiation medicine professionals. The impact of this transformation on the quality of patient care and personalized practice has great potential. Realization of this potential, from a data perspective, rests squarely on the shoulders of medical physicists.

CHAPTER 2 Waiting Time Uncertainty — A Universal Challenge

Where there are encounters, there are waits. Whether it's waiting for a service personnel (electrician, plumber, internet service provider) to come and deal with an issue, waiting to pick someone up from the airport, waiting for a response (by email or phone), waiting in line to purchase goods, or waiting to see a doctor after a call about test results, waiting is something we experience on a daily basis. In most situations, waiting is inevitable and unavoidable (e.g. insufficient resources to meet the demand or general lack of control of the situation). The problem therein lies in the uncertainty of waiting, such as unexpected delays or simply not knowing how long to wait. Situations like these may appear unpredictable and out of control. As a result, waiting may develop into an anxious and painful experience, that can have significant psychological, social, financial, or physical costs on the entire service, including those affected by the wait.

Waiting time uncertainty is a universal challenge that can occur anywhere, and the healthcare system is no exception. This chapter first outlines some of the key implications of waiting for healthcare services. Concerns about waiting times have been linked to patient health, patient outcomes, and overall quality of life. The remainder of this chapter describes waiting in the context of radiation oncology at the MUHC and concludes with a method of addressing waiting time uncertainty – through auditing electronic medical records to learn from the experiences of previous patients.

2.1 The Effects of Waiting in Healthcare

Delays in healthcare services are one of the most challenging problems facing the Canadian healthcare system today (Fraser Institute, 2013). The impact of delayed treatment varies by condition. For conditions such as cardiac disease, cataract surgery, and cancer, longer waiting times can often lead to more serious and sudden adverse events such as deterioration, disability or even death (Sobolev et al., 2008; Kulkarni et al., 2009; Hodge et al., 2007; Guttmann et al., 2011). For orthopaedic surgeries, increased delays may have lifelong impacts on health including mental and emotional well-being (Wright and Menaker, 2011). For some conditions, longer waits can be devastating for individuals in terms of personal burdens on families and friends (Fraser Institute, 2013).

Waiting for healthcare in not a benign process nor are delays a mere inconvenience for patients. Delays and the lack of explaining them can be extremely harmful to patients. In addition to important medical, psychological, and personal consequences of longer waits, the uncertainty in waiting times can lead to increased stress, dissatisfaction, poorer outcomes of care and can impact the healthcare system overall (Fraser Institute, 2013; CIHI, 2012).

2.2 Waiting For Care in Radiation Oncology

In radiation oncology, waiting time concern is a well-known determinant of overall patient satisfaction (Paul et al., 2012) but the effect of waiting time uncertainty is often overlooked. In a recent local survey (Rossy Cancer Network, AOPSS survey, private communication), 73% of MUHC cancer patients who expressed dissatisfaction with their care stated that their initial consultation delay was not explained to them. 68% of MUHC cancer patients felt that staff did not do everything they could to make their radiation therapy waits comfortable. In the absence of progress in reducing waiting times, informing patients of their expected delays and providing explanations for them is one way in which dissatisfaction may be reduced (Famiglietti et al., 2013).

The radiation oncology department at the MUHC operates on an electronic "tasking" system. These tasks consist of a time issued, due date and time, and time completed timestamps, along with a description of the work required, the name of the sender and the name of the receiver, and are sent amongst staff members and doctors as records of events. Appointments in the radiation oncology department also operate electronically, in which they include the same metadata timestamps (i.e. the same timestamp information) as tasks, and usually include information on the attending physician, rooms to be booked, and patient information. Where there are recorded timestamps, there are records of waits. The following two subsections describe the patient radiation treatment pathway from start to finish and the periods in which waiting can occur.

2.2.1 The Patient Pathway

The patient's pathway in radiation oncology at the MUHC begins with a consultation request sent via fax from the referring physician to the booking centre. Two pieces of information are typically provided in the incoming request: (1) information on the referring institution and cancer type, and (2) detailed patient information including tissue biopsies, diagnosis, and date of diagnosis. After receiving the request, the booking staff choose from eight consultation types, based on whether treatment is to be curative or palliative (symptom-relief), patient hospitalization, and whether the patient is a new or returning patient. The booking staff determine a time for the chosen initial consultation according to doctor and patient availabilities, and the patient then arrives at the set time for consultation.

When arriving at the clinic, patients check in either at the reception or at a self-check in machine (using their medicare card). The room, the time they are seen, and the attending physician are all recorded electronically into their electronic medical record. After the patient is seen, the time the appointment was completed is recorded.

After the initial consultation, the patient's radiation oncologist reviews the patient's chart, goes over the patient's history, and ideally on the same day, fills out a radiation oncology requisition (ROR). The ROR outlines a basic recommended treatment plan, including an idea of prescribed dosage, tumour/treatment site, and how long the patient can wait to be treated (priority level). Either treatment is not indicated, in which case the patient's radiation treatment pathway ends there, or the patient has missing lab tests, in which case, tests are ordered and the patient follows up. If radiation treatment is indicated, the patient is classified according to the provincial priority level (four levels), and the patient start date must be set accordingly from the date the patient is medically ready for treatment. From this point, a computed tomography (CT) scan is tasked to the liaison office by the primary physician. The patient is assigned both a CT simulation scanner for CT simulation scan and a treatment machine for the duration of their treatment. The liaison office communicates with other departments to determine, for example, when chemotherapy finishes and radiation treatment is ready (or sometimes chemotherapy can be administered adjuvantly with radiation). After the patient receives the CT simulation scan, they are sent home and a series of steps–all mediated by the electronic tasking system–take place before they are contacted for the beginning of radiation treatment.

In external beam radiation therapy, radiation treatment planning starts with the medical dosimetrists who contour the CT images for the physician, labelling the organs at risk of radiation damage. The physician is responsible for contouring the tumour region (gross tumour volume, GTV), an additional margin for sub-cellular disease around the tumour (clinical target volume, CTV), and subsequently a further region for planning setup and internal movement variations (planning target volume, PTV) (Jones, 1994). After the physician notes organ radiation limits and other restrictions in the CT planning sheet, the plan is ready for dose calculation to ensure the PTV gets the full dose of radiation and the surrounding tissues get minimal radiation, as per the physician's intent. After dose calculation, the physician approves the treatment plan (electronically) and writes a prescription note. The plan is ready then for the medical physicists to approve the physical aspects. At any step, the plan may need to be changed and must be sent back in the pipeline.

Finally, when the treatment plan is approved, the patient is contacted for daily fractionated treatments, where a fraction of the total prescribed dose is delivered on a daily basis. The patient routinely checks in at the clinic on a daily basis to receive radiation treatment. On average, patients are treated in 25 fractions (five fractions per week, Monday to Friday, for five weeks). However, fractionated treatment can range between one and 35 fractions, depending on the treatment intent and site. During the treatment session, Radiation therapists set the patient up on the treatment table according to the planned positions. Radiation delivery is administered by the radiation therapists. Setup positions, images, the opening and closing of patient's charts, and radiation delivery are electronically recorded into the patient's electronic medical record. The radiation oncologist typically verifies patient positioning using the recorded images on a weekly basis, depending on the treatment site.

2.2.2 Waiting Time Occurrences

In the context of radiation oncology at the MUHC, patients encounter three types of waiting time during their treatment pathway, as indicated by the colours in Figure 2.1. The first type of wait is the interval between the time the patient receives their CT-scan and their first treatment appointment. During this time, the patient is waiting at home, usually for days to weeks, for their treatment plan to be prepared by staff and physicians. The second wait occurs on-site in the clinical waiting room, after the patient has been called to start daily radiotherapy treatment. The patient arrives at the clinic on a daily basis to receive fractionated treatment and may wait between minutes to hours. The third type of wait occurs in the waiting room pre-, during, and posttreatment, when the patient is scheduled for a visit with the physician for consultation, intra-treatment examination, and follow-up, respectfully, and may wait between minutes to hours. These waiting times are difficult for staff to predict and typically rough estimates are provided, based on the personal experience of the staff involved. The uncertainty inherent to these estimates is a source of stress for staff who field inquiries from concerned patients/relatives without confidence in the answers they provide. Due to this uncertainty, most patients are left unable to plan their calendars and daily lives, making the waiting experience in radiation oncology uncomfortable, even painful.

2.3 Waiting Time Studies

Over the past few decades, waiting times in Radiation Oncology have been studied extensively (Robinson et al., 2005; Fortin et al., 2006; Jack et al., 2007). The main focus of these studies was to investigate waiting time variations between pre-treatment events (such as initial consultation, diagnosis,



Figure 2.1: An overview of the patient's radiotherapy pathway with colour-coded waiting periods. In radiation oncology, patients experience three types of waiting times. One type of wait occurs when the patient is waiting at home for their treatment plan to be prepared by staff after receiving a CT-scan (purple). Another waiting period occurs in the waiting room for daily fractionated radiotherapy (blue). Similarly, another waiting period occurs in the waiting room to see a physician in consultation, intra-treatment examination, or follow-up (green).

surgery or chemotherapy) to the start of radiotherapy treatment and the factors that cause variations among waiting times (Jack et al., 2007). A retrospective study in Quebec found that variations in the waiting time between surgery and the start of post-operative radiotherapy for breast cancer patients were due to the number of cases, demand and supply of treatment resources, localized cancer stage and proximity to a radiotherapy facility (Fortin et al., 2006). Another study hypothesized that a factor contributing to lengthening waiting times could be the increase in treatment complexity associated with recent widespread applications of three-dimensional conformal and intensitymodulated radiotherapy (Robinson et al., 2005). These studies provide much incentive to try to decrease waiting times for radiotherapy patients by attempting to describe features that influence them.

