Knowledge Discovery in Databases of Radiation Therapy Treatment Planning

by

Yang Sheng

Graduate Program in Medical Physics
Duke University

Date:_____
Approved:

_____
Q. Jackie Wu, Co-Supervisor

_____
Fang-Fang Yin, Co-Supervisor

_____
Qiuwen Wu, Chair

_____
Yaorong Ge

_____
Anuj Kapadia

_____
Bridget Koontz

_____
Taoran Li

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the Graduate Program in Medical Physics
in the Graduate School of Duke University

2017

ABSTRACT

Knowledge Discovery in Databases of Radiation Therapy Treatment Planning

by

Yang Sheng

Graduate Program in Medical Physics
Duke University

Date:_____
Approved:

_____
Q. Jackie Wu, Co-Supervisor

_____
Fang-Fang Yin, Co-Supervisor

_____
Qiuwen Wu, Chair

_____
Yaorong Ge

_____
Anuj Kapadia

_____
Bridget Koontz

_____
Taoran Li

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the
Graduate Program in Medical Physics in the Graduate School of Duke University

2017

Abstract

Radiation has been utilized in medical domain for multiple purposes. Treating cancer using radiation has increasing popularity during the last century. Radiation beam is directed to the tumor cells while the surrounding healthy tissue is attempted to be avoided. Radiation therapy treatment planning serves the goal of delivering high concentrated radiation to the treatment volume while minimizing the normal tissue as much as possible. With the advent of more sophisticated delivery technology, treatment planning time increases over time. In addition, the treatment plan quality relies on the experience of the planner. Several computer assistance techniques emerged to help the treatment planning process, among which knowledge-based planning (KBP) has been successful in inverse planning IMRT. KBP falls under the umbrella of Knowledge Discovery in Databases (KDD) which originated in industry. The philosophy is to extract useful knowledge from previous application/data/observations to make predictions in the future practice. KBP reduces the iterative trial-and-error process in manual planning, and more importantly guarantees consistent plan quality. Despite the great potential of treatment planning KDD (TPKDD), three major challenges remain before TPKDD can be widely implemented in the clinical environment: 1. a good knowledge model asks for sufficient amount of training data to extract useful knowledge and is therefore less efficient; 2. a knowledge model is usually only applicable for the specific treatment site

and treatment technique and is therefore less generalizable; 3. a knowledge model needs meticulous inspection before implementing in the clinic to verify the robustness.

This study aims at filling in the niche in TPKDD and improves current TPKDD workflow by tackling the aforementioned challenges. This study is divided into three parts. The first part of the study aims to improve the modeling efficiency by introducing an atlas-based treatment planning guidance. In the second part of the study, an automated treatment planning technique for whole breast radiation therapy (WBRT) is proposed to provide a solution for the area where TPKDD has not yet set foot on. In the third part of the study, several topics related to the knowledge model quality are addressed, including improving the model training workflow, identifying geometric novelty and dosimetric outlier case, building a global model and facilitating incremental learning.

I. Improvement of the modeling efficiency. First, a prostate cancer patient anatomy atlas was established to generate 3D dose distribution guidance for the new patient. The anatomy pattern of the prostate cancer patient was parameterized with two descriptors. Each training case was represented in 2D feature space. All training cases were clustered using the *k*-medoids algorithm. The optimal number of clusters was determined by the largest average silhouette width. For the new case, the most similar case in the atlas was identified and used to generate dose guidance. The anatomy of the atlas case and the query case was registered and the deformation field was applied to

the 3D radiation dose of the atlas case. The deformed dose served as the goal dose for the query case. Dose volume objectives were then extracted from the goal dose to guide the inverse IMRT planning. Results showed that the plans generated with atlas guidance had similar dosimetric quality as compared to the clinical manual plans. The monitor units (MU) of the auto plan were also comparable with the clinical plan. Atlas-guided radiation therapy has proven to be effective and efficient in inverse IMRT planning.

II. Improvement of model generalization. An automatic WBRT treatment planning workflow was developed. First of all, an energy selection tool was developed based on previous single energy and dual energy WBRT plans. The DRR intensity histograms of training cases were collected and the principal component analysis (PCA) was performed to reduce the dimension of the histogram. First two components were used to represent each case and the classification was performed in the 2D space. This tool helps new patient to select appropriate energy based on the anatomy information. Secondly, an anatomy feature based random forest (RF) model was proposed to predict the fluence map for the patient. The model took the input of multiple anatomical features and output the fluence intensity of each pixel within the fluence map. Finally, a physics rule based method was proposed to further fine tune the fluence map to achieve optimal dose distribution within the irradiated volume. Extra validation cases were tested on the proposed workflow. Results showed similar dosimetric quality between auto plan and clinical manual plan. The treatment planning time was reduced from

between 1-4 hours for the manual planning to within 1 minute for the auto planning. The proposed automatic WBRT planning technique has proven to be efficient.

III. Rapid learning of radiation therapy KBP. Several topics were analyzed in this part of the study. First of all, a systematic workflow was established to improve the KBP model quality. The workflow started with identifying geometric novelty case using the statistical metric "leverage", followed by removing the novelty case. Then the dosimetric outlier was identified using studentized residual and then cleaned. The cleaned model was compared with the uncleaned model using the extra validation cases. This study used pelvic cases as an example. Results showed that the existence of novelty and outlier cases did degrade the model quality. The proposed statistical tools can effectively identify novelty and outlier cases. The workflow is able to improve the quality of the knowledge-based model.

Secondly, a clustering-based method was proposed to identify multiple geometric novelty cases and dosimetric outlier cases at the same time. One class support vector machine (OCSVM) was applied to the feature vectors of all training cases to generate one class of inliers while cases falling out of the frontier belonged to the novelty case group. Once the novelty cases were identified and cleaned, the robust regression followed by outlier identification (ROUT) was applied to all remaining cases to identify dosimetric outliers. A cleaned model was trained with the novelty and outlier free case pool and was tested using 10 fold cross validation. Initial training pool included

intentionally added outlier cases to evaluate the efficacy of the proposed method. The model prediction on the inlier cases was compared with that of novelty and outlier cases. Results showed that the method can successfully identify geometric novelty and dosimetric outliers. The model prediction accuracy between the inliers and novelty/outliers was significantly different, indicating different dosimetric behavior between two groups. The proposed method proved to be effective in identifying multiple geometric novelty and dosimetric outliers.

Thirdly, a global model using the model tree and the clustering-based model was proposed to include cases with different clinical conditions and indications. The model tree is a combination of decision tree and linear regression, where all cases are branched into leaves and regression is performed within each leaf. A clustering-based model used *k*-means algorithm to segment all cases into more aggregated groups, and then the regression was performed within each small group. The overall philosophy of both the model tree and the clustering-based method is that cases with similar features have similar geometry-dosimetry relation. Training cases within small feature range gives better model accuracy. The proposed method proved to be effective in improving the model accuracy over the model trained with all cases without segmenting the cases.

At last, the incremental learning was analyzed in radiation therapy treatment planning model. This study tries to answer the question when model re-training should be invoked. In the clinical environment, it is often unnecessary to re-train the model

whenever there is a new case. The scenario of incrementally adapting the model was simulated using the pelvic cases with different number of training cases and new incoming cases. The result showed that re-training was often necessary for small training dataset and as the number of cases increased, re-training became less frequent.

In summary, this study addressed three major challenges in TPKDD. In the first part, an atlas-guided treatment planning technique was proposed to improve the modeling efficiency. In the second part, an automatic whole breast radiation therapy treatment planning technique was proposed to tackle the issue where TPKDD has not yet resolved. In the final part, outlier analysis, global model training and incremental learning was further analyzed to facilitate rapid learning, which lay the foundation of future clinical implementation of radiation therapy knowledge models.

