2017 16th IEEE International Conference on Machine Learning and Applications

Predicting waiting times in Radiation Oncology using machine learning

A. Joseph, T. Hijal, J. Kildea
Department of Radiation Oncology
McGill University Health Centre
Montreal, Canada

L. Hendren, D. Herrera
Department of Computer Science
McGill University
Montreal, Canada

***Abstract*— We describe a method for predicting waiting times in radiation oncology using machine learning. The patient waiting experience remains one of the most vexing challenges facing healthcare. At our comprehensive cancer centre, waiting periods that arise throughout a patient's course of treatment are generally difficult for staff to predict and only rough estimates are typically provided based on personal experience. To the patient, waiting times feel long and are seemingly unpredictable. Delays for treatment at our centre depend on the durations of preceding patients scheduled in the queue. To that end, we have incorporated the treatment records of all previously-treated patients into a machine learning framework in order to predict treatment durations to infer an overall waiting time. We found that the *Random Forest Regression* model provides the best predictions for daily fractionated radiotherapy treatment durations. Using this model, we achieved a median residual (actual minus predicted duration) of 0.25 minutes and a standard deviation residual of 6.1 minutes to retrospective treatment data. Waiting times are derived by summing the predicted durations. The main features that generated the best fit model (from most to least significant) are: Allocated appointment time, radiotherapy fraction number, median past duration of treatments, the number of treatment fields, and previous treatment duration.**

Keywords—radiation oncology; machine learning; waiting times

I. INTRODUCTION

Waiting time uncertainty is a universal challenge that can occur anywhere, and the healthcare system is no exception. In healthcare, waiting time uncertainty can cause patients, who are already sick and in pain, to worry about *when* they will receive the care they need [1]. In the context of radiation oncology, the most recurrent waiting period patients encounter is awaiting daily fractionated treatment. This waiting period occurs at the hospital in the waiting room, after the patient has been called to start radiotherapy treatments. The patient arrives at the clinic on a daily basis (for several weeks) to receive a portion of their total treatment (known as fractionated treatment) and may wait between minutes to hours after arrival. Although treatment appointments are scheduled at our clinic, delays for treatment 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 also 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.

Waiting times in radiation oncology have been extensively analyzed for the purpose of understanding and explaining the factors that affect delays in order to decrease waiting times for radiation therapy patients [2][3][4]. These studies focus on waiting time variations between pre-treatment events (such as initial consultation, diagnosis, surgery or chemotherapy) to the start of treatment and the factors that cause variations among waiting times. Another study hypothesized that delays before the start of treatment are due to the overall failure to provide inadequate resources [5]. That is, the waiting time for treatment is caused by the insufficient supply of machine and staff resources to the patient demand.

Another study addressed waiting times for purely walk-in appointments (i.e. not scheduled in advanced) [6]. In this study, treatment delays were calculated using Queueing Theory. Patient-flow predictions were proposed as one of the most essential clinical management tools since it is the patient queue that drives and defines the clinical workflow, and subsequently patient wait forecasting.

However, until this work, there have been no reports in the literature that computationally approach the uncertainty in waiting times for scheduled daily radiotherapy treatments. In the present era of electronic health records (EHRs), waiting times for these treatments 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 system 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 i

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.00-16

outcomes [7]. Because of their 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. The essence of machine learning is to predict future outcomes by learning from previous experience. The application of machine learning to fractionated treatment data in radiation oncology has the potential to produce personalized waiting time predictions such that the uncertainty may be removed from the radiotherapy patient's waiting experience. Thus, today's machine learning technology can play a vital role in improving the quality of patient care.

II. Materials & Methods

Our radiation oncology department operates on an electronic "tasking" system. These tasks consist of timestamps to record the data and time issued, the due date and time, and the date and time completed, along with a description of the work required, the name of the sender and the name of the receiver. They are created by staff members and doctors and sent between themselves as reminders and records of events. Appointments in our radiation oncology department also operate electronically, in which they include the same timestamp information as tasks, and usually include additional information on the attending physician, rooms to be booked, patient information, etc. An important advantage of electronic tasks and appointments is that where there are recorded timestamps, there are records of waits.