Other studies address waiting times for daily radiotherapy appointments. In one prospective study (Chan et al., 2010), treatment delays (calculated as the difference between the scheduled appointment time and the actual appointment time) were mainly attributed to an indirect cause of catching up from previous delayed appointments. The mean wait time was found to be longest for midday appointments. They also found that patients treated in the pelvic region experienced the longest wait time due to its specific treatment protocol, suggesting variations in wait time by treatment site.

2.4 The Significance of Electronic Health Records

Waiting times in radiation oncology have been extensively analyzed for the sole purpose of understanding and explaining the factors that affect delays in order to decrease waiting times for radiation therapy patients (Robinson et al., 2005; Fortin et al., 2006; Jack et al., 2007). However, until now, there are no reports in the literature that computationally approach the uncertainty in radiation therapy waiting times; harnessing previous patient data from electronic health records (EHRs) to provide predictions for future patients.

In the present era of EHRs, waiting times need not be uncertain. EHRs contain digital information about a patient's history, their treatment pathways, and their encounters including, but not limited to, demographics, diagnoses, radiographic images, appointment dates, treatment plans, and lab test results. The increased availability and growth of electronic health data can drive deeper insights and guide decision-makers to improve performance and the overall quality of care. For example, in large amounts, data within cohorts of similar patients tend to cluster towards representative values which may reveal statistical patterns and trends. Similar patients, according to their health records, may experience similar waiting times and thus personalized prediction may be possible based on the data of similar previously-treated patients.

EHRs are creating new opportunities to use health data to personalize care, support decision-making and improve patient outcomes (CIHI, 2013).

Because of its volume, electronic health data must be analyzed and transformed into information that enables personalized medicine in real-time to benefit patients and clinicians. New analytical methods, more efficient processing and automation tools are making it easier to draw insights from health data. Technological innovations, such as machine learning, are enabling computer systems to learn from large, potentially complex datasets and support real-time predictive functions. Thus, today's technology can play a vital role in improving the quality of patient care. The potential to reduce waiting time uncertainty through the use of data has never been better.

CHAPTER 3 Machine Learning—The Personalized Solution

3.1 Overview

Machine learning (ML) is an area of computer science and artificial intelligence that aims to develop sophisticated computer algorithms that learn from data to solve particular tasks without being explicitly programmed (i.e. hardcoded). These "soft-coded" algorithms are, in essence, designed to emulate human intelligence by adapting to changing environments through repetition (i.e. experience) such that they become better and better at reaching a desired goal. The concept of adaption is referred to as *training*, in which samples of input data are provided to the algorithm, along with the desired outcomes for the algorithm to produce. During training, the algorithm self-optimizes in such a way that it is not only able to produce the desired outcome from input datasets but also able to generalize outcomes for new, previously-unseen data.

There are two benefits to a successful machine learning algorithm. The first is that it can be used to substitute laborious and repetitive tasks. The second is that it can potentially detect hard-to-discern patterns from large, noisy or complex datasets better than the average human observer. These capabilities are particularly beneficial to radiation therapy. For example, contouring is an integral part of the treatment planning process in which several specialists outline tumours and organs at risk on radiographic images. Contouring is a time consuming procedure that is highly dependent on the observer's familiarity and knowledge of the human anatomy on radiographic images. This
familiarity, though, has its limits, and consequently, there are risks of subjectivity and inter-observer variability in the resulting contours. A machine learning algorithm can aid in detecting patterns and subtleties in textures and shapes by incorporating data from multiple observers, thereby combining the experience to reduce the contouring uncertainty.

A significant portion of radiation oncology practice is dynamic and involves not only intercommunication amongst treatment team members, but also several stages of human-machine interactions and decision-making, which naturally invite suitable machine learning algorithms to automate and optimize these interactions. The ability of machine learning algorithms to learn from current situations (and generalize into unseen data) allows improvements in the overall quality and safety of radiation therapy practice and should lead to better outcomes.

3.2 Learning From Data

Learning from data requires inference of unknowns from knowns. Using treatment appointment durations as an example, the main components of a machine learning problem are as follows. There exists an input vector, \mathbf{x} , where each x_i represents a unique "feature" that describes the output, y (for example, patient information that influences a patient's appointment duration), an unknown target function $f : X \to Y$, which corresponds to the ideal formula to predict the duration of an appointment, where X is the input space (the set of all possible inputs \mathbf{x}), and Y is the output space (the set of all possible outputs y, in this case, a real-valued duration). There exists a dataset D of input-output training examples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$, where $y_n = f(\mathbf{x}_n)$ for $n = 1, \ldots, N$ (inputs corresponding to previous patient data and their corresponding appointment durations known in hindsight). Finally, there exists a learning algorithm that uses the dataset D to pick a formula $g : X \to Y$



Figure 3.1: An overview of the machine learning setup. Adapted from (Abu-Mostafa et al., 2012).

that approximates f. The algorithm chooses g from a set of possible formulas, called the hypothesis set H. To achieve that, the algorithm chooses the gthat best matches f on the training examples of previous patient data, with the expectation that it will closely match f on new patient data. When a new patient checks in for their appointment, their waiting time is inferred by applying machine learning to estimate the appointment durations of those patients ahead in the queue who have yet to be seen (as described in section 4.3). The predicted individual durations of those appointments is based on g (the hypothesis that the learning algorithm produced), not on f (the ideal target function which remains unknown). Figure 3.1 illustrates the components of the learning setup.



Figure 3.2: Machine learning according to the nature of the input training data. In unsupervised learning, the outputs are not defined (i.e. unlabelled). In semi-supervised learning, some of the outputs are defined (i.e. semilabelled). In supervised learning, all the outputs are defined (i.e. labelled).

3.2.1 Types of Learning

Depending on the nature of the problem, machine learning can be divided into supervised, unsupervised, and semi-supervised learning, as shown in figure 3.2. The learning problem that was discussed above (in the context of treatment appointments) is an example of *supervised* learning. In supervised learning, the algorithm is trained on a pre-defined set of training examples, where the output is defined (i.e. labelled data). The algorithm's job is to estimate an unknown (input, output) mapping from the known (input, output) training samples.

In an unsupervised setting, the training data does not contain any output information at all. In other words, a machine learning algorithm is only given input examples $\mathbf{x}_1, \ldots, \mathbf{x}_N$. Unsupervised learning can be interpreted as the task of finding clusters and patterns in the input data space. For example, if our task is to categorize a set of foreign coins, and we only use general properties of these coins (size, mass, etc.), then we can identify coins that have similar properties and put them together, without knowing the underlying denomination.

Semi-supervised learning is a combination of both supervised and unsupervised learning where the input data is partially labelled and the labelled part is used to infer the unlabeled portion. An example of semi-supervised learning is studying for a written exam. The labelled data are the few example problems that the teacher solves in class. The unlabelled data are the set of unsolved problems the teacher provides either as: (a) a take-home exam (in which the general goal is do well on them); or (b) as similar problems that will be encountered during the real exam.

3.3 Learning Algorithms

Depending on the format of the desired output, machine learning algorithms can be divided into two classes: classification and regression. Classification involves estimating a discrete value, or more generally, a categorical variable. For example, modelling a diagnostic image to predict whether a tumour is malignant or benign. The output variables in this case are one of two categories: the tumour is malignant, or the tumour is benign. In the case of regression problems, a machine learning algorithm maps the input data onto a continuous output value. For example, predicting the time for tumour recurrence. In this case, the output (i.e. time) is real-valued.

3.3.1 List of Common Machine Learning Algorithms

The following is a list of common machine learning algorithms that can be applied to almost any data problem. Table 3.1 provides a summary of algorithms along with the type of learning.

ML Algorithm	Model Type	Learning Method
Linear Regression	Regression	Supervised
Logistic Regression	Classification	Supervised
Decision Tree	Classification & Regression	Supervised
Support Vector Machine	Classification & Regression	Supervised
K-Means	Classification	Unsupervised
		-
Random Forest	Classification & Regression	Supervised

 Table 3.1: List of common machine learning algorithms along with the model type (classification or regression) and learning method (supervised or unsupervised)

Linear Regression

As the name suggests, linear regression is based on continuous variables. In this problem, a relationship between independent (i.e. features) and dependent (i.e. target output) variables is established by fitting a best-fit line. The bestfit line is known as the regression line and is represented by a linear equation y = a * x + b, where y is the dependent variable, a is the slope of the line, x is the independent variable, and b is the intercept. The coefficients a and b are derived based on minimizing the sum squared difference of distance between the data points and the regression line.

A simple example of linear regression is the relationship between a person's height and weight, shown in figure 3.3. In this example, the best fit line is determined. Using the equation of the best fit line, a person's weight can be determined from their height.

Logistic Regression

Despite its name, logistic regression is a classification algorithm, not a regression algorithm. It it used to estimate discrete values typically in the form of binary classifications (true/false, 0/1, yes/no, etc.), based on a set of independent variable(s). In essence, a logistic regression algorithm predicts the



Figure 3.3: An example linear regression model. In this example, a person's weight is a function of one feature; the person's height. With a regression line defined (red), a new person's weight can be predicted from their height. Adapted from (Buitinck et al., 2013).

probability of an event occurring by fitting data to a *logit* function. Since it predicts a probability, the output value lies between 0 and 1, as expected.

As an example, assume the relationship between the probability of passing an exam and the hours spent studying, shown in figure 3.4. There are only two outcomes in this scenario – pass or fail. However, in logistic regression, the scenario is modelled as a probability of achieving either outcome. In this example, if a person studied for three hours, they are around 80% likely to pass. On the other hand, if a person studied for one hour, the probability of passing is only 5%.