Contents

List of Tables

List of Figures

List of Abbreviations

ANN: artificial neural network

AP: Auto Planning

ART: adaptive radiation therapy

ASTRO: American Society for Radiation Oncology

AUC: area-under-curve

BEV: beam's eye view

CBCT: cone-beam CT

CI: conformity index

CNN: convolutional neural network

COM: center of mass

CT: computed tomography

CTV: clinical target volume

DARC: dynamic conformal arc

DRR: digital reconstructed radiograph

DTH: distance-to-target histogram

DVH: dose-volume histogram

ECOMP: electronic compensation

EM: expectation maximization

FDR: false discovery rate

FiF: field-in-field

gEUD: generalized equivalent-uniform-dose

GMM: Gaussian Mixture Model

GTV: gross tumor volume

HI: homogeneity index

HN: head-and-neck

IGRT: image-guided radiation therapy

IMRT: intensity modulated radiation therapy

KBP: knowledge-based planning

KDD: Knowledge Discovery in Databases

LINAC: linear accelerator

LN: lymph node

MCO: multi-criteria optimization

MI: mutual information

MLC: multi-leaf collimator

MRI: magnetic resonance imaging

MU: monitor unit

OAR: organ-at-risk

OCSVM: one-class support vector machine

OVH: overlap volume histogram

PC: principal component

PCA: principal component analysis

PET: positron emission tomography

PTV: planning target volume

QM: quality metric

RBF: radial basis function

RF: random forest

ROC: receiver-operating-characteristic

ROUT: robust regression followed by outlier identification

RSDR: Robust Standard Deviation of the Residuals

SABR: stereotactic ablative body radiotherapy

SAM: segment aperture morphing

SBRT: stereotactic body radiation therapy

SRS: stereotactic radiosurgery

SV: seminal vesicles

SVM: support vector machine

SWO: segment weight optimization

TPKDD: treatment planning Knowledge Discovery in Databases

TPS: treatment planning system

VMAT: volumetric modulated arc therapy

WBRT: whole breast radiation therapy

WSAR: Weighted Sum of Absolute Residual

WSR: Weighted Sum of Residuals

Acknowledgements

I would like to express my sincere thankfulness to my research advisors, Dr. Jackie Wu and Dr. Fang-Fang Yin, for their constructive suggestion and advice for my research project, and also the patient guidance and incentive encouragement for my career in Medical Physics. I would also like to thank Dr. Qiuwen Wu for serving as my committee chair and his insightful comments and critiques about this research project. I also thank Dr. Yaorong Ge for the constructive critique provided and mentoring for my research training. I would also like to extend my gratitude to Dr. Taoran Li for his constructive recommendation during this study, as well as the mentorship throughout my Medical Physics career. I would also thank Dr. Bridget Koontz for her participation and comments about this project in spite of her busy clinical schedule, and Dr. Anuj Kapadia for encouragement and being on my committee. I also thank Dr. Lulin Yuan, Dr. You Zhang, Dr. Chunhao Wang, Dr. Jiahan Zhang, Tianyi Xie, Xiao Liang, and Kyle Lafata for the encouraging discussion for my research projects. My gratitude also extends to Ms. Wendy Harris and Dr. Anna Rodrigues who participated in the discussion and proof read the manuscripts.

Last but not least, I would like to thank my parents for their continuous support for my education since I was a kid. They are great educators and I enjoy being a family member. Looking back the last twenty years, I am deeply grateful for them bringing me

up, for the best education I had, for the comfort they gave me when I was depressed, and for everything they do for me which shapes me who I am.

1. Background

1.1 Radiation therapy treatment planning

Since the discovery of x-ray by Roentgen, several radiation formats (photon, electron, proton, neutron etc.) have been applied to medical domain. Kilo voltage x-ray can be used for imaging purposes, e.g. scout x-ray imaging, computed tomography (CT), cone-beam CT (CBCT) etc. High voltage x-ray, i.e. mega voltage x-ray, is being used for treatment purposes, especially in treating cancer patients. The whole radiation therapy treatment process is composed of several serial components: imaging, treatment planning, treatment delivery and treatment assessment (Figure 1). A radiation therapy treatment plan is needed to deliver prescribed radiation dose to the tumor volume while minimizing unnecessary radiation to surrounding healthy tissue. The treatment planning process takes place after the organs being contoured, including tumor volume and surrounding organs-at-risk (OARs), and before the treatment delivery. Based on available imaging modalities, for example CT, magnetic resonance imaging (MRI) and positron emission tomography (PET), the physician will contour the gross tumor volume (GTV) and transfer to CT if contoured on MRI or PET. A clinical target volume (CTV) which encompasses GTV is then contoured to include sub-clinical malignant volume. The radiation therapy plan aims to deliver prescribed dose to the planning target volume (PTV), which is an expansion from CTV considering patient setup error during treatment delivery. CTV-PTV margin is dependent on the treatment modality

and the on-board imaging device availability. Treatment planning is the virtual process where the radiation is delivered to the patient and the dose accumulated in the patient is summarized on a computer. By taking a "snapshot" of the patient's anatomy, radiation can be virtually delivered to the CT volume and the dose can be calculated based on certain dose calculation model. How the radiation is delivered (intensity, beam shape etc.) is adjusted driven by the dose accumulated. The treatment planning process requires the planner's knowledge to adjust treatment parameters, including beam angle, beam weight, beam intensity etc., to achieve an optimal dose distribution. Depending on the treatment modality and patient anatomy, the effort required for the treatment planning process varies. For 3D treatment plan, the planner usually adjusts the beam weight to conform the high dose region to the tumor volume. For more advanced techniques which call fluence modulation, such as intensity modulated radiation therapy (IMRT), volumetric modulated arc therapy (VMAT) and dose painting, an optimization process is often warranted. In IMRT optimization, the achievable dose distribution depends on the patient's anatomy, and is often blind to the planner *in priori*. Therefore, how the dose volume constraints are configured affects the final dose distribution. Inadequate fluence modulation may introduce extra unnecessary dose spill to OARs, while over modulation may compromise the tumor volume dose coverage. It is reported that the plan quality of IMRT plan is related to the experience of the planner. And often, the planning process is iterative and time consuming.

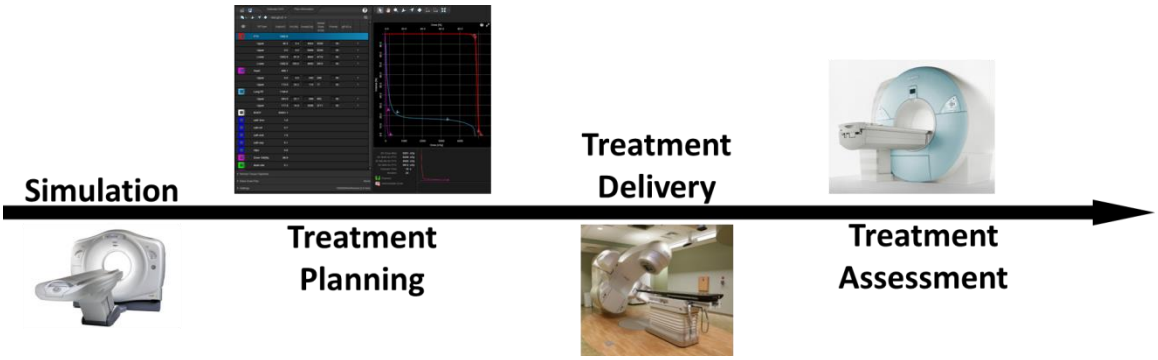

Figure 1: Typical workflow of LINAC-based radiation therapy treatment.

1.2 Computer assistance in radiation therapy treatment planning

Conventional radiation therapy treatment planning process relies on hand calculation to obtain dose distribution. Later, the treatment process was simulated on a CT simulator to mimic the treatment process while the treatment region is localized. With the advent of the computer, faster and more accurate dose calculation is feasible. In addition, the treatment planning process enters the era of virtual planning, i.e. the volumetric CT image of the patient is acquired and then the planning process is performed on the virtual image. With virtual planning, the planner is capable of tuning different planning parameters, including beam configuration, beam intensity etc., to achieve the planning goal. It allows more flexibility of generating clinically acceptable plan.

More sophisticated computation power allows advanced treatment delivery techniques, such as IMRT and VMAT. Unlike classic 2D or 3D treatment, more planning parameters are open to change to achieve the planning goal. For example, IMRT is capable to tuning the fluence intensity using either multi-leaf collimator (MLC) or

multiple segments. Using cross-firing radiation delivery, the radiation dose can better conform to the treatment volume while minimizing the dose to surrounding health tissue. The planner will set dose volume constraints for specific organ to regularize the fluence optimization process, which is often referred as the inverse planning. The planner starts with setting the dosimetric endpoints to achieve optimal fluence map. However, setting dose volume points is a non-intuitive process and requires extensive experience in IMRT planning. Often, the planner has to iteratively fine tune the objectives to achieve optimal dose distribution, thus increases the treatment planning time. And the plan quality is dependent on the experience of the planner as well.

In order to improve the treatment planning efficiency and the plan quality consistency, several computer assistances in treatment planning have been introduced, including knowledge-based planning (KBP), multi-criteria optimization (MCO) and heuristic optimization.

1.2.1 Knowledge-based planning (KBP)

Knowledge-based planning was a model based application to apply previous knowledge to guide future application (Appenzoller, Michalski, Thorstad, Mutic, & Moore, 2012; Fogliata, Wang, et al., 2014; Krayenbuehl, Norton, Studer, & Guckenberger, 2015; B. Wu et al., 2011; Yuan et al., 2012a). The philosophy of KBP is to build a model which is able to capture the relation between the patient's anatomy information (input) with the dosimetric information (output). And once the model is built, it can predict the

dosimetric information for the query patient based on its anatomy. It borrows the idea in machine learning which uses certain algorithms to capture the input and output relation and predicts future practice which is hard or expensive for human to operate.

In order to build a predictive model, descriptive features for both the anatomy and dosimetry information is necessary. Usually, domain experts will identify or hand craft certain features to import into the model. Domain knowledge helps the model to better capture the useful relation between the anatomy and geometry. Then the model will select certain features which are predictive for the endpoint and train the model based on the selection. The model is able to record the mean behavior of all training cases and use the knowledge to make predictions. A successful model relies on good clinical plans which embed good knowledge and appropriate model training to capture the knowledge. The model will give guidance in the treatment planning process, such as dose volume objectives and 3D dose distribution.