A. *Waiting for Daily Radiotherapy Treatments*

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 he/she actually starts treatment. This delay for a particular patient is primarily the result of the treatment durations (and duration overruns) of those patients who immediately precede him/her. Hence, to provide an arriving patient with an estimate of his/her expected waiting time, we use our machine learning algorithm to predict the treatment *durations* of those preceding patients who have yet to be treated. Fig. 1 provides a detailed illustration of how a personalized daily radiotherapy waiting time estimate is derived.

B. *Modelling waiting times*

Our challenge was to design a learning algorithm that can adapt to a dynamically changing situation, as is encountered in a busy radiation oncology clinic. Modelling waiting times in radiation oncology not only requires previous patient data as experience for learning but also requires putting the proper machine learning implementation steps into practice. At our centre, we have incorporated the treatment records of all previously-treated patients, hosted on our clinical database (ARIA, Varian Medical Systems, Palo Alto, California), into

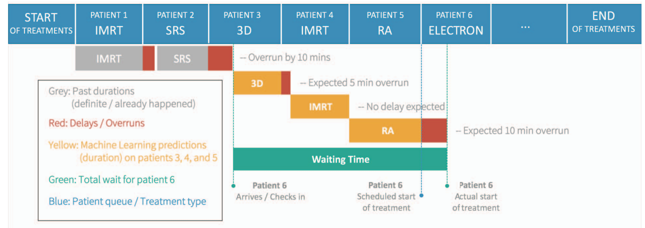

Fig. 1. Defining waiting in the context of daily radiotherapy appointments. In this example, patient 6 arrives just as patient 2's treatment appointment has finished. Both patient 1's and patient 2's appointments have incurred an overall 10-minute delay upon patient 6's arrival. Machine learning is performed 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 is the time of arrival to the estimated start time inferred from the cumulative preceding patient durations.

our machine learning model. The model provides personalized duration predictions to define an overall waiting time for newly-arriving patients at our centre. The treated patient data then become a part of the learning dataset to refine and rebuild the model for future patients. Fig. 2 describes the basic cycle of the machine learning process in the context of waiting times for daily radiotherapy treatments.

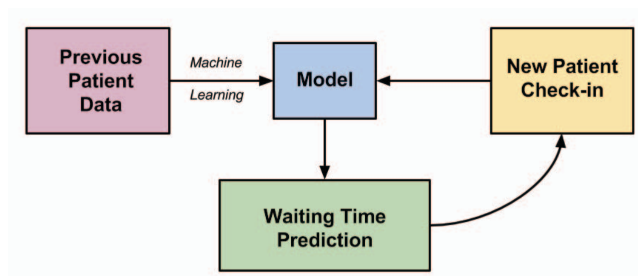

Fig. 2. Schematic of the machine learning process. A model is developed on the collective knowledge and experience learned from previous patient treatment durations. A waiting time prediction, based on durations (fig. 1), is then delivered to a newly-arriving patient. After treatment, the new data become part of the dataset to refine the model for future patients

To develop the prediction model using our large dataset (with more than 100k sample points) a variety of off-the-shelf machine learning algorithms from the Python package scikit-learn (or *sklearn* for short) were tested [8]. Sklearn 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 to 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.

C. *Feature Set*

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 two sources of information: (1) feedback from professionals working in the radiation oncology clinic and (2) other readily-available data in the clinical database. 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 they expected that appointments would last as they predicted.

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 set up 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 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. 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.

Other readily-available data, such as patient demographics and time of the day and week were used as features. Table 1 presents the proposed list of features used for this study.

D. *Machine Learning Model*

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 50 840 retrospective appointment data samples initially extracted from the ARIA database, only 44 225 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.

III. Results

Table 2 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.

TABLE I. List of features extracted for each input sample set. Features were either categorical (as string variables) or continuous (as float variables).