Decision Tree

A decision tree is a supervised learning algorithm that is used for both classification and regression type problems. In this technique, each node (i.e. decision point) represents a single feature, x_i , and a split is made on the most



Figure 3.4: An example logistic regression model. In this classification example, there are two outcomes – either a person passes an exam or fails. The two outcomes are a function of one feature; the amount of hours spent studying. A logit function (red) models the scenario as a prediction probability of achieving either outcome. For example, if a person studies for three hours, they are around 80% likely to pass. Adapted from (Buitinck et al., 2013).



Figure 3.5: A decision tree example. In this classification example, the objective is to predict whether a person is male or female. According to the decision nodes, the first split is based on the most significant feature; a person's height. If the person is shorter than 170 cm, then a second decision is made based on the second most significant feature; their weight. This is a crude example of course, but essentially new outcomes are predicted by traversing the tree model and evaluating the specific input features at each node (blue) until an end leaf node (green) is reached.

significant feature that makes each child node as distinct as possible. The leaf nodes (i.e. the end nodes) of the tree contain an output variable, y, which is used to make the prediction.

An example of a decision tree model is shown in figure 3.5. Suppose a dataset contains two features: height and weight, and the objective is to predict a gender classification: male or female. With this example, making a prediction is relatively straightforward. Given a new input, the tree is traversed by evaluating the specific features starting at the root node of the tree.

All split points of a decision tree are evaluated and chosen in a greedy fashion (i.e. the best split feature is chosen at each node each time). For regression trees, the procedure of splitting the dataset involves minimizing the sum squared error across all training samples that fall within the node.

Decision trees are very susceptible to overfitting. Unless some stopping procedure is specified, splitting can occur repeatedly until every possible decision is made to fit the training data perfectly. Overfitting the training data is likely to have poor performance on the testing set. The most common stopping criterion is to specify the minimum number of training instances assigned to each leaf node. If a split results in a count that is less than some minimum, then the split is rejected and the current node is taken as the final leaf node.

Support Vector Machine

Support vector machine (SVM) algorithms are used for both classification and regression type problems. In this technique, data points are plotted in n-dimensional space (where n is the number of features) with the value of each feature being the value of a particular coordinate. In a classification setting, a support vector machine constructs a hyper-plane in high dimensional space to achieve good separation that has the largest distance to the data points of any class. Conversely, in a regression situation, a support vector machine seeks to construct a hyper-plane that best fits the data to within some margin.

An example of a classification SVM problem is shown in figure 3.6. In this two-dimensional example, there are two features and two classes. Each axis represents a feature and each colour represents a class (blue and red). The objective is to determine some *line* that best splits the two classes of data. This will be the line such that the closest point in each class will be furthest away from the line. With a line defined, the closest points in each class that are the furthest apart from the line are known as *support vectors*. Support vectors determine the margin in which separation occurs.



Figure 3.6: A support vector machine classification example. In this example, there are two features and two classifications. Each input feature represents one coordinate and each target class is labelled as colours (red and blue). The objective is to predict a class from new data. A SVM model determines the best split on the training data such that the closest training point(s) in each class are furthest away from the line. These closest points are known as *support vectors* and are defined as margins (dotted lines) in which separation occurs (Buitinck et al., 2013).



Figure 3.7: A support vector machine regression example. In this example, the target output is plotted as a function of one feature. An SVM model fits the training data to within some margin defined by an epsilon (ϵ) parameter (dotted band). The optimization process ignores training points that are located within the band and attempts to minimize the errors on the training points outside the band. Adapted from (Buitinck et al., 2013).

An example of a regression SVM problem is shown in figure 3.7. In this one-dimensional example (similar to linear regression), each training sample is plotted along one feature and its corresponding output, on the x- and yaxes respectively. The objective is to construct a line that best fits these points to within some margin. This margin is known as an *epsilon* band. The optimization process ignores errors which are situated within the certain distance of the true value (epsilon band) and minimizes the errors on the training points outside of the band. An important and beneficial feature of SVM is that the construction of a hyper-plane occurs in the smallest dimension that linearly separates (or fits) the data. For example, if an ellipse (non-linear) bound is required to separate two-dimensional data points, then a *kernel* function (i.e. non-linear mapping) is applied on the data that maps the data into a higher dimension in which linear separability is possible.

K-Means

K-means is an unsupervised algorithm that is used to classify data through a certain number of clusters (k clusters). In essence, k-means is performed by the following four steps: (1) the algorithm initially picks k number of points (at random) for each cluster, known as *centroids*; (2) the algorithm calculates the sum of squares distance of each data point in a cluster to the centroid; (3) the algorithm finds a new centroid data point in each cluster based of minimizing the within-cluster sum of squares distance; (4) with new centroids, repeat steps 2 and 3. Repeat until convergence occurs (i.e. centroids do not change).

A visual example of a fully-formed k-means clustering is shown in figure 3.8.

Random Forest

Random forest is a versatile machine learning technique capable of performing both classification and regression. A random forest is a random ensemble of decision tress (hence the term "forest"). To classify an object, a random forest model constructs multiple decision trees with randomly-selected input samples and initial features (both sampled with replacement). Each tree decides (or "votes") on a classification or regression output based on the randomly selected features. The forest chooses the output having the most votes over all the trees in the forest. An illustration is shown in figure 3.9.



Figure 3.8: A visual representation of a k-means example model. In this unsupervised example, ten centroids are defined from unlabelled training data. Clusters are developed based on the minimum sum of squares distance from each data point to a centroid. With a fully-formed k-means model, a class can be predicted based on which cluster new data points fall into (Buitinck et al., 2013).



Figure 3.9: An example of a random forest model. A random forest is a collection of decision trees. When developing a model on training data, each decision tree decides on an output based on randomly selected input samples and features. The forest chooses the output based on a majority vote from all decision trees.

One of the benefits of random forests, compared to a single decision tree, is its inherent capability of achieving better accuracy and model stability. As more decision trees are added to the ensemble, each with its own bias and variance, noise and error is reduced in the final prediction. It is also effective in maintaining accuracy when a portion of the data is missing.

Another benefit is its ability to handle large datasets with high dimensionality. A random forest model can handle thousands of input features and identify the most significant feature(s). Feature importance is a very useful dimensionality reduction technique that comes with random forest models.

One drawback to using a random forest model is the loss of intuition when explaining the data. A single decision tree has the benefit of graphically depicting exactly how and where a particular decision was made. In a random forest mode, the graphical representation is lost (due to the ensemble of trees) and there is very little control on what the model does.

3.3.2 Performance Evaluation

Once a model is obtained using one or more machine learning algorithms, it is important to estimate its performance. The performance of a classification model is measured in terms of specificity, sensitivity, accuracy, and area under the curve (AUC). *Specificity* is defined as the proportion of true negatives that are correctly observed, whereas *sensitivity* is given by the proportion of true positives that are correctly identified by the classifier. *Accuracy* and *AUC* are metrics used to evaluate the overall performance of the classifier. Specifically, accuracy is a measure related to the total number of correct predictions. In contrast, AUC is a measure that is based on a receiver operating characteristic (ROC) curve that plots the tradeoffs between sensitivity and 1-specificity (false positive rate). An example plot of various ROC curves is illustrated in figure 3.10.



Figure 3.10: Example ROC plots of various classification models. A perfect model would have an area under the curve (AUC) of 1.0 (i.e. 0.0 false positive rate and 1.0 true positive rate). The best model in this example is the blue curve having an AUC of 0.91. A perfect random classifier (shown by the dotted line) would have an AUC of 0.5. Adapted from (Buitinck et al., 2013).

The performance of a regression model is measured in terms of mean absolute error (MAE), median absolute error (MedAE), mean squared error (MSE), and the coefficient of determination (also known as R^2). The term "error" here represents the difference between the predicted value and the true value; this is also known as the *residual*. The absolute value or the square of this difference is measured to capture the total magnitude of error across all instances, as the difference between the predicted value and the true value may be negative. The error metrics evaluate the predictive performance of a regression model either in terms of the mean deviation or the median deviation of its predictions from the true values. The lower the error values, the more accurate the model is at making predictions. An overall error metric of zero means that the model fits the data perfectly.

Mean Absolute Error

If \hat{y}_i is the predicted value of the *i*-th sample, and y_i is the corresponding true value, the mean absolute error, estimated over $n_{samples}$ is expressed as

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i|$$
(3.1)

Median Absolute Error

The median absolute error is expressed as

$$MedAE(y,\hat{y}) = median(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$
(3.2)

Mean Squared Error

The mean squared error is defined as

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2$$
(3.3)

Coefficient of Determination

The R^2 metric provides an overall indication of the goodness of fit of a set of predictions to the actual values. The best possible value is 1, whereas the worse value is 0. R^2 is expressed as

$$R^{2}(y,\hat{y}) = 1 - \frac{\sum_{i=0}^{n_{samples}-1} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=0}^{n_{samples}-1} (y_{i} - \bar{y}_{i})^{2}}$$
(3.4)

where $\bar{y} = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} y_i$.

3.3.3 Sampling

The performance evaluation metrics introduced in the previous section are computed from a testing set which is a subset of the overall input dataset. In order to obtain reliable results in terms of the prediction performance of a model, training and testing sets should be sufficiently large and independent, while the outputs of the testing samples should be known (in the case of supervised learning). The most commonly-used methods for splitting the input datasets into subsets are: (a) cross-validation, (b) bootstrapping, (c) holdout method, and (d) random sampling. In *cross-validation*, each sample is used the same number of times for training and only once for testing. The total accuracy of this method is calculated as the average of all accuracies during each validation cycle. In the *bootstrapping approach*, the input dataset is separated into the training and testing sets. The samples chosen for training are placed again into the entire dataset for resampling. In the *holdout* method, the samples are simply partitioned into the two separate sets, typically in a 80/20 fashion for training and testing respectively. *Random sampling* is a repetition of the holdout method, choosing the training and testing samples randomly in order to better estimate the accuracy of the model.