The KBP has been commercialized in RapidPlan (Varian Medical System, Palo Alto, CA). It allows the planner to import previous IMRT/VMAT plans to train the model. Certain model parameters are also provided for the planner to inspect and oversee the model training. Once the model is finalized, it can be used to make dose volume objectives prediction for future patients.

1.2.2 Multi-criteria optimization (MCO)

Radiation therapy treatment planning is often associated with competing constraints, meaning one criterion cannot be improved without compromising another one. For example, the radiation dose to the OAR cannot be infinitely reduced without compromising the target volume coverage. The process of generating a plan with balanced criteria/constraints is called multi-criteria optimization (MCO) (Craft, Halabi, Shih, & Bortfeld, 2007; David, Tarek, & Thomas, 2005). For a specific patient, the optimal combination of criteria is often unknown prior to the planning process, therefore generating a large number of plans is a typical way to identify the optimal criteria. A plan is called "Pareto optimal" when such a plan cannot improve one criterion without compromising another one. This subset of plans generates the Pareto front. If sufficiently large amount of plans are accessible, the physician will be able to select a plan that has better tradeoff effect among all criteria. However, due to the nature of MCO, a substantial amount of time is necessary to generate a series of plans and therefore to construct the Pareto front. Unlike KBP, MCO can generate truly optimal plans with the cost of substantially increased treatment planning time. The concept of MCO has been commercialized in RayStation (RaySearch Laboratories, Stockholm, Sweden), where the planner is able to interactively modify the goal constraints to obtain an optimal plan.

1.2.3 Heuristic optimization

Another approach to accelerate the planning process is to explore the possible solution of the plan with the automation provided by the computer. This explorative process is known as the heuristic optimization (Purdie, Dinniwell, Fyles, & Sharpe, 2014; Purdie, Dinniwell, Letourneau, Hill, & Sharpe, 2011). Thanks to the computation power, searching through the solution space becomes feasible, which may cost extreme amount of time for human planner to test different planning parameters. One example of heuristic optimization is provided by RayStation. It utilizes heuristic optimization to help identify optimal beam configuration for whole breast irradiation. Once the boundary of the breast volume is identified on the CT volume, the optimization modual will explore all possible beam angles setup and identify the optimal one which has less penetration through the OARs such as lungs. Another example of the heuristic optimization is the Auto Planning (AP) modual in Pinnacle Treatment Planning System (TPS, Philips, Inc., Andover, MA). AP mimics the human planning process, such as setting dose volume objectives, fine tuning dose volume objectives during the optimization process, contouring assistant structures for further optimization etc. This approach reduces human intervention in the planning while the solution is not guaranteed global optimal.

1.2.4 Comparison

Aforementioned three approaches are all effective in improving radiation therapy treatment planning efficiency. The ideal technique for treatment planning is the MCO approach, where the patient specific tradeoff is known and the physician can make adjustment between competing constraints. However, the MCO approach is usually time consuming in the sense that it requires sufficient amount of time to construct a faithful Pareto front. KBP is an alternative approach for guiding treatment planning. The overall philosophy of KBP is to extract the knowledge in the previous plans and use the knowledge in the future practice. Unlike MCO, the knowledge embedded in the KBP model does not guarantee global optimal solution, due to the plan quality variations within the training dataset. Therefore, it is very important to inspect the training data quality for the KBP approach. In addition, the KBP approach is able to serve as estimation for the global optimal solution. Since the MCO approach is time consuming, the KBP approach can help limit the Pareto front searching since in most scenarios the clinical acceptable solution is within a confined range of the Pareto surface. Once the KBP is called, the guidance can be used to further approximate the Pareto solution. The KBP-MCO approach would reduce the solution searching time while guarantees optimal solution. The heuristic optimization replaces the human planning process. In mimics the human planning process to gradually reach the acceptable plan. Due to the superior computation power, this approach can achieve the best treatment

parameter which is usually time-consuming for the human planner to achieve. For example, the optimal beam angle can be searched and identified based on the efficiency of the beam angle. The heuristic optimization also serves as an alternative approach for treatment planning automation.

The KBP approach also contrast the other two approaches in the sense that it asks for an appropriate subset of features to learn the knowledge. The MCO searches a sufficiently large pool of solutions where the optimal solution can be identified. The heuristic optimization also directly works on the task to reach a solution. The KBP approach provides guidance/knowledge for the specific task and usually does not work on the task directly. For example, RapidPlan provides dose-volume objectives while the planner still needs to run the optimization. Learning the knowledge within the task asks for feature identification. Since there are numerous features available, currently domain expert empirically identifies the predictive features to aid the knowledge learning process. In addition, due to the limited data samples available for radiation therapy treatment plans, the number of features that can be used has to be confined in a limited range (empirically as the square root of the number of samples). The feature identification process could potentially be replaced by the sophisticated machine in the future. Highly non-linear model such as the neural network/deep learning uses linear combination of a large input feature vectors to make predictions. Some linear combinations of all features are predictive for the output although it may lack the

clinical interpretation. The future development of radiation therapy KBP may rely on fast computation and more sophisticated model to make sense of all features available and to provide a more generalized strategy for feature identification.

1.3 Clinical significance

Radiation has been utilized to treat cancer for decades. Treatment facilities vary from Cobalt source unit, i.e. gamma knife, to LINear ACcelerator (LINAC). LINAC based external photon beam treatment has been mainstream for treating multiple cancer sites patients, e.g. intra-cranial cancer, head-and-neck (HN) cancer, lung cancer, pelvic cancer etc. State-of-the-art LINAC is equipped with multiple high accuracy components which enable advanced treatment techniques. Such components include on-board imaging which enables image-guided radiation therapy (IGRT), high definition MLC, 6D couch which provides more choices of beam angles. Fluence modulation plus on-board imaging enables highly conformal dose delivered to the patient. IMRT, VMAT and dose painting have been widely implemented to spare normal tissue while providing adequate dose coverage to the tumor volume. These treatment modalities have proven to minimize the radiation induced injuries.

With the advent of computerized knowledge learning methods, automation of radiation therapy treatment planning has been made possible. It not only saves human planner's effort, but also guarantees the quality of the plan as long as the quality of the model is carefully inspected, reducing the possibility of generating a suboptimal plan.

The automation of treatment planning also provides the physician more choices to choose from as each plan can be readily available within short amount of time. This is an important step forward towards precision-medicine, which aims to improve patient care.

1.4 Knowledge Discovery in Databases (KDD)

Data exists in all fields and expands at an incredible pace. With an increasing pace, data is being collected with different types and manners. People were trying to learn knowledge from huge volume of data so as to utilize the knowledge in future application. Fayyad et al. (U. Fayyad, G. Piatetsky-Shapiro, & P. Smyth, 1996; Usama Fayyad, Gregory Piatetsky-Shapiro, & Padhraic Smyth, 1996; U. Fayyad & Stolorz, 1997; U. M. Fayyad, 1996; Usama M. Fayyad, 1997) introduced the concept of Knowledge Discovery in Databases (KDD) which describes the overall process of learning knowledge from raw dataset. The entire process starts with treating untouched raw dataset and outputs the abstract and more useful knowledge. The pipeline consists of five steps: 1. data selection, 2. data preprocessing, 3. data transformation, 4. data mining and 5. knowledge interpretation and evaluation (Figure 2).

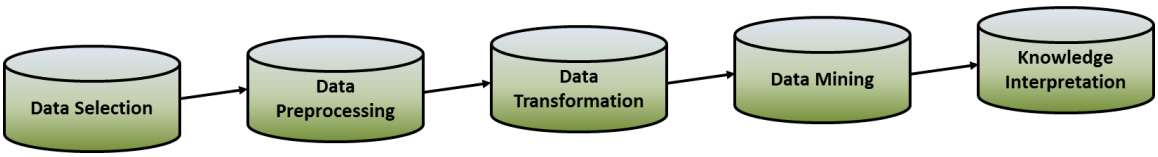

Figure 2: Workflow of conventional KDD process

Denote D as the database, L as certain language (statistical) and ε as some certainty measure. The KDD process is to extract the pattern P in the language L which summarizes the relation among D_s (a subset of D) with a certainty ε_0. The pattern P shall be simpler than enumerating all the facts in D. The first step is selecting a subset of data/input variable D_s from D to start the learning process. The second step is in general removing noise data and handling missing data labels. The third step is transforming high dimensional data into low dimensional representations. This step is necessary as considering the computation issue for the next step. The fourth step is data mining, which is using machine learning algorithm to learn the pattern P in the dataset and finding the knowledge. Data mining and machine learning has been increasingly popular in recent years. It composes one part of the entire KDD process. The success of the KDD process depends not only on the appropriate data mining methods, but also the other four steps to inspect the knowledge learned. The fifth and last step is evaluating the knowledge extracted and performing fine-tuning of the learning process if necessary. The entire process could be closed-loop iterative process. KDD has been widely implemented (Butte & Kohane, 1999; Chu, Cárdenas, & Taira, 1995; Usama Fayyad et al., 1996; U. Fayyad & Stolorz, 1997; U. M. Fayyad, 1996; Usama M. Fayyad, 1997; Frawley, Piatetsky-Shapiro, & Matheus, 1992; J. Han, Cai, & Cercone, 1992; J. Han & Fu, 1994; C. S. G. Khoo, Chan, & Niu, 2000; Kovalerchuk, Vityaev, & Ruiz, 2000; Mackinnon & Glick, 1999; Matheus, Chan, & Piatetsky-Shapiro, 1993; Ohsaki, Abe,