Feature	Data Type
Patient's diagnosis	Categorical
Patient's physician	Categorical
Course of treatment	Categorical
Treatment machine	Categorical
Patient's gender	Categorical
Patient's age at the time of treatment	Continuous
Appointment day of the week	Categorical
Appointment hour	Categorical
Appointment month	Categorical
Type of radiotherapy treatment plan	Categorical
Patient's body orientation on treatment table	Categorical
Radiation type	Categorical
Number of treatment beams	Continuous
Duration of previous treatment appointment	Continuous
Number of pre-treatment images taken for patient setup	Continuous
Duration of pre-treatment images	Continuous
Radiation therapist who is treating the patient	Categorical
Total monitor units (measure of radiation)	Continuous
Total monitor units coefficient (measure of radiation specific to treatment plan)	Continuous
Fraction number of radiotherapy treatment	Continuous
Allocated appointment time	Categorical
Medial past duration of the appointment	Continuous

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

Regression Model	Preprocessing Method	Mean Absolute Error [min]	Median Absolute Error [min]	Standard Deviation Error [min]	R^2
Linear regression	One-hot encoding	4.9	3.9	6.4	0.25
SVM	One-hot encoding	4.9	3.4	7.2	0.13
Decision tree	Numerical categorization	4.7	3.3	6.8	0.42
Random forest	Numerical categorization	4.6	3.3	6.1	0.47

The best trained model, according to the evaluations done on the testing set, was found to be the random forest regression (RFR) model. A residual (i.e. error) histogram is shown in Fig. 3. The median error was found to be 0.25 minutes with a standard deviation of 6.1 minutes. This means that most of our estimates are within 6.1 minutes of the actual treatment duration. The mean absolute error and the median absolute error were found to be 4.6 and 3.3 minutes respectively. 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 3.

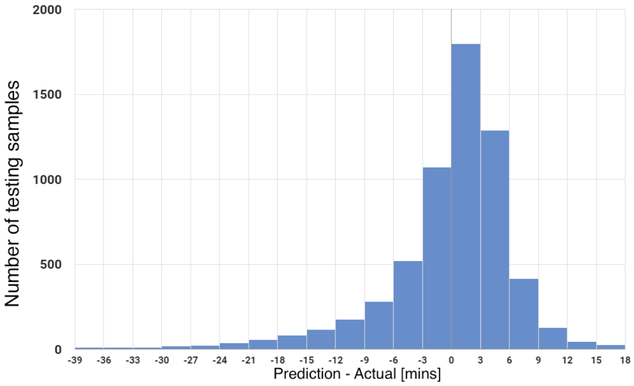

Fig. 3. Histogram of daily radiotherapy treatment duration residuals; Residuals are calculated by subtracting the predicted duration from the true duration.

TABLE III. TOP FIVE FEATURES AS DEFINED BY SKLEARN'S FEATURE IMPORTANCE METHOD USING THE RANDOM FOREST REGRESSION MODEL. OTHER PROPOSED FEATURES CORRESPONDED TO A TOTAL REMAINDER IMPORTANCE OF 8%

Feature	Relative Importance
Allocated appointment time	40%
Fraction number	24%
Median past duration	17%
Number of treatment beams	6%
Previous duration	5%
Others	8%

The main features that generated the best fit model (from most to least significant) are as follows:

- Allocated time – the length of time a patient is scheduled for treatment,
- Fraction number – first fraction is usually the longest (setup time, etc.),
- Median past duration – for fractions > 1, the appointment is likely to last around the times of the previous fractions,
- Number of treatment fields – defines the complexity of the treatment, and
- Previous duration – defines the immediate preceding duration of the previous fraction appointment

A. Comparison Against Simple Averages

The best performing machine learning algorithm (random forest regression model) was compared to simple averages of several categorical features. For 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 4 presents performance metrics of several categorical features, individually used as predictors.

TABLE IV. 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).

Predictor	Mean Absolute Error [min]	Median Absolute Error [min]	Standard Deviation Error [min]	R^2
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	8.4	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**

IV. DISCUSSION

Providing patients with estimates of their overall waiting time is of significant 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 center. With the abundance and availability of patient EHR data at our center, it was possible for us to 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.

A. 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 frequently 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 (Fig. 3).