In the case of the holdout method, there are two competing concerns when dividing the overall input dataset into training and testing sets. With less training data, the feature estimates will have a greater variance. With less testing data, the performance statistics will have greater variance. For a sufficiently large input datasets, there is very little dependence between choosing a 80/20 ratio versus a 70/30 or 90/10. Thus, 80/20 is a typically used ratio. However, for small input datasets, there may be no split that would result in a satisfactory variance in the estimates. Thus, in this case, other sampling methods, such as cross-validation or bootstrapping, can be more appropriate.

3.3.4 Feature Selection

Feature selection is an essential processing step prior to applying a machine learning algorithm. Features are predictors that influence the overall performance of the system. The objective of feature selection is three-fold: (1) improving the performance of machine learning algorithms by the removal of redundant and irrelevant input information, (2) dimensionality reduction by providing faster and more cost-effective predictors, and (3) providing a better understanding of the overall learning process. Ultimately, the degree in which the output varies according to each feature value is what determines the best and most influential feature.

Features can be selected *ab initio* according to an observer's intuitive understanding of the overall system. In other words, before constructing a learning algorithm, a human observer can decide which attributes influence the specific task at hand, thereby using a pre-determined feature set. This has the benefit of narrowing the dimensionality of the system by introducing human experience since humans know first-hand or at least have an idea of the interdependencies of the overall system. However, the drawback to this approach is possibly missing relevant, hidden information that the algorithm itself may or may not find.

There are two common approaches in which a machine learning algorithm evaluates the usefulness of features: (1) the wrapper method and (2) the filter method. The *wrapper* approach uses the prediction performance of the learning algorithm to determine the usefulness of subsets of features (Guyon and Elisseeff, 2003). A wide range of strategies are available to determine the best possible features but the most commonly used are greedy strategies that are computationally advantageous and resilient against overfitting. For instance, the algorithm can start with no features and successively add predictors, or it



(b) Filter method.

Figure 3.11: Two methods for evaluating features for a machine learning algorithm.

can start with all features and successively eliminate the least promising one. The former is known as *forward selection* and the latter is known as *backward elimination* (Guyon and Elisseeff, 2003).

The *filter* method is an alternative approach to feature selection. The filter method considers features independent of the learning algorithm that will use them. This methodology relies on general characteristics (such as correlation) of the training set to select the most relevant variables and exclude others. Compared to the wrapper method, the filter approach is computationally faster. However, the filter method tends to select redundant features because it does not consider the relationship between features (Langley et al., 1994). Figure 3.11 illustrate the wrapper and filter feature selection process.



Figure 3.12: The machine learning process in radiation oncology. Previous patient data are fed into a machine learning model in order to deliver a waiting time estimate to new, checked-in patients. After the patient undergoes treatment, their data is supplied to update the machine learning model.

3.4 Modelling Waiting Times in Radiation Oncology

The current challenge is to design a learning algorithm that can adapt to a dynamically changing situation, as is encountered in a busy radiation oncology practice. Modelling waiting times in radiation oncology not only requires previous patient data as experience for learning but also requires putting the proper ML implementation steps into radiation oncology practice. Figure 3.12 describes the basic cycle of the machine learning process in the context of waiting times in radiation oncology. The next two subsections describe how to properly implement machine learning as well as introduce ways in delivering waiting time estimates to patients using modern technology.

3.4.1 Steps to Applying Machine Learning in Clinical Practice

The first step for the application of machine learning in general, and in radiation oncology in particular, is to define the nature of the problem in terms of the input data and the desired outputs. In this dissertation, the problem at hand is modelling waiting times for daily fractionated treatments. A detailed description of what goes into the machine learning algorithm, what comes out, and why is explained in section 4.3. Essentially, what determines a patient's overall wait is the sum of the appointment durations of the preceding patients who have yet to be treated. Therefore, the input data are a set of features that influence a patient's treatment duration and the output is the estimated duration.

The second step is to filter out bad data from the training dataset. Machine learning is robust to noise, however, it should be noted that all models are built on approximations (i.e. empirical solutions). In other words, "All models are wrong; some models are useful" (Box et al., 1987). The quality of the output is determined by the quality of the input. This is known as GIGO (garbage in, garbage out). If bad data are introduced into the system, the output is likely to be incorrect and uninformative.

A major limitation to the acceptance of machine learning is the "blackbox" stigma. Today, there are numerous off-the-shelf machine learning software packages that are readily available without knowing the underlying theory. These algorithms are essentially fed input data and performance metrics can be extracted using simple helper methods provided by the software packages. Although these software packages reduce significant costs in terms of time, resource and ability, most pre-existing machine learning algorithms are unable to provide intuitive interpretation of the learning process (except in the case of decision trees) that could potentially aid clinical practitioners in better understanding their data and the resulting model and implementing real-life changes that might improve practice. This is an undeniable limitation. In this regard, caution should be taken into consideration when implementing pre-existing machine learning packages in health informatics research.

CHAPTER 4 Design and Implementation of a Clinical Machine Learning Solution

Despite the prevalent use of electronic charting and tasking in our radiation oncology department, our clinical database does not lend itself well to research and discovery. There are several technical challenges that need to be addressed to properly design and perform machine learning on our clinical dataset. First, the use of relevant data requires accessibility (i.e. knowing where the data are and how to access them). Second, managing the data requires standardization (i.e. translating clinical information codes into meaningful taxonomies) to neatly categorize similar data. Third, the use of off-the-shelf machine learning software packages requires preprocessing (i.e. mapping the data into a representable format) for proper machine learning implementation.

This chapter describes the methodology of addressing these technical challenges including automating data collection, calculating waiting times in the context of daily radiation treatment appointments, processing the data for learning, and communicating waiting time estimates to patients.

4.1 Data Background

EHR data for this project originated in Aria, a record-and-verify database developed by Varian Medical Systems (Palo Alto, California). Members of the radiation oncology team at the MUHC use Aria for every aspect of day-to-day clinical workflows, including tasking, scheduling, treatment planning, imaging and treating. The Aria database includes a large number of fields corresponding to all records of practice as well as detailed patient histories. Although Aria is an excellent system for recording and managing patient health records, it does not come with any tools to analyze the collected data in aggregate. However, the sheer amount of patient data that is available in Aria does lends itself well to data-mining techniques to extract information and/or patterns and in turn improve the quality and efficiency of treatment. Not all information within Aria were necessary or relevant for the purposes of this current study. A subset of Aria data was used and was extracted into a custom MySQL database, which will be described below.

4.2 Data Collection

To collect data from the Aria database, a custom in-house software was developed, known as AEHRA (Automatic Electronic Health Record Auditing). AEHRA was designed for two reasons: (1) to conduct studies on a centralized registry where information can be pulled from multiple clinical databases and (2) to obviate the need for direct access to clinical database(s) when performing regular (and oftentimes heavy) queries for these studies. To store data, AEHRA uses MySQL, a free, open-source database management system (Widenius and Axmark, 2002). Our centralized database is hosted on a Ubuntu operating system server (ubuntu.com) inside our hospital's intranet with automatic mirroring to an independent fail-over server and nightly backups to a backup server. Figure 4.1 presents a schema of the custom MySQL database known as $AEHRA_db$.

A schematic of the entire AEHRA software is illustrated in figure 4.2. Select data are transferred from Aria to AEHRA_db on a routine basis using a "cron job" scheduler, hosted on our Ubuntu server, that copies across appointments, treatment information, patient demographics, and other relevant information (see AEHRA_db schema below). For the purpose of this study, the cron job was scheduled for nightly data transfer. The exact metadata that were copied across was set using an *Alias* interface. As described below, the



Figure 4.1: The AEHRA database. Most tables are connected to each other through the patient serial number as a foreign key. The database includes individual data for each patient such as gender, date of birth, information about the doctors and machines assigned to the patient, the diagnosis for each treatment course the patient went through, as well as events which are separated into tasks and appointments. Treatment plans and treatment delivery records are also included.



Figure 4.2: A flowchart of the AEHRA architecture. Select data from Aria (and potentially other hospital EMRs) are copied across to our centralized database, AEHRA_db. The times at which transfers occur are determined by the cron job scheduler and set using the Cron Control Interface. The exact metadata that are copied is set using the Alias Interface, in which internal clinical codes are translated and grouped into standard taxonomies. AEHRA is a standalone software in which many types of studies can be created (using the Study Interface) for multi-purpose data analysis.

Alias interface is a webpage that allows users of the software to decide which data are collected into AEHRA_db.

4.2.1 Aliasing

The Alias interface is used to map internal clinical codes into standardized taxonomies to categorize similar data. For example, daily treatment appointments are coded in Aria as: *.EBC Daily Rx*, *.EBC Daily Rx* (*Electron Boost*), *.EBC-Daily Rx* (*TBI*), to name but a few. These codes describe the type of daily treatment appointment but ultimately, they are all daily treatment appointments. An alias is created that lists these internal appointment codes under one standard taxonomy called *Daily Treatment Appointment*. Creating an alias results in a trigger to pull the internal codes (along with associated metadata such as timestamps, statuses, stakeholders, etc.) from Aria and transfer them with appropriate labels to AEHRA_db. By default, not all data



Figure 4.3: Defining waiting in the context of daily radiotherapy appointments. In this example, patient 6 arrives just as patient 2's treatment appointment has finished. Both patient 1's and patient 2's appointments have incurred a combined 10-minute delay at the time of patient 6's arrival. Machine learning is performed to calculate the treatment durations of patients 3, 4, and 5, generating a cumulative duration that may or may not run over into patient 6's appointment. In this example, it has. The total waiting time for patient 6 is thus the time of arrival to the estimated start time inferred from the cumulative preceding patient durations.

from the Aria database are copied across. Only data that have been aliased using the Alias interface are transferred.