Tsumoto, Yokoi, & Yamaguchi, 2007; Piatetsky-Shapiro, 2000; Piatetsky-Shapiro, Matheus, Smyth, & Uthurusamy, 1994; Prather et al., 1997; Soibelman & Kim, 2002; Tan, Yu, Heng, & Lee, 2003; Tsumoto, 2000a, 2000b; Wille, 2002; Yoon & Kerschberg, 1993), e.g. providing marketing strategies based on customers' previous behaviors, predicting game results based on each team's statistics, detecting fraud credit card transactions and so on. Frawley et al (Frawley et al., 1992) gave a more statistical view of KDD. Knowledge discovery is the process of extracting and summarizing unknown information which could be potentially useful from existing database.

KDD deals with the large volume database, which is also often referred as data warehouse. Since the data volume is large, the appropriate selection of data is critical to mine the underlying pattern. Additional domain knowledge and user-defined biases can be provided to assist the discovery process (Frawley et al., 1992). Database differs from traditional statistics in four aspects (Mackinnon & Glick, 1999): 1. The volume of the data is massive; 2. The data is highly heterogeneous (collected by multiple people, different time series etc.); 3. The type of data varies across the database; 4. The pattern does not remain static over time. Database technologists are more concerned about how to store and retrieve the data more efficiently. However, machine learning and industry community is more concerned about learning knowledge from the data.

1.5 Treatment planning KDD (TPKDD)

KDD was not first introduced in medical domain but it has been increasingly popular in recent years. In medical domain, data is spread in all formats and types, e.g. lab test results, physician diagnosis, imaging data etc. Efforts are being made to combine available data in the hospital to provide guidance of treatment regimen selection and treatment outcome prediction. Knowledge discovery in the database of congenital malformation was developed using rule induction methods (Tsumoto, 2000b). Similarly, discovery of positive and negative knowledge in clinical databases was performed using rule induction based method (Tsumoto, 2000a). Rule based reasoning has also been analyzed among computer-aided systems, experts and the databases to eliminate insistent knowledge discovered in the medical diagnosis such as breast cancer diagnosis (Kovalerchuk et al., 2000). An evolutionary classifier was developed to prevent unwanted medical events (Tan et al., 2003). A relevance network was developed to find the correlation between variables in large medical database so that the laboratory tests could be meaningful (Butte & Kohane, 1999). Textual database was also explored and a knowledge extraction and knowledge discovery system which extracts causal knowledge was developed (C. S. G. Khoo et al., 2000). Clinical data warehouse was explored and the knowledge was mined to discover the contributing factors for preterm birth (Prather et al., 1997). KDD is a necessary step towards the precision-medicine, which provides patient specific and tailored medical care for the patient.

Efforts on mining knowledge in radiation therapy treatment planning were previously made. The database of radiation therapy treatment planning consists of various data types, including patient imaging I, structure set S, treatment modality M, treatment machine m, patient medical diagnosis R, prescription Rx, radiation plans R_P, radiation dose R_D etc. Treatment planning KDD (TPKDD) is the process of automatic learning of the mapping f from patient's anatomical feature and medical condition to the treatment plan parameters under certain modality.

$$R_{P/D} = f_M(I, S, m, R, R_x, \dots) \qquad \text{Eq.(1)}$$

where I is the patient imaging, S is the structure set attached to the imaging I, m is the treatment machine, R is the diagnosis, R_x is the prescription and f_M is the mapping of all above features to the radiation plan/dose related parameters under the treatment modality M.

Previous research related to TPKDD was performed to learn the knowledge of treatment planning using prior treatment plans, to facilitate the automation of treatment planning in the future. Several studies focused on predicting dosimetric endpoints $R_{P/D}$, in either 2D format, i.e. dose-volume histogram (DVH) or 3D format, i.e. optimal dose distribution for the patient. Either format will benefit the planning process, as once an explicit goal is set for the planner, trial-and-error will be most likely avoided. On the

other hand, previous studies have explored some of the popular data mining techniques to solve this task. Depending on what data is available and what knowledge is needed, different data mining techniques may be used. In general, the data used for learning includes patient imaging (CT), structure set, dose distribution, beam fluence, beam configuration etc. The goal is to learn the relation between patient anatomy (imaging and structure set) and optimal dose distribution.

1.5.1 Anatomy-based method

For patients with the same disease and staging, class solutions have been proposed to provide dose objective guidance and/or other optimization parameter guidance. Class solution is defined as a set of treatment planning parameters related to one anatomical site to cover certain anatomy variation range in such site (Lessard, Kwa, Pickett, Roach, & Pouliot, 2006). Class solutions for 3D prostate plan treatment planning were provided using three-field, four-field and six-field plan templates (V. S. Khoo, Bedford, Webb, & Dearnaley, 2003). IMRT planning dose constraints templates were developed to provide inverse treatment planning process for nasopharyngeal treatment (Xia et al., 2004). Similarly, inverse planning parameters were also developed for simultaneous boost IMRT for hypofractionated prostate treatment (Mott, Livsey, & Logue, 2004). This set of planning parameters reduces the planner's intervention in the planning process. In Brachytherapy, inverse planning optimization parameters were provided to mimic the experienced dosimetrist (Lessard et al., 2006). Such parameters

include parameters associated with the organs, parameters associated with the implants and these parameters were tuned to fulfill clinical requirement. For prostate stereotactic ablative body radiotherapy (SABR) using VMAT, class solution for the choice of arc angle and CTV-PTV was analyzed (Murray et al., 2014). For early stage Hodgkin's lymphoma treated with VMAT, a single arc or a 3 arcs plan was preferred class solution for patients with mediastinal disease (Fiandra et al., 2012). Class solutions were also provided to guide IMRT treatment planning. These methods focused on finding the anatomical match between the existing and the new patient cases based on certain similarity metric. The new case then used part or all of the matched plan parameters (dose, fluence maps, MLC segments, etc.) as starting point of the optimizer, and additional manual refinement is required to generate the final plan. Generalizable class solution for spinal stereotactic body radiation therapy (SBRT) using IMRT was generated by retrospectively analyzing the spine SBRT plan quality (Weksberg et al., 2012). Plans were grouped based on the tumor location in the spine and the shape of the CTV. Dosimetrically superior plans were chosen for corresponding group to generate class solution including beam angles, IMRT optimization objectives and weights, planning structures etc. Similarly, beam angle class solution was provided for central nervous system malignancies treated with IMRT, where the dosimetric quality is sensitive to beam angle settings (Likhacheva, Palmer, Du, Brown, & Mahajan, 2012). Previous IMRT plans were grouped based on anatomical shape and superior plans were

identified to generate class solution. Schreibmann et al. developed class solution for prostate IMRT beam orientation using multi-objective optimization (Schreibmann & Xing, 2004). Yuan et al. developed a cluster-based method to generate lung IMRT beam angle bouquet (Yuan et al., 2015). Chanyavanich et al. used 95 mutual information (MI) match of the beam's eye view (BEV) projections of the structure contours to find reference case from a library of previous clinical prostate plans; while Wu et al matched the new patient case to the database using the overlap volume histogram (OVH) calculated over the PTV-OAR geometry. These methods relied on a sufficiently large library, consisting of typically over 100 cases 100, so that a very close match can be found and the existing plan dosimetric characteristics can be directly applied to guide the planning of the new case. This group of method falls into the category of nearest neighbor and case-based reasoning system. It is referred to as "***anatomy-based method***" in this proposal. Under the framework of TPKDD, the five-step workflow can be summarized as: 1. previous treatment plans R_P together with imaging I and structures S are selected as the subset of database D_S; 2. patients with missing fields are excluded for processing; 3. define a similarity measure Δ; 4. use classification or clustering method to find the pattern between R_P and I/S; 5. validate the pattern using query cases. The treatment planning parameter R_P of the query case q will be the best choice selected from the most similar p one from the database.

$$R_P^q = R_P^{p_{min}}, subject\ to\ p_{min} = argmin(\Delta_{qp}) \qquad \text{Eq.(2)}$$

where q is the query case and p is the existing case in the database. Δ is the similarity measure between any two cases. The most similar case in the database p_{min} and its plan parameters $R_P^{p_{min}}$ will be used for the query case q.

Chapter 3 and 4 will use the anatomy-based method to improve the data mining techniques in the KDD workflow.