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.

B. *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 center, they tend to last their allocated appointment time. The majority of treatment appointments at our cancer center 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 set up and treat the patient in a shorter amount of time compared to the first treatment. This matches the intuition made by the radiation therapists when asked about an initial feature set. 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, regardless of the specific treatment protocol. In other words, regardless of other features specific to the patient and their treatment, such as diagnosis, treatment machine, treatment technique, these features can be encapsulated 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.

C. *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. As part of a bigger, ongoing project, our team is developing an app for smartphones and for the web, called "Opal" – the Oncology Portal and Application – for radiation oncology patients at our center [7]. One of the many features of the app is to directly provide patients with their personalized waiting time estimates upon check-in at the clinic. Fig. 4 provides screenshots of the smartphone app.

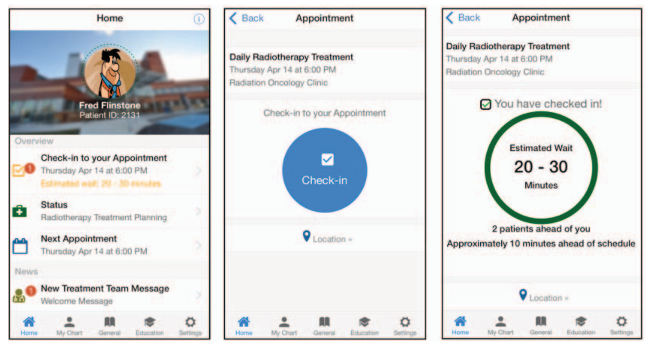

Fig. 4. Opal on a smartphone. Left: home page; Centre: appointment information with an option to check-in; Right: appointment view after checking in with additional information on waiting time estimates.

Opal 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. In addition, patients can be called in for treatment remotely on their phones using Opal.

D. *Further Observations*

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.

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 Y is scheduled after patient X and before patient Z and patient X's treatment finishes early but patient Y is not yet checked-in and patient Z is waiting in the waiting room, then patient Z would likely be called in earlier, due to the absence of patient Y. A fully-useful waiting time estimate algorithm should be able to account for such early call-ins.

V. Conclusions

This study involved developing a machine learning solution to address the pain of waiting in radiation oncology at our comprehensive cancer center. The study investigated waiting for daily radiation treatment appointments. Although treatment appointments are scheduled at our center, 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 treatment durations to within 6.1 minutes. This is a significant improvement on the rough estimates usually given to patients, and 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.

Acknowledgments

AJ acknowledges support by the CREATE Medical Physics Research Training Network grant of the Natural Sciences and Engineering Research Council (Grant number: 432290), from the 2014 Q+ Initiative of the McGill University Health Centre and the Montreal General Hospital Foundation, and from the Henry R. Shibata Fellowship grant of the Cedars Cancer Foundation.

References

[1] Paul C., Carey M, et al, Eur. J Cancer Care 2012; 21, 321–329, Fitch M.I., Gray R.E., Psycho-Oncology 2003; 12: 664–674

[2] Robinson, D., Massey, T., Davies, E., Jack, R., Sehgal, A., and Moller, H.: 2005, British journal of cancer 92(7), 1201

[3] Fortin, B., Goldberg, M. S., Mayo, N. E., Valois, M.-F., Scott, S. C., and Hanley, J.: 2006, Healthcare Policy 1(2)

[4] Jack, R. H., Davies, E. A., Robinson, D., Sainsbury, R., and Moller, H.: 2007, BMC Cancer 7(1), 1

[5] CIHI: 2013, Better Information for Improved Health: A Vision for Health System Use of Data in Canada

[6] Munro, A. J., and S. Potter. "Waiting times for radiotherapy treatment: not all that mysterious and certainly preventable." Clinical Oncology 6.5 (1994): 314-318.

[7] [7] Pianykh, Oleg S., and Daniel I. Rosenthal. "Can we predict patient wait time?." *Journal of the American College of Radiology* 12.10 (2015): 1058-1066.

[8] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825–2830, 2011

[9] opalmedapps.com