4.3 Waiting Time Calculation

It is important to define "waiting" in the context of radiation oncology, specifically for daily radiotherapy treatments. In practice, (unplanned) waiting boils down to the delay between the time a patient is scheduled to start treatment and the time they actually start treatment. This delay for a particular patient is primarily the result of the treatment durations (and duration overruns) of those patients who immediately precede them. Hence, to provide an arriving patient with an estimate of their expected waiting time, we use our machine learning algorithm to predict the treatment *durations* of those preceding patients who have yet to be treated. Figure 4.3 provides a detailed illustration of how a personalized waiting time estimate is derived.

4.4 Learning Process

To predict treatment durations, an off-the-shelf machine learning Python (python.org) package was explored, known as *scikit-learn* (Pedregosa et al., 2011). Scikitlearn, or *sklearn* for short, was designed for non-specialists to use a wide range of state-of-the-art machine learning algorithms for supervised and unsupervised learning problems. Using sklearn for our machine learning problem provided us with two benefits: (1) due its' frequent use and maintenance by the community (with over 700 contributors), it obviated the need to develop learning algorithms from scratch that would otherwise require significant resources, and (2) sklearn readily provides performance metrics and feature selection tools to evaluate models and automatically determine the most important features, respectively.

Having the algorithms already implemented in Python, the preliminary step was to define a set of features (predictors) that influence daily treatment appointment durations. An initial feature set was put together based on feedback from professionals working in the radiation oncology clinic. Practitioners, namely the radiation therapists who directly treat patients, were asked to estimate the duration of a select number of appointments on their schedule and to provide reasons for their estimates. Essentially, they were asked to make a prediction based on their personal experience and to come up with a list of reasons (i.e. features) why appointments would last their predicted response.

Interestingly, the radiation therapists came up with a few common predictors. If the patient was treated for the first time, they concluded that the main contribution to treatment time was the treatment setup. In order to setup the patient for treatment, the correct positioning and the condition of the patient, in terms of the severity of the disease, must be taken into consideration. This

Feature	Data Type
Diagnosis	Categorical
Physician	Categorical
Course	Categorical
Machine	Categorical
Gender	Categorical
Age	Continuous
Appointment Day of the Week	Categorical
Appointment Hour	Categorical
Appointment Month	Categorical
Plan	Categorical
Body Orientation on Treatment Table	Categorical
Radiation Type	Categorical
Number of Treatment Beams	Continuous
Previous Duration	Continuous
Number of Images Taken	Categorical
Imaging Duration	Continuous
Radiation Therapist	Categorical
Total Monitor Units	Continuous
Total Monitor Units Coefficient	Continuous
Fraction Number	Categorical
Allocated Appointment Time	Categorical
Median Past Durations	Continuous

 Table 4.1: List of features extracted for each input sample set. Features were either categorical as string variables or real valued (continuous).

suggested that treatment positions and patient diagnoses were potential factors in the overall treatment duration. Furthermore, if the patient had been treated more than once, then the radiation therapists felt more comfortable in their estimates since they were already familiar with the patient. In other words, their estimates were mainly based on how long the treatment took last time. The main purpose of machine learning is to emulate human thought. Defining an initial feature set based on the real-life experience of radiation therapists can potentially guide the machine learning algorithm to accomplish its task. Table 4.1 presents the proposed list of features used for this study.

With a preliminary feature set defined, the next step was to extract these features from AEHRA_db for learning. This is detailed in the next subsection.

4.4.1 Data Extraction

Retrospective data, including the feature set, along with the corresponding appointment durations, were extracted from AEHRA_db on a nightly basis for nightly training using a Python script. This was done by: (1) performing a relationship (also called an inner join) between the *Patient*, *Diagnosis*, Doctor, Resource, Appointment, Course, Plan, Radiation and RadiationHstry tables by the keys that those tables have in common (see schema in figure 4.1), and (2) building a Python *list* (i.e. an array) for each retrieved record. Aria and, in turn, AEHRA_db, do not provide a direct connection between appointments and treatment information (Appointment and RadiationHstry tables, respectively), which was pertinent in extracting the proper dataset. The Appointment table contains information about the start and end times of the appointment, status of the appointment, etc. The RadiationHstry table contains information about the start and end times of the treatment, fraction number, monitor units, etc. One can reach either table starting from the Patient table however there is no key linking the two tables directly. These tables were linked according to the dates patients were treated. In other words, the dates of Appointment. Scheduled Start Time and Radiation Hstry. Treatment Start Time provided the connection between each patient's scheduled appointment and their treatment records.

The *physician* and *radiation therapist* features were defined as being their primary radiation oncologist and primary radiation therapist, as registered in the database, respectively. In other words, it may be the case that multiple physicians and therapists treat an individual patient but the precise information on the degree of collaboration is not well defined in the database. Thus, the primary metadata were used.

The previous duration and median past duration features were not explicitly recorded in the database. Gathering these data required additional subqueries and helper functions to order the appointments by date (for the

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1M} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \dots & x_{NM} \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Figure 4.4: Preprocessing the dataset using sklearn for learning required constructing a training input matrix of N samples and M features and a separate matrix of N target values.

previous duration feature) and to calculate the median (for the median past duration feature).

After extraction and careful construction of the retrospective dataset, additional steps were required in order to properly represent the data for machine learning using sklearn. The next subsection describes the preprocessing step.

4.4.2 Preprocessing

To facilitate the use of our retrospective dataset set with sklearn, several data formatting conventions needed to be established. First, the input data and the labelled outputs were structured as separate matrices. In other words, the input dataset was constructed as an $M \ge N$ matrix, where M is the number of features and N is the number of treatment appointment records. The labelled output data (i.e. the treatment appointment durations) were put into a separate $N \ge 1$ matrix (essentially into an array). Figure 4.4 demonstrates this process.

Next, all data types must be converted to floating point (fractional) numbers. Integer values can easily be converted to floats by assigning a decimal point (e.g. 1 to 1.0). Categorical features, typically in the format of strings, must be mapped in one of two ways depending on the machine learning model used. One way to map categorical features is to use *one-hot encoding*. One-hot encoding transforms a single feature variable with n distinct values to n binary variables, where a value of "1" is assigned to the appropriate column and zero

$$"dog" = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, "cat" = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, "mouse" = \begin{bmatrix} 0\\0\\1\\0 \end{bmatrix}, "bird" = \begin{bmatrix} 0\\0\\1\\0 \end{bmatrix}$$

Figure 4.5: Example of one-hot encoding. In this example, there are four distinct values for one feature of "animals." Therefore, four binary variables are used where a value of "1" is assigned to the appropriate value and zero otherwise. Note that in this example, the feature space increases by a factor of four.

otherwise. An example of one-hot encoding is illustrated in figure 4.5. There is a trade-off to using one-hot encoding. Although each feature can be defined as one of two states (on or off), the feature space itself is extended which can be costly to server memory and CPU run-time performance.

Another method for mapping categorical data is to assign a numerical value to each category. For example, [house, dog, car] can simply be converted to [0,1,2]. Although this is a simple mapping that contains the same cardinality as the original set, this method imparts an ordinal property to the variables (i.e. house < dog < car) which may or may not make sense. Some learning algorithms (for example decision trees) can still operate independent of this property. However, for most learning algorithms, this method may result in inaccurate machine learning.

Once the retrospective data were sampled and formatting conditions were established, various sklearn machine learning models were implemented. This is described in the next subsection.

4.4.3 Machine Learning Models

The goal of model selection was to select the best model and output the best hypothesis from that model. Specifically, the best model should have the lowest MAE, MedAE, and MSE, while delivering the highest R^2 score. To determine these performance metrics and validate the models, a subset of the



Figure 4.6: Sklearn's "cheat sheet" for choosing the right algorithm(s). This study involved predicting daily treatment appointment durations. In other words, the outputs were real-valued. Thus, various supervised regression models (light blue section) were used in this study. Image taken from http://scikit-learn.org/.

retrospective dataset was held-out for testing. The random sampling method was used in a 80%/20% fashion representing the training and testing sets respectively. With this partitioned dataset, various sklearn machine learning models were applied to learn on the training set, were verified on the testing set, and the best model was chosen. Figure 4.6 was used as reference to choose the right sklearn regression models for our data.

4.4.4 Feature Selection

Based on the best model chosen, sklearn helper methods were used to determine the most significant features and to improve the model's accuracy. By default, sklearn removes all zero-variance features (i.e features that have the same value in all samples). Furthermore, features were considered unimportant and removed if the corresponding *feature_importance* helper method returned values that were below a provided *threshold* parameter. The threshold was initially chosen arbitrarily because ultimately, the *feature_importance* method returned relative percentages. In other words, features were assigned a percentage of importance relative to all feature samples.

4.5 Communicating Waiting Time Predictions

In an effort to ultimately provide waiting time estimates to patients, two novel communication routes were developed in parallel to the machine-learning research: (1) a patient portal and smartphone application called *Opal* and (2) a waiting time simulator to predict waiting times in real time.