1.5.2 Statistics-based method

Another group of methods took the machine-learning-based approach, where the relationships between certain anatomy features (as input) and dosimetry features (as output) were formulated. Zhu et al. and Yuan et al. proposed a machine learning model to provide patient-specific OAR sparing goals. This method did not require matching specific query case with a database of expert cases. Rather, the OAR sparing goals were computed by a prediction algorithm. Lian et al. further expanded the model to analyze inter-technique and inter-institutional performance of these models, demonstrating highly conformal dose distribution was achieved with Tomotherapy and static-gantry IMRT techniques at a comparable level. Appenzoller et al. also proposed patient-specific OAR sparing models based on the correlation between the expected DVH within certain distance range to the PTV surface. This model was further developed for stereotactic radiosurgery (SRS) plans to predict DVH-based quality metrics (QMs). The model is

capable of identifying suboptimal plans. Apart from aforementioned 2D dose volume endpoint prediction, 3D dose distribution prediction was also studied. Shiraishi et al. used artificial neural network (ANN) to predict 3D dose matrix for specific patient taking the input of patient's anatomy parameters and treatment parameters. They validated the model on prostate cases and SRS cases and highly accurate dose distribution was achieved. This group of method falls into the category of regression and classification category. It is referred to as "***statistics-based method***" in this proposal. Under the framework of TPKDD, the five-step workflow can be summarized as: 1. previous treatment plan dose R_D with the same treatment modality together with imaging I and structures S etc. are selected as the subset of database D_S; 2. patients with missing fields are excluded for processing; 3. reduce the dimension of the data if necessary (T); 4. use classification or regression method to find the relation between R_D and anatomical variables I, S, etc; 5. validate the pattern using the query cases. The plan dose R_D of the query case q will be predicted from the model.

$$R_D^q = f\left(I_q, S_q, \ldots\right) \qquad \text{Eq.(3)}$$

where I_q is the imaging of case q, S_q is the structure set of case q.

Chapter 5 to Chapter 9 will use the statistics-based method of TPKDD. These Chapters will focus on the data selection and data preprocessing components of the KDD workflow.

1.6 Challenges facing the TPKDD and clinical implementation

In TPKDD process, five steps that compose the entire workflow are equally important. Previous studies mainly focus on exploring various data mining algorithms although the endpoint is sometimes similar. There remain many niches in TPKDD considering the model efficiency, different application for various treatment sites and translation to clinical implementation.

1.6.1 Model efficiency

Most of previous studies have proven effective in the designated task. However, some of the models are less efficient due to the required training data size. For example, the anatomy-based method proposed by Chanyavanich et al. asked for a large library in order to cover possible prostate anatomy. This method requires large training data size p and directly applying the plan parameters (fluence) from previous plan may not be the optimal solution. The idea of TPKDD is to mine the pattern within the dataset while such pattern is simpler than enumerating all facts in the database. A large training data size which approximates the entire database D violates this assumption. Recall equation 2, the choice of p can be further simplified to refine the knowledge. A more descriptive system based on the anatomy of the patient could be developed. Such system should

ideally be efficient in preparing the data and constructing the library/atlas. And more importantly the plan guidance (fluence/dose) could be tailored towards each query anatomy to improve the plan quality in terms of target coverage and OAR sparing.

1.6.2 Model generalization

The model developed in TPKDD addressed certain treatment site and treatment modality, with the exception that some models are generalizable to other treatment sites. For example, the anatomy-based method intrinsically requires that the library/atlas is able to cover most of possible anatomies of certain treatment site. This requirement can be easily met for prostate cancer patients. However, some other cancer sites exhibit various morphologies such as lung cancer and HN cancer. A large library/atlas is often needed and sometimes constructing such library/atlas is not easy task due to the limited number of clinical cases available.

Statistics-based method such as the machine learning model proposed by Yuan et al. is able to predict the DVH endpoints for certain treatment site. The model is often targeted towards IMRT or VMAT treatment. A model trained using cases from single site is able to predict decently for the query case from the same anatomy, while predicting cases from other treatment site is less favorable.

Both aforementioned methods adopt the domain expert knowledge in data selection. Usually it is driven by clinical needs. Thus the extracted knowledge is confined within the selected subset of data D_S and is not applicable for the dataset $D_{!S}$.

Such subset S is confined by the treatment site and treatment modality. In radiation therapy treatment planning, many subsets (treatment sites and modalities) have not been addressed by TPKDD yet. One example is the whole breast radiation therapy (WBRT) using tangential fields. Two near-opposing fields are setup from medial and lateral direction to provide uniform dose distribution within the entire breast volume. Conventionally, physical wedge was used to compensate the missing tissue in the BEV direction. However, hotspot or cold spot was often observed in the irradiated volume. Virtual compensator was then developed, which utilizes the MLC to modulate the fluence in order to achieve a uniform dose distribution within the volume. The hotspot volume can be substantially reduced using fluence modulation than using wedge. Current clinical practice asks planner to manually paint the fluence which is iterative and time-consuming.

1.6.3 Model implementation

Most of previous studies focused on exploring various data mining methods to solve treatment planning tasks. Almost all of them prove to be effective. In order to translate them to clinical implementation, one major question remains to be answered is "how we can guarantee the quality of the model". In this phase of study, we are not only interested in how to build an effective model, but also how to build a good model. We need to inspect the entire workflow of TPKDD, starting from data selection, data processing, data transformation to data mining and model evaluation. Possible research

areas include outlier identification for the model which addresses “data processing”, building the model using cases from multiple treatment sites which addresses “data selection”. Removing outliers could potentially purify the knowledge discovered. Building the model using cases from multiple treatment sites will minimizes the domain expert’s effort in selecting the data for knowledge discovery. And how to accumulate and adapt the knowledge as the database grows is another interesting topic for analysis. This is an important step towards the automated nature of KDD to find the useful information (Soibelman & Kim, 2002).

2. Study focus and the overall structure

This study will investigate new technologies to address the limitations in TPKDD. The structure of the study is divided into three parts. The first two parts will focus on developing new technology to accelerate the treatment planning process using prior knowledge and address the first two challenges of TPKDD. More specifically, in Chapter 3, a novel anatomy feature driven technique is introduced to guide prostate IMRT planning. A dose atlas is constructed based on the anatomical similarity metrics. For a new query case for planning, the anatomy of the case will identify an atlas case to generate 3D dose guidance. In Chapter 4, a novel anatomy and physics rule driven technique is developed to accelerate the whole breast radiation therapy treatment planning. The whole process is composed of three steps. First, an energy selection tool is proposed. The second step is generating initial fluence map using prior knowledge. A third step will follow to fine tune the fluence map to achieve optimal dose distribution based on the physics rules.

The third part of the study will focus on data selection and data preprocessing in the TPKDD process. More specifically, in Chapter 5, a systematic workflow is established to improve the knowledge-based model quality. The workflow will identify geometric novelty and then the dosimetric outlier. A refined model could potentially have better prediction accuracy for the new cases. This workflow helps identify novelty and outlier one by one. In order to identify novelty and outlier in a group fashion, a

clustering-based method is developed in Chapter 6. A one class support vector machine (SVM) followed by ROUT (robust regression followed by outlier identification) is developed to identify and exclude novelty and outlier cases. Chapter 7 will introduce the model tree to build a global model including all cases with different clinical conditions and indications. The model tree will be trained with cases from different treatment sites and compared with the model trained with single site cases. In Chapter 8, an alternative method for training the global model using the clustering-based method is proposed. All cases will be clustered based on the geometric features using the *k*-means algorithm and each cluster builds the model locally. In the last Chapter 9, the knowledge-based model will be trained in an incremental fashion to discover the solution of how to address new clinical cases. The experiment will answer the question when re-training the model is necessary.

3. Development of a novel anatomy features driven automatic IMRT treatment planning for prostate cancer

3.1 Introduction

As discussed in section 1.6.1, the model training efficiency depends on the complexity of the model. The ideal scenario is one-fit-for-all, meaning there exist a reference case that can generate guidance for all other cases. In reality, it is usually impractical due to the large variation in the patient anatomy. However, this concept can be further explored to generate an atlas to guide the treatment planning process. A small case cohort atlas will improve the model training efficiency as compared to other knowledge-based modeling techniques.

Atlas-based medical image segmentation techniques have been successfully implemented to address the anatomical variations between reference and query cases (X. Han et al., 2008; Stapleford et al., 2010; Strassmann et al., 2010; X. Yang et al., 2014). Classifications based on multiple anatomy patterns divided the datasets into various clusters to further enhance the similarity fidelity between reference and query cases (Albano et al., 1978; Jain, Duin, & Mao, 2000). To match a new case to the most similar atlas in such multi-case atlas scheme, pattern recognition was used to find regularities in the image data. Existing segmentations of the 50 atlas were then used to predict the new image's segmentation (Langerak, Berendsen, Van der Heide, Kotte, & Pluim, 2013).