4.5.1 Opal — The Oncology Portal and Application

Opal is a smartphone app and patient portal that was developed for radiation oncology patients at our cancer centre (submitted for publication: Physics in Medicine and Biology, August 2016). It is a novel communication tool that aims to empower patients with their personal health information, including their waiting time estimates. Opal was initially developed from inside our hospital system to automatically personalize the provision of information to patients according to diagnosis and stage of treatment using each patient's data within the hospital's electronic medical record. Although this dissertation solely focusses on addressing daily treatment appointment waiting times, Opal is the main avenue in which waiting times will be communicated to patients at our centre. Figure 4.7 illustrates Opal's architecture.

4.5.2 Waiting Time Delivery Simulator

To monitor and evaluate the machine learning model's performance on prospective data, a live web-based waiting time delivery simulator was developed. Effectively, this software simulates the delivery of waiting time estimates to all



Figure 4.7: An overview of Opal's communication architecture. Select data from Aria and other hospital EMRs are transferred to a centralized database known as *OpalDB*. Data transfer occurs using a "cron job" Auto Update scheduler, that copies across approved appointments, tasks, documents, and lab test results. Approval and timing of data transfer are determined by a list of "publishing rules" that are set using a *Publish Manager* web interface. Secure serving of data through the hospital's firewall to Opal is facilitated by *Firebase*, a cloud database operated by Google Inc. (firebase.com). Firebase it designed such that all applications connected to it are served data in real time. Data is encrypted once before sending to Firebase and encrypted again by Firebase's own encryption protocols. Data ultimately reaches out to the patient in the outside world on their mobile phone or web browser. A demo of Opal can be found online at depdocs.com/opal.

patients scheduled to be treated on the current date. Every morning, appointment data from AEHRA_db are used by the trained model to get duration estimates for each patient to be treated that day. These estimates are stored back in AEHRA_db to be used by the simulator. Every 30 seconds during clinical hours, the simulator probes AEHRA_db for any changes in the database and constructs queues based on patients who are checked-in on the same treatment machine. Depending on the patient's position in the queue, a waiting time estimate is printed and logged, along with the number of patients in the queue and the overall schedule status of the clinic. A detailed version of figure 4.3 is presented in figure 4.8, providing a more illustrative description of the simulated delivery process.

The possible time block estimates that can be delivered are: < 10 minutes, 10-20 minutes, 20-30 minutes, 30-45 minutes, 45-60 minutes, and > 60minutes. These ranges were chosen to account for the errors in the summation of treatment duration predictions (i.e. the more patients in the queue, the larger the error in the total waiting time estimate). Most treatments at our centre typically last 15 minutes and thus in most cases, a long waiting time estimate would be the result of several patients in the queue. However, in some cases, a single patient can be treated for a very long with a lower duration estimate uncertainty than several shorter length treatments in a queue.

As mentioned above, the waiting time simulator logs estimates every 30 seconds. Therefore, timeline plots were implemented in this software to monitor logs for each patient in real time. Figure 4.9 presents a timeline log of the number of scheduled and actual patients queued before a selected patient. Figure 4.10 shows a log of the estimated start time over time for a particular patient. Figure 4.11 presents a log of the clinical schedule over time for a selected patient.



Figure 4.8: A magnified and detailed version of figure 4.3 describing the waiting time delivery simulation process. In this example, patient 3 has started treatment. Patients 4, 5, and 6 have yet to be treated on the same treatment machine. For each pending patient, the simulator prints out a live time block estimate (light blue) based on the sum of predicted durations (yellow) of preceding patients, along with an indication on the amount of patients in the queue and the status of clinical overruns by the time their respective treatments start. For instance, patient 4 is after patient 3. The waiting time delivery simulator prints a waiting time estimate range of "10-20 mins" since the actual wait is $15 \text{ mins} \pm error$. Patient 5 is second in the queue. Therefore, the simulator prints a waiting time estimate range of "30-45 mins" since the actual wait is $(15 + 15) = 30 \text{ mins} \pm error$, etc.


Figure 4.9: A screenshot of a timeline log for the number of scheduled (black) and actual (blue) patients in the queue for a selected patient. The beginning of the timeline log represents the time the patient checkedin. In this example, when the patient checked-in there was one patient already in the queue. Over time, the scheduled patients arrived, checked-in, and the number of actual patients in the queue grew to match the scheduled number of patients. As the preceding patients underwent treatment, the number of pending patients decreased. The end of the timeline log represents the time the patient was called in for treatment. In this screenshot, the patient was called in before a preceding patient finished treatment. The vertical line represents the scheduled start time for this particular patient.



Figure 4.10: A screenshot of an estimated start timeline log for a selected patient. The beginning of the timeline log represents the time the patient checked-in and the end of the timeline log represents the time the patient was called for treatment. Red represents behind schedule and green represents ahead of schedule. In this screenshot, preceding patients were taking longer than their allocated scheduled time. Thus, the patient's estimated start was pushed over time. The information displayed in the blue box (upper right) represents details about the data point at the end of the timeline. When the patient was called in for treatment, a waiting time estimate of "< 10 minutes" was logged.



Figure 4.11: A screenshot of the clinical schedule timeline log for a selected patient. The beginning of the timeline log represents the time the patient checked-in and the end of the timeline log represents the time the patient was called for treatment. Red represents behind schedule and green represents ahead of schedule. Over time, preceding patients were taking longer than their allocated scheduled time. Thus, by the time the patient was called in for treatment, the clinic was 25 minutes behind schedule. At the end of the day, the machine learning model's performance is evaluated according to the standard performance metrics (MAE, MedAE, R^2) on the predicted durations. This essentially evaluates the generalizability of the machine learning model on prospective (i.e. unseen) data.

Note that the main purpose of this software was to *simulate* the delivery of waiting time estimates to patients. Ultimately, the real communication of estimates will be delivered via Opal. This simulator was developed as a helper tool for monitoring and evaluating real-time estimates on a large scale (i.e. as if every patient had the Opal app).

CHAPTER 5 Results–Evaluation of Waiting Time Estimates

5.1 Prediction Models

Four machine learning regression models from the sklearn python package were considered in this supervised learning problem: (1) linear regression, (2) support vector machine (SVM), (3) decision tree, and (4) random forest. Each model was fit to a training set and evaluation of its performance was done on a testing set. The training set was constructed by randomly sampling 80% of the whole retrospective dataset, while the remaining 20% was used for testing. Of the 50840 retrospective appointment data samples initially extracted from the AEHRA_db database, only 44225 of those were used due to missing data for one or more of the proposed features. Depending on the applied regression model, categorical features were preprocessed either using one-hot encoding or by mapping each distinct category to a numerical value. Continuous features were left as is.

Table 5.1 displays a comparison of the sklearn regression models used along with their performance metrics and the type of preprocessing implemented. The accuracy of the regression models was compared by calculating the mean and median absolute errors of the prediction from the actual value as well as the standard deviation of these errors and the R^2 value. The mean squared error was not considered due to the sufficient evaluation provided by the mean absolute error.

The best trained model, according the evaluations done on the testing set, was found to be the *random forest regression* (RFR) model. A residual (i.e.

Regression Model	Preprocessing Method	Mean Absolute	Median Absolute	Standard De- viation Error	\mathbb{R}^2
		Error	Error		
		[min]	[min]	[min]	
Linear regression	One-hot encod-	4.9	3.9	6.4	0.25
CI VI I	One hat arread	4.0	2.4	7.0	0.19
S V IVI	ing	4.9	3.4	1.2	0.15
Decision tree	Numerical cate- gorization	4.7	3.3	6.8	0.42
Random forest	Numerical cate- gorization	4.6	3.3	6.1	0.47

 Table 5.1: Performance evaluation of each machine learning regression model used for this study. The standard deviation error is the standard deviation of the calculated residuals.

Feature	Relative Importance
Allocated Appointment Time	40%
Fraction Number	24%
Median Past Duration	17%
Number of Treatment Beams	6%
Previous Duration	5%
Others	8%

Table 5.2: Top five features as defined by sklearn's *feature_importance* methodusing the random forest regression model. Other proposed featurescorresponded to a total remainder importance of 8%.

error) histogram is shown in figure 5.1. The mean absolute error was found to be 4.6 minutes with a standard deviation of 6.1 minutes. The median absolute error was found to be 3.3 minutes and the R^2 value was determined to be 0.47. According to the RFR model, only five out of the 22 features initially extracted were determined to be most influential. The relative importance of each feature was determined using sklearn's *feature_importance* method. A summary of the top five features used, along with their relative feature importance, is displayed in table 5.2.

5.2 Feature Trends

In an attempt to describe top predictors for daily treatment appointment durations, correlation plots were obtained for the features used in this study.

Figure 5.2 displays average treatment duration as a function of diagnosis. Average treatment durations according to appointment day of the week and



Figure 5.1: A residual plot histogram for the performance of the random forest regression model on the testing set.

appointment hour are depicted in figures 5.3 and 5.4, respectively. Trends for fraction number and allocated appointment time are plotted in figures 5.5 and 5.6, respectively. The plot of appointment duration against the median past duration is displayed in figure 5.7.

5.2.1 Imaging-Phase Study

A decision was made to sample post-2015 retrospective data for learning due to a change in practice that followed a relocation of our radiation oncology department in early 2015. New treatment machines were installed at our new facility, incorporating state-of-the-art equipment for both treatment and imaging. As was described in section 1.4, imaging is an essential part of the treatment process to ensure proper patient setup and to monitor any anatomy shifts that may have occurred since last treatment.

Before the relocation (pre-2015), imaging was typically performed on a weekly basis. Thus, during the conduction of the research, when *all* preand post-2015 retrospective data were included for learning, it was found that



Figure 5.2: Plot of average treatment duration versus diagnosis for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion).

imaging was a significant feature on the overall treatment time. That is, imaging caused variance on the treatment durations. On the days that imaging occurred, time was spent by radiation therapists imaging the patient, causing an added time to the overall treatment duration compared to days when imaging did not occur.