Akin to atlas-based image processing, several adaptive radiation therapy (ART) techniques have employed a similar approach to re-fine the adaptive plans, thus reducing the planning time and therefore accelerating the clinical workflow. By treating the original plan as the 55 reference (atlas), plan parameters were adjusted to re-conform the dose distribution to daily anatomy thus correcting for inter-fractional anatomical variations (Ahunbay et al., 2008; Ahunbay, Peng, Godley, Schultz, & Li, 2009; Ahunbay et al., 2010; T. Li et al., 2011; T. Li et al., 2010; Thongphiew et al., 2009; Q. J. Wu et al., 2008). In this process, the original dose distribution served as single atlas to predict the plan parameters or adjustments corresponding to daily anatomy variation. For example, Ahunbay et al. developed an online adaptive treatment planning technique using segment aperture morphing (SAM) and segment weight optimization (SWO) to adapt IMRT plans for inter-fractional motion correction (Ahunbay et al., 2008; Ahunbay et al., 2009; Ahunbay et al., 2010). Wu et al. and Li et al. also developed a fast re-optimization technique for on-line ART (T. Li et al., 2011; T. Li et al., 2010; Thongphiew et al., 2009; Q. J. Wu et al., 2008). This ART method used the original plan as the base dose distribution and applied a deformable registration technique to generate the *goal dose* distribution for subsequent treatment fractions that require re-optimization. The *goal dose*, which took the daily anatomical variations into account, can guide the optimizer to automatically generate new fluence maps in 1-2 minutes. Plans resulting from this method 70 showed

daily target coverage and OAR sparing comparable to those planned by expert planners (Q. J. Wu et al., 2008).

As an extension to the aforementioned atlas-based ART methods, which were shown to work well for correcting plans for inter-fractional variation for a single patient, prediction of dose 75 across patients for general treatment planning is challenging because inter-patient anatomical variations are often more substantial. The present work investigates the feasibility and benefit of developing and using a multi-dose atlas for IMRT planning. To address the inter-patient anatomical variation, pattern recognition was incorporated to classify the anatomy into multiple clusters to increase the fidelity in matching anatomy and thus dose patterns. Patient 80 anatomical variations were represented via distinguishing patterns and classified into clusters. Under this framework, a new case was first classified to an anatomy cluster. The representative dose of that cluster, the atlas dose, was then warped to the new anatomy via deformable registration. The warped dose served as guidance for subsequent treatment planning.

The purpose of this research is to offer an alternative that does not require large plan library to begin with. Departing from aforementioned methods, shape analysis was the basis used for creating guidance for treatment planning in this study. The pattern recognition originating from image segmentation was introduced to treatment planning. The introduction of multi-dose atlas, combined with shape analysis of patient anatomy, aimed at providing a new and simpler solution to accelerating the treatment planning

process for new patient while guaranteeing clinically acceptable plan quality. The authors believe this is the first attempt using shape pattern recognition to create and use atlas for guiding treatment planning.

3.2 Materials and methods

Ninety prostate IMRT cases were retrospectively studied under an IRB protocol. 70 cases were randomly selected to build the multi-dose atlas and the remaining 20 cases were used to evaluate the performance of the AGP technique. The CTV included both the prostate and the seminal vesicles (SV). The PTV was expanded from the CTV with a 5 mm margin.

The workflow for the AGP (Figure 3) consists of two major parts: (1) construction of the multi-dose atlas and (2) atlas-guided planning. The first part was performed only once, while the second part was performed for each new patient and included query-atlas matching, deformable dose warping, and automated plan optimization. Each component is explained in detail in the following sections.

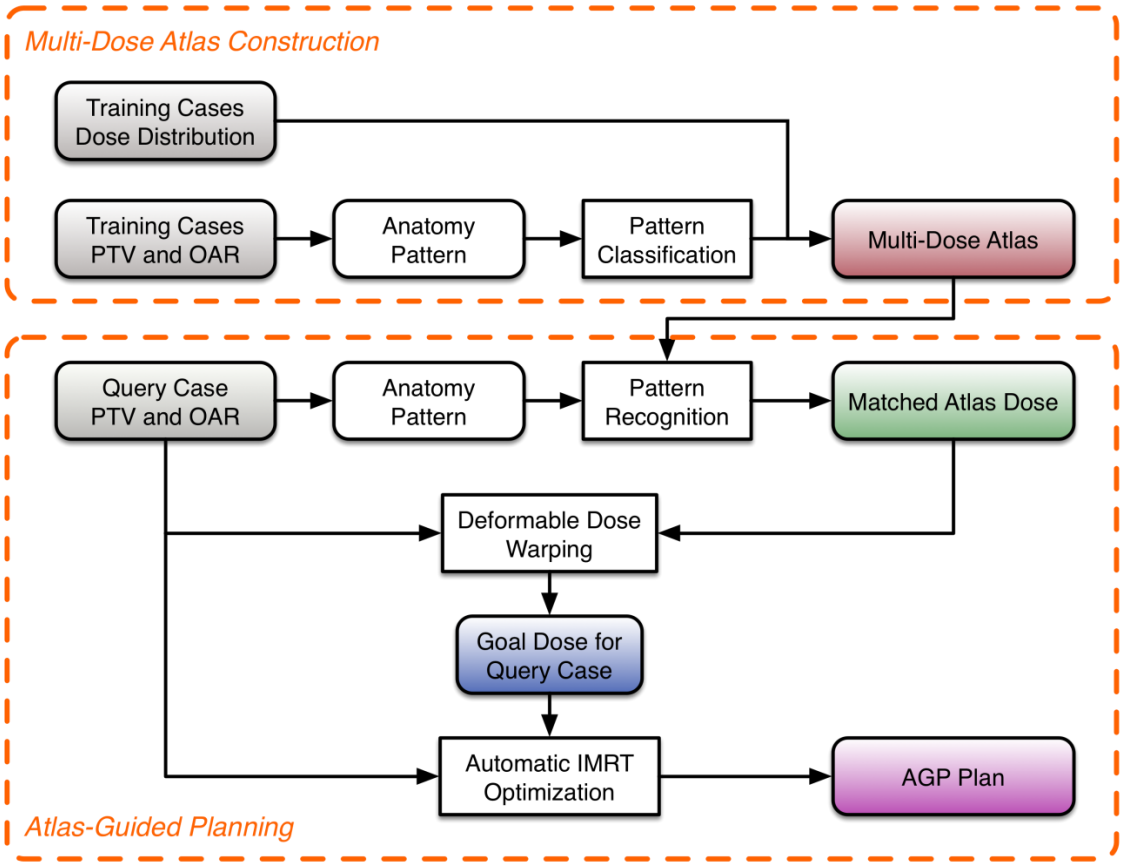

Figure 3: Flowchart of the AGP technique.

3.2.1 Construction of the multi-dose atlas

Multi-dose atlas was chosen to reduce the demand of transforming reference high conformal dose distributions across the large anatomy variations in the patient population. The training cases were classified based on their anatomy shape patterns, rather than the similarity of volumes. The shape pattern analysis considered the topology of different anatomies, and condensed them to a pattern map that summarizes the overall anatomy variations. In this study, two patterns were chosen to distinguish and represent anatomical variations among prostate cancer patients. As shown in Figure

4(a), the first pattern measured the concaveness of the posterior wall of the PTV, encoding dose falloffs around the target in this region. Hunt et al. found that the concaveness of the PTV was strongly associated with normal tissue sparing. We have also found that for prostate IMRT, the PTV-rectum boundary shape was a significant anatomical feature affecting dosimetry. To extract this shape parameter, the PTV was partitioned into two components: PTV_SV (SV including the 5 mm margin except inferiorly where the prostate and the SV are connected), and PTV_prostate (the remaining portion of the PTV). For PTV_SV, the axial slice with the most posterior point $(x_{SVx\_extreme}, y_{SVx\_extreme})$ was located. On the same slice, (x_{SVmin}, y_{SVmin}) and (x_{SVmax}, y_{SVmax}), the two points where the PTV_SV contour intersected with the anterior-posterior (AP) axis, were also located. This first pattern was defined as the PTV_SV concaveness angle θ located between the AP axis and the line connecting points $(x_{SVx\_extreme}, y_{SVx\_extreme})$ and (x_{SVmax}, y_{SVmax}) (Equation (4)).

$$\theta = arctan\left(\frac{|y_{SVx\_extreme}|}{|x_{SVx\_extreme} - x_{SVmax}|}\right) \qquad \text{Eq.(4)}$$

Hence, θ indicates the concaveness of the PTV posterior wall, and is a determining factor for the dose falloff/gradient around the anterior rectal wall.