After the relocation (post-2015), new protocols required that imaging be done daily. Thus, imaging no longer became a significant feature due to its matching frequency with daily treatment durations. Figures 5.8 and 5.9 show the frequency of imaging for all patients according to their diagnosis and fraction number.

5.3 Comparison Against Simple Averages

The best performing machine learning algorithm (random forest regression model) was compared to simple averages of several categorical features. For



Figure 5.3: Plot of average treatment duration versus treatment appointment day of the week for all retrospective data (1 = Sunday, 7 = Saturday). Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion).



Figure 5.4: Plot of average treatment duration versus treatment appointment hour for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to single value data points.



Figure 5.5: Plot of average treatment duration versus fraction number for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to a single value data points.



Figure 5.6: Plot of average treatment duration versus allocated treatment appointment time for all retrospective data. Error bars represent the standard deviation of each categorical distribution (i.e. the dispersion). Missing error bars are due to a single value data points.



Figure 5.7: Plot of treatment duration versus median past duration for all retrospective data. The red line represents a linear regression of all data points.



Figure 5.8: Histograms of imaging occurrence as a function of fraction number per diagnosis (pre-2015). The y-axis represents the frequency of imaging and the x-axis represents the fraction number. Note that every patient has a first fraction (i.e. treated for the first time) and imaging is always done on the first fraction. The frequency at which imaging occurs in later fractions decreases due to the decreasing number of patients that are treated at higher fractions.



Figure 5.9: Histograms of imaging occurrence as a function of fraction number per diagnosis (post-2015). The y-axis represents the frequency of imaging and the x-axis represents the fraction number. Note that every patient has a first fraction (i.e. treated for the first time) and imaging is always done on the first fraction. The frequency at which imaging occurs in later fractions decreases due to the decreasing number of patients that are treated at higher fractions.

example, the mean treatment duration of all patients treated for breast cancer was used to predict the treatment duration of future patients with breast cancer. In other words, for each categorical feature, the mean of each category was used to predict treatment appointment durations. Table 5.3 presents performance metrics of several categorical features, individually used as predictors.

5.4 Results from Waiting Time Simulator

In order to evaluate performance on real, unseen data, the RFR model was executed on live appointment data using the waiting time simulator software. Every 30 seconds, during regular treatment hours, waiting time estimates were generated for all patients waiting for treatment (i.e. checked-in patients). After every execution, the software logged these waiting time estimates along with additional information regarding the number of patients in the queue and

Predictor	Mean Absolute	Median Absolute	Standard De- viation Error	\mathbb{R}^2
	Error	Error		
	[min]	[min]	[min]	
Allocated Appointment Time	5.4	4.2	7.5	0.18
Diagnosis	5.8	4.6	8.1	0.05
Physician	5.9	4.8	8.0	0.02
Radiation Therapist	5.9	4.8	8.3	0.02
Appointment Day of the Week	6.1	5.0	8.5	0.002
Appointment Hour	6.0	5.2	8.2	0.002
Fraction Number	5.6	4.4	7.9	0.10
Number of Treatment Beams	5.7	4.6	7.9	0.11
RFR Model	4.6	3.3	6.1	0.47

Table 5.3: Evaluation and comparison of the random forest regression (RFR) model against simple averages. For each predictor listed here, the average of each distinct category was used to predict treatment appointment durations. The standard deviation error is the standard deviation of the residual (predicted minus actual).

Regression Model	Mean Absolute Error [min]	Median Absolute Error [min]	Standard De- viation Error	\mathbb{R}^2
	[]	[]	[]	

Table 5.4: Performance evaluation of the random forest regression model on
prospective data. Performance metrics were collected on a daily ba-
sis, between June 2016 and October 2016. The average and standard
deviation were taken for each metric.

the clinical schedule (see section 4.5.2). At the end of each treatment day, the regression performance metrics were calculated and recorded into the database. Table 5.4 provides a summary statistic of the RFR model's performance on prospective data since the waiting time delivery simulator software's launch in June 2016.

Another evaluation that was performed using the waiting time delivery simulator software was capturing the logged information that was recorded the moment patients were called in for treatment. Waiting time estimates logged moments before the start of treatment was essential in determining what could have been delivered to patients had the RFR model was used for predictions. Figure 5.10 plots the waiting time estimates that were simulated to checked-in patients prior to treatment.





Figure 5.10: Pie chart showing the estimate details that were delivered to checkedin patients immediately prior to treatment.

5.5 Prediction Provision

As was mentioned in section 4.5.1, the development of a successful in-house technology paved the way to deliver waiting time estimates. This section describes the smartphone app (Opal) that will be released in open-beta (late November, 2016) to patients at our treatment facility.

5.5.1 Opal

The app's Home screen, figure 5.11, provides the patient with easy access to waiting time communications. On the day of the patient's next appointment, a "Check-in for your appointment" button is provided. It is only enabled when the patient is geo-located within 200 m of the hospital, as determined by the phone's GPS. Once checked-in, the patient is provided with a waiting time estimate as well as information pertaining to the busyness of the clinic and the number of patients ahead of them (figure 5.11).



Figure 5.11: Screenshots of waiting time communications in Opal. On the Home screen (left), a patient is presented with a summary of their appointment(s). On the day of the appointment, a check-in button is provided (center) and is enabled only when the patient is geo-located to within 200 m of the hospital. Once checked-in (right), the patient is provided with a waiting time estimate, as well as information regarding the number of preceding patients and the busyness of the clinic, updated on a live basis.

CHAPTER 6 Discussion

Providing patients with estimates of their overall waiting time is of current interest in healthcare. In radiation oncology, however, no major successes have been reported to date and only rough estimates are communicated to patients (with very little confidence) at our comprehensive cancer centre. With the abundance and availability of patient EHR data at our centre, one can meet the challenge using machine learning. The results of our machine learning model, applied to daily radiation treatment durations, provide better waiting time estimates than simple averages and rough estimates. However, the performance metrics on the random forest regression model suggest room for improvement, as described below.

6.1 Model Performance

The random forest regression (RFR) model, as applied to our data, tends to underestimate treatment durations. This may be due to bad data in the sampled dataset. The appointment start and end timestamps are recorded in the database based on the opening and closing of patient electronic charts, respectively, on the treatment console. Occasionally, charts are left open before a pause in practice (e.g. during lunch break or in between shifts), or they are opened before the next patient enters the treatment room, or they are left open on another treatment console, making treatment durations appear longer than reality. Conversely, patient charts are occasionally opened *after* the patient is fully set up on the treatment table rather than when the patient *enters* the treatment room. This means that treatment durations are recorded as shorter than reality. However, from the data, the former occurs more frequently than the latter, albeit sporadically, but not frequent enough to affect predictions. In other words, the RFR's model's responds poorly to noise in the case of longer-than-reality treatment durations (figure 5.1).

Another source of underestimation in the model is how appointments are completed in the database when more than one time slot is allocated to a single patient. Occasionally, patients are booked in multiple time slots for billing purposes. For example, a patient can be booked for two 30-minute blocks instead of one hour block. In theory, these time slots can be predicted individually (even for the same patient), but in practice, most of the time, only one appointment time slot is recorded as completed in the database on behalf of all time slots for the same patient. In other words, the other appointment time slots are left uncompleted (i.e. no end timestamp in the database) and therefore unusable for machine learning. If a patient is indeed booked over multiple time slots and remains in the same treatment room, there is very little incentive for the radiation therapists to close the patient's chart (which would record a completed timestamp for the first appointment time slot) and reopen the same chart for the second appointment time slot (which would record a start timestamp for the second appointment time slot). Thus, the patient's chart is usually left open under the first appointment time slot until the overall treatment is complete. A completed timestamp is recorded for this first appointment time slot (and not the second) resulting in a seemingly long appointment duration compared to its allocated time slot. This leads to an underestimation in the predicted treatment durations since allocated appointment time is the most significant predictor for treatment durations.

In the context of prospective data, the RFR model only slightly underperforms on unseen data compared to retrospective testing data. One of the challenges in machine learning is to avoid overfitting and to generalize into unseen data. A retrospective testing set is a close representation of unseen data. However, it may or may not reflect the true, unseen data. Our RFR model demonstrates similar accuracies on real-time data compared to retrospective testing data (table 5.4). Thus, we can conclude that it is able to generalize well.

6.2 Feature Set

The top features of the random forest regression model agree well with human intuition. As a first guess, since appointments are scheduled at our centre, they tend to last their allocated appointment time (figure 5.6). The majority of treatment appointments at our cancer centre are booked under 15, 30, or 45 minute time slots. These time slots are chosen based on the personal experience of the booking staff. However, data suggest there is room to book shorter appointment times for longer allocated appointments.

The next important predictor is fraction number. The first fraction tends to take the longest because the radiation therapists must familiarize themselves with the patient and the treatment setup. As daily fractionated treatments go on, the therapists become more comfortable with the patient and can readily setup and treat the patient in a shorter amount of time compared to the first treatment. This is reflected in the data (figure 5.5) and matches the intuition made by the radiation therapists when asked about an initial feature set (see section 4.4). The first two significant features (allocated appointment and fraction number) are of great importance when a patient undergoes treatment for the first time. However, after the first few fractions, the third most important feature (median past duration) becomes increasingly relevant.

The median past duration is a useful determinant in predicting treatment appointment durations due to the fact that treatments tend to last around their historical times (figure 5.7), regardless of the specific treatment protocol. In other words, regardless of other features specific to the patient and their treatment, such as diagnosis (figure 5.2), treatment machine, treatment technique, these features can be encapsulated (see table 5.2) by previous durations since the patient is treated in the exact manner as previous treatments. The *median* past duration is taken instead of the *mean* past duration because, statistically, medians are pulled less by outliers than averages.

The appointment time and appointment day of the week have no significant effect on the overall treatment time (figures 5.3 and 5.4).

As mentioned in section 5.2.1, imaging was initially taken as a feature. However due to a recent change in practice, it is no longer a significant predictor. Before the relocation (pre-2015), imaging was performed either on every fifth fraction (1, 5, 10, 15, etc.) or performed after every five fractions (1, 6, 11, 16, etc.). This can be seen in the data on several plots representing several diagnoses (figure 5.8). State-of-the-art treatment and imaging machines are used at our new facility, delivering better image quality. Thus, imaging is done before treatment on a daily basis for most diagnoses (figure 5.9). Essentially, this suggests that a constant is added to every treatment time, due to daily imaging, resulting in an invariable (and therefore insignificant) feature.

6.3 Real-Time Delivery of Waiting Time Estimates

Waiting time estimates from this study have not yet been actually delivered to patients at our radiation oncology department. Opal, the oncology portal and application, is one way in which waiting time estimates will be communicated to patients. Predicting waiting times using machine learning and delivering them on this unique platform is one way in which the stress due to the uncertainty of waiting times may be reduced. There is opportunity for patients to grab a coffee at the nearest cafeteria or go to the restroom, if waiting time estimates are to be delivered in this fashion – patients can be called in for treatment remotely on their phone using Opal.

It is undetermined whether estimating waiting times can *reduce* waiting times overall. At our radiation oncology facility, appointments are scheduled. Therefore, there is an inherent expectation of a definite wait until one's scheduled start time. The main concern to patients is whether their appointments start on time or not; and if not, a realistic estimate and an explanation on a delay (if it occurs) is vital to their overall satisfaction and health. In this study, waiting time estimates were recorded for checked-in patients only (i.e. patients who were already present in the waiting room). It would be more valuable to deliver estimates to all patients (checked-in or not) so that even patients at home (or generally outside the hospital) can decide when to arrive to their appointment with little time to wait in the waiting room.

Based on the analysis that was performed to predict waiting times in real-time using the waiting time delivery simulator, there is a great deal of investigation that should be done to ensure patients receive the most accurate waiting time estimates. It was recognized that multiple patients waiting in the waiting room are called in by radiation therapists (at the same time) to wait in another area (closer to the treatment room) and not necessarily called in for immediate treatment. This does not affect treatment duration estimates. However, this does cause a discrepancy in our simulated delivery of waiting time estimates (figure 5.10) because there is nothing capturing the event that patients may or may not be waiting in this "buffer" area before treatment.

It was also recognized that patients are sometimes called in before others due to their immediate availability in the waiting room. For example, if patient X is scheduled before patient Y, but patient X is checked-in but unavailable and patient Y is waiting in the waiting room, then patient Y would likely be called in, due to the immediate absence of patient X. Waiting time estimates should be delivered in a proper queue-like fashion based on the number of patients who have yet to be treated *in order* of their scheduled time. Perhaps with a way to call patients remotely (i.e. using Opal) there can be a more ordered call-in and patients would not have to worry about being skipped/missed.

Further to the discussion on waiting time delivery using the waiting time delivery simulator, the majority of patients received (note: by simulation) "< 10 minutes left" and "10-20 minutes left" as their waiting time estimate the moment they were called in by staff. This analysis, of course, is incomplete. Whether a waiting time estimate is provided when they are called, or while they wait, or when they check-in, or (favourably) when they are at home, all patients should receive an estimate that reflects the time they will be treated, regardless of when the estimate is provided. However, due to time limitations of this research, the waiting time delivery simulation was evaluated only at the time a patient was called in for treatment (figure 5.10). Of course, in this case, all patients should receive "< 10 minutes left."

CHAPTER 7 Conclusions and Future Work

This research project involved developing a computational solution to address the pain of waiting in radiation oncology at our comprehensive cancer centre. The study investigated waiting times in terms of daily radiation treatment appointments, where patients wait in the waiting room for their treatments on a daily basis. Although treatment appointments are scheduled at our centre, appointments may start before or after their scheduled time. What is important when calculating waiting times is to predict the *duration* of treatments for patients who precede a current, checked-in patient. In other words, machine learning was implemented to predict treatment durations in order to infer an overall waiting time.

Several off-the-shelf machine learning algorithms from the sklearn python package were studied in this project and were found to produce varied performance metrics. The random forest regression model was found to be the best performing algorithm. On average, we can estimate true treatment durations to within 7.2 minutes. This is a significant improvement on the rough estimates usually given to patients, and also on the estimates one would obtain by simple averages of all patients with certain features. These initial predictions are a promising start, but there is much work to be done to improve our model. Although the accuracy of the model was improved through feature selection and feature importance (i.e. the inputs of the system), the focus should be on the labeled outputs (i.e. the true treatment durations) when training the model. There is potentially a lot of inaccurate data in the database that add variance to the set of data to be trained on. For example, the start and end timestamps recorded in the database for each treatment appointment are triggered on the time a patient's chart was opened and closed on the treatment console, respectively. Occasionally, patient charts are left open and recorded in the database as appointments that last longer than reality. It would be helpful to define further rules, perhaps even a slight change in practice, to make better use of this data.

The next step is to test and evaluate Opal, the smartphone app, that will be in the hands of our radiation oncology patients by the end of autumn 2016. Opal was developed as a parallel project to automatically deliver personal health information to patients as well as to incorporate waiting time estimates. Waiting time information was reported as one of the most needed features in the app. Therefore, future reports should (a) investigate the impact of these waiting time estimates on clinical workflow, (b) conduct patient surveys about these waiting time estimates and monitor patient satisfaction, and (c) determine the costs and benefits of implementing a fully automated and personalized system.

If waiting is experienced and expected virtually anywhere, the problem therein lies in the uncertainty of waiting. Waiting is a symptom of delayed action caused by a preceding chain of events. Thus, it is worth evaluating the lengths of these events (i.e. durations) rather than confronting waiting times directly. This work in radiation oncology, if successful, might lead to similar work in other specialties, including chemotherapy, surgery, and even emergency departments.

REFERENCES

- Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T.: 2012, Learning from data, Vol. 4, AMLBook Singapore
- Box, G. E., Draper, N. R., et al.: 1987, Empirical model-building and response surfaces, Vol. 424, Wiley New York
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G.: 2013, in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp 108–122
- Canadian Cancer Society: 2015, Canadian Cancer Statistics 2015, Canadian Cancer Society
- Chan, K., Li, W., Medlam, G., Higgins, J., Bolderston, A., Yi, Q., and Wenz,
 J.: 2010, Journal of Medical Imaging and Radiation Sciences 41(3), 145
- CIHI: 2012, Health Care in Canada, 2012: A Focus on Wait Times
- CIHI: 2013, Better Information for Improved Health: A Vision for Health System Use of Data in Canada
- El Naqa, I., Li, R., and Murphy, M.: 2015, *Machine Learning in Radiation* Oncology: Theory and Applications, Springer International Publishing
- Famiglietti, R. M., Neal, E. C., Edwards, T. J., Allen, P. K., and Buchholz, T. A.: 2013, International Journal of Radiation Oncology* Biology* Physics 87(1), 148
- Fortin, B., Goldberg, M. S., Mayo, N. E., Valois, M.-F., Scott, S. C., and Hanley, J.: 2006, *Healthcare Policy* 1(2)

Fraser Institute: 2013, Reducing Wait Times for Health Care

- Guttmann, A., Schull, M. J., Vermeulen, M. J., and Stukel, T. A.: 2011, *Bmj* 342, d2983
- Guyon, I. and Elisseeff, A.: 2003, *Journal of machine learning research* **3(Mar)**, 1157
- Hodge, W., Horsley, T., Albiani, D., Baryla, J., Belliveau, M., Buhrmann,
 R., O'Connor, M., Blair, J., and Lowcock, E.: 2007, *Canadian Medical Association Journal* 176(9), 1285
- Jack, R. H., Davies, E. A., Robinson, D., Sainsbury, R., and Møller, H.: 2007, BMC Cancer 7(1), 1
- Jones, D.: 1994, *Medical physics* **21(6)**, 833
- Kulkarni, G. S., Urbach, D. R., Austin, P. C., Fleshner, N. E., and Laupacis,A.: 2009, *The Journal of urology* 182(4), 1318
- Langley, P. et al.: 1994, in Proceedings of the AAAI Fall symposium on relevance, Vol. 184, pp 245–271
- Macdonald, F., Ford, C., and Casson, A.: 2004, *Molecular Biology of Cancer*, Advanced Texts, Taylor & Francis
- Paul, C., Carey, M., Anderson, A., Mackenzie, L., SANSON-FISHER, R., Courtney, R., and CLINTON-MCHARG, T.: 2012, European journal of cancer care 21(3), 321
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: 2011, *Journal of Machine Learning Research* 12(Oct), 2825
- Robinson, D., Massey, T., Davies, E., Jack, R., Sehgal, A., and Møller, H.: 2005, *British journal of cancer* **92(7)**, 1201
- Sobolev, B. G., Fradet, G., Hayden, R., Kuramoto, L., Levy, A. R., and FitzGerald, M. J.: 2008, *BMC health services research* 8(1), 1

- Washington, C. and Leaver, D.: 2015, Principles and Practice of Radiation Therapy, Elsevier - Health Sciences Division
- Widenius, M. and Axmark, D.: 2002, MySQL reference manual: documentation from the source, "O'Reilly Media, Inc."
- Wright, J. G. and Menaker, R. J.: 2011, Canadian Medical Association Journal 183(9), E559