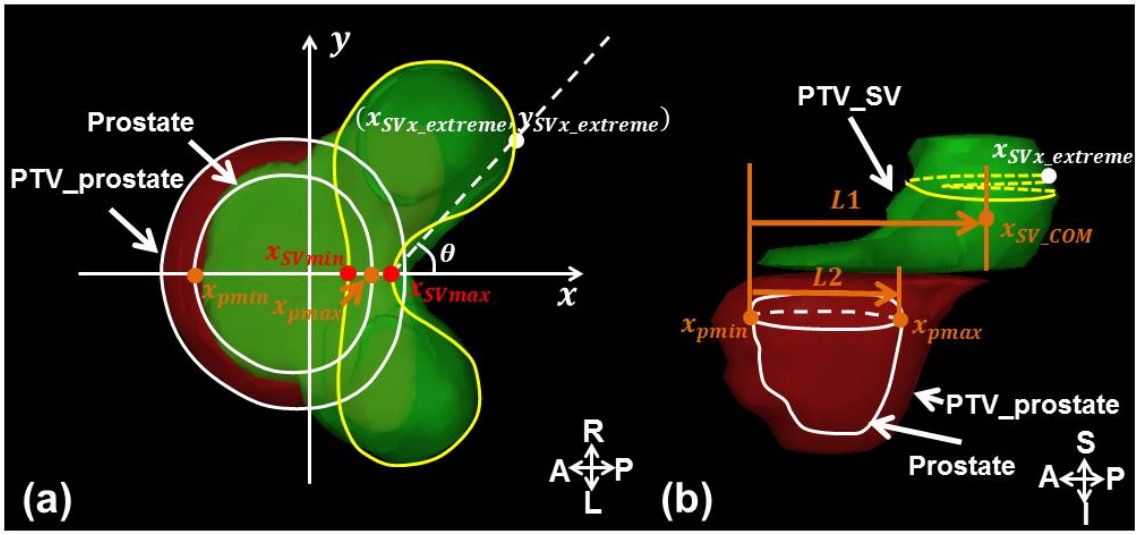

Figure 4: (a) Axial view showing the slice where the PTV_SV (yellow contour) has the most posterior point of the SV; the concaveness angle was defined as the angle between AP axis and the dashed line connecting the distal point $(x_{SVx\_extreme}, y_{SVx\_extreme})$ and the point (x_{SVmax}, y_{SVmax}); (b) Sagittal view through a central slice depicting the PTV_SV (green shadow), which is the SV plus the 5 mm margin (except inferiorly), and the PTV_prostate (red shadow). The Percent Distance to the Prostate (PDP) was defined as the ratio of L1 (the AP distance between the SV COM and the prostate anterior border) and L2 (the AP dimension of the prostate).

The second pattern focused on the superior-inferior direction, summarizing the overall drop of the SVs from the prostate along the AP direction (Figure 4(b)). The center of mass (COM) of the SV (x_{SV_COM}, y_{SV_COM}) and the range of the prostate in the AP direction, x_{pmin} and x_{pmax}, were calculated. The relative position of the SV with respect to the prostate was defined as the distance between SV's COM to the prostate anterior border, relative to the AP dimension of the prostate. This metric was referred as the Percent Distance to the Prostate (PDP) (Equation (5)).

$$PDP = \frac{|x_{SV\_COM} - x_{pmin}|}{|x_{pmax} - x_{pmin}|} \times 100\% \qquad \text{Eq.(5)}$$

The PDP and the concaveness angle θ formed a 2-dimensional pattern map, with each axis representing one anatomy pattern and each case's position on the map determined by the numerical values of the pattern. The *k*-medoids algorithm (Kaufman & Rousseeuw, 1987) classifies the training cases by minimizing the sum of the intra-cluster Euclidean distance (Equation (6)) of the patterns,

$$arg\ \min_S \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - S_i\| \qquad \text{Eq.(6)}$$

where S is the selected training set of atlas, x_j is the case in the cluster S_i. Generally, a medoid is defined as the object in a cluster, with the average distance to all the other objects in the same cluster being the minimal. Each medoid is the representative case for its cluster and hence the atlas case.

To determine the optimal size *k* of the atlas, the average silhouette width for each *k*-medoids scheme was calculated. The silhouette value for each data (Equation (7)) was defined by Rousseeuw (Rousseeuw, 1987):

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \qquad \text{Eq.(7)}$$

where $a(i)$ is the average dissimilarity of i with all other data in the same cluster, and $b(i)$ is the lowest average dissimilarity of i to any other cluster. The Euclidean distance between two data points on the pattern map was used as the similarity measure. The average silhouette width, the arithmetic mean silhouette value of all data, reflects how well the data were partitioned and clustered with higher average silhouette width corresponding to better classification (Rousseeuw, 1987). Similar clustering algorithm, the *k*-means clustering algorithm, was used in leaf sequencing by breaking the optimal fluence into smaller groups or clusters in Pinnacle treatment planning system (Philips Healthcare, Andover, MA).

With the *k*-medoids classification, the final classification result was dependent on the selection of *k* objects as the initial medoids to start the classification iteration. To avoid the iteration from being trapped in a local minimum, the classification procedure was repeated 5~6 times to find the final medoids with the lowest sum of within-cluster distances. A different set of randomly selected initial medoids was used for each repetition.

3.2.2 Atlas-guided plan generation

The remaining 20 cases outside the training pool were used as query cases to evaluate the AGP technique. The process of applying AGP to a new prostate case is described in the following steps:

3.2.2.1 Matching the query case to the atlas

First, the query case's anatomy pattern was extracted the same way as described in 2.1 and the PDP and θ values determined the case's position on the pattern map. The closest atlas was then matched to the query case.

To account for local and fine variations between the query and atlas cases, deformable image registration (T. Li et al., 2011; T. Li et al., 2010; Thongphiew et al., 2009; Q. J. Wu et al., 2008) was performed to enhance anatomy-dose correlations between the query and atlas cases. In this study, the MIM Maestro™ system (MIM Software Inc, Cleveland, OH), which uses the free-form deformation algorithm, was used to deform the matched atlas case's anatomy onto the query case's anatomy using contour-based deformable image registration (Q. J. Wu et al., 2008). The resulting 3D deformation vector fields from the registration process were applied to warp the atlas dose distribution towards the query case's anatomy. Because the deformation vector fields encode the local variation between the atlas case's anatomy and the query case, the warped dose follows the same transformation and conforms to the query case's anatomy (T. Li et al., 2011; T. Li et al., 2010; Q. J. Wu et al., 2008). This dose was referred to as the *goal dose,* representing a highly conformal 3D dose distribution to be achieved subsequently using inverse planning (Q. J. Wu et al., 2008).

3.2.2.2 Atlas-guided planning

In our routine clinical planning protocol, inverse planning or optimization objectives are specified as dose-volume objectives. Following the same planning protocol, the AGP optimization objectives were sampled from the *goal dose* and imported into the Eclipse (Varian Medical Systems, Palo Alto, CA) TPS. The optimization process was subsequently run without human intervention. In this study, fluence-based optimization (Dose Volume Optimizer, DVO) provided in our TPS was used. The AGP technique provides the planner with the dose-volume objectives to start the fluence optimization.

Existing clinical plans for the query cases, which were manually designed for clinical treatment, were used as benchmarks to evaluate AGP plan quality. Beam configurations for both sets of plans follow our institutional template. The only difference between the generation process of AGP plans and clinical plans is that for AGP plans the dose-volume objectives are generated automatically and fixed during optimization, whereas for clinical plans the dose-volume objectives are manually generated and adjusted during optimization. Apart from the difference in human intervention, the algorithms for fluence optimization and leaf sequence calculation are identical between AGP and clinical plans.

3.2.2.3 AGP plan quality evaluation

First, the spatial dose distribution was visually inspected by an experienced physicist to identify any hotspots in sensitive OARs to ensure that the overall dose distributions met clinical treatment quality. If any of the hotspots were located in unfavorable locations, the AGP plan was noted as a major deviation from the clinical planning quality.

Quantitatively, the paired AGP and clinical plans were compared. All dosimetric comparisons were tested for significance using the one-sided Wilcoxon Signed Rank test. The null hypothesis was that the plans generated by AGP were not inferior to the clinical plans in the dosimetric parameters. A p-value below 0.05 was considered statistically significant.

Key DVHs parameters of the PTV and two main OARs (the bladder and the rectum) were also evaluated. For PTV, the homogeneity index (HI) (Q. Wu, Mohan, Morris, Lauve, & Schmidt-Ullrich, 2003) was analyzed,

$$HI = \frac{D_2 - D_{98}}{D_{R_x}} \times 100\% \qquad \text{Eq.(8)}$$

where D_2 and D_{98} are doses to 2% and 98% of the PTV volume, respectively, and D_{R_x} is the prescription dose. The conformity index (CI) described by Paddick (Paddick, 2000) was compared between the paired plans,

$$CI = \frac{PTV_{PIV}^2}{PTV \times PIV} \quad \text{Eq.(9)}$$

where PTV is the planning target volume, PIV is the prescription isodose volume and PTV_{PIV} is the prescription isodose volume in the PTV. The CI indicates how well the prescription dose (and similarly high dose) conforms to the PTV.

The OAR sparing was compared using the generalized equivalent-uniform-dose (gEUD) defined by Niemierko (Niemierko, 1999) (Equation (10)).

$$gEUD = \left(\sum_i v_i D_i^a\right)^{\frac{1}{a}} \quad \text{Eq.(10)}$$

where v_i, D_i correspond to the percent volume of each voxel and the absolute dose of each voxel in the OAR, respectively, and the a value was set to 6 for both the bladder and rectum. It was used as a summary metric of the entire DVH curve and as an overall sparing quality index (Censor, Bortfeld, Martin, & Trofimov, 2006).

Additional dosimetric parameters, such as volumes receiving 100% and 65% of the prescription dose, were compared. These parameters represent specific dose ranges that physicians examine when they review plans and are used in our clinical protocol for plan quality assessment. Finally, the combined monitor units (MUs) of each plan were compared to assess delivery efficiency.

3.3 Results

3.3.1 Atlas construction

70 cases were classified using the *k*-medoids algorithm. The average silhouette width was calculated for various classification schemes (Figure 5(a)). An atlas consisting of 5 cases was chosen as its classification had highest average silhouette width. The result atlas is shown in Figure 5(b), where the 5 atlas cases are shown in solid-red dots. The rest of training cases (i.e. non-atlas cases) are shown in black circles, connected to their atlas cases or medoids with the red lines. Further, the validation cases are plotted on this map as solid-blue dots (Figure 5(c)), and the red lines indicate the atlas cases they were matched to.

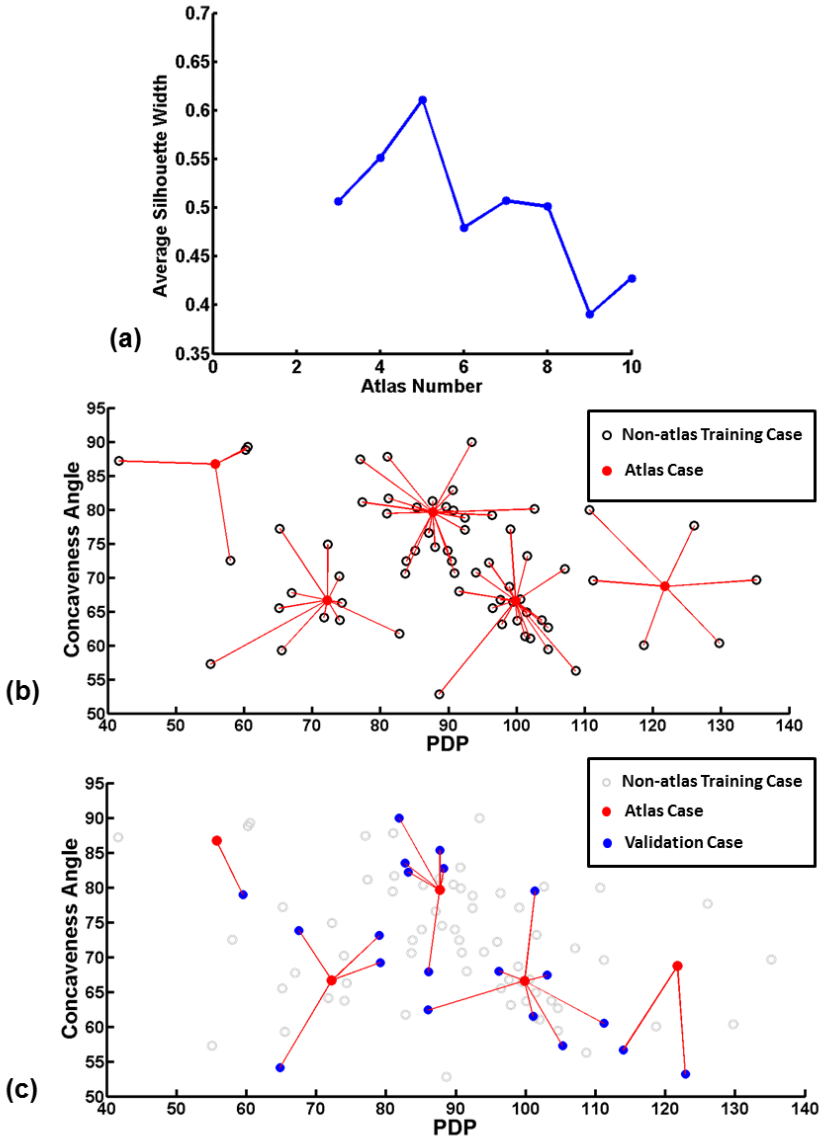

Figure 5: (a) The average silhouette width for classifications with different class numbers. A higher average silhouette width corresponds to a better classification. (b) The anatomical pattern map for all 70 training cases used for classification and atlas selection. 5 out of the 70 cases were selected as atlas cases (solid-red dots) based on the *k*-medoids classification results in (a). The rest non-atlas cases (black circles) are connected to their corresponding atlas cases. (c) 20 test cases (solid-blue dots) are plotted on top of the classification in (b) where non-atlas cases are shown as grey circles. The red lines connecting the atlas and test cases indicate the atlas-query case matching.

3.3.2 Goal dose distribution

Figure 6 shows an example of a query case (b) matched to an atlas case (a) and the 3D-warped *goal dose* (c) from the atlas to the query case anatomy. As shown, the selected atlas and query case were similar in key features, e.g. displaying similar concavity of the PTV posterior border, but not necessarily in volume (PTV volume of atlas vs. query case: 118.4 cc vs. 193.0 cc).

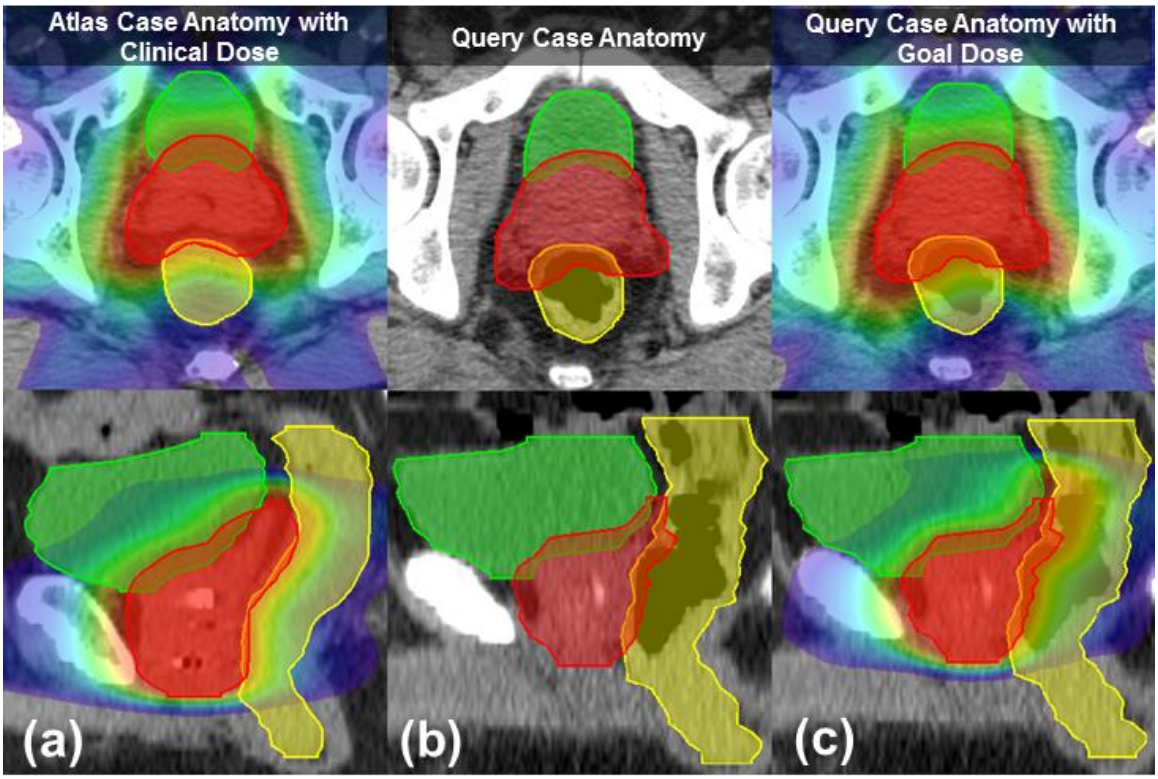

Figure 6: Example of a goal dose warped from the atlas dose. The bladder, rectum, and PTV volumes are delineated in green, yellow, and red, respectively. (a) The atlas case's anatomy with clinical dose distribution. (b) The query case's anatomy with similar shape pattern to the matched atlas case. (c) The query case's anatomy with the goal dose warped from the atlas dose.

3.3.3 PTV coverage and OAR sparing

Visual inspection of the AGP plans showed no major deviations from typically observed clinical plan quality, i.e. no excessive hotspot was observed in sensitive OARs. Figure 7 shows 3 plans that describe the range of observed AGP performance in this study. Overall, the AGP plans provided similar high dose sparing for the OARs as the clinical plans, which was consistent throughout the query case cohort. For the medium to low dose regions, variations were observed in some cases. For case #07 (Figure 7(a)), the AGP plan displayed better bladder sparing in the medium dose range. The AGP plan for case #11 (Figure 7(b)) showed worse low dose sparing for the rectum. In most cases (case #16 shown as an example in Figure 7(c)) similar plan quality was observed between two plans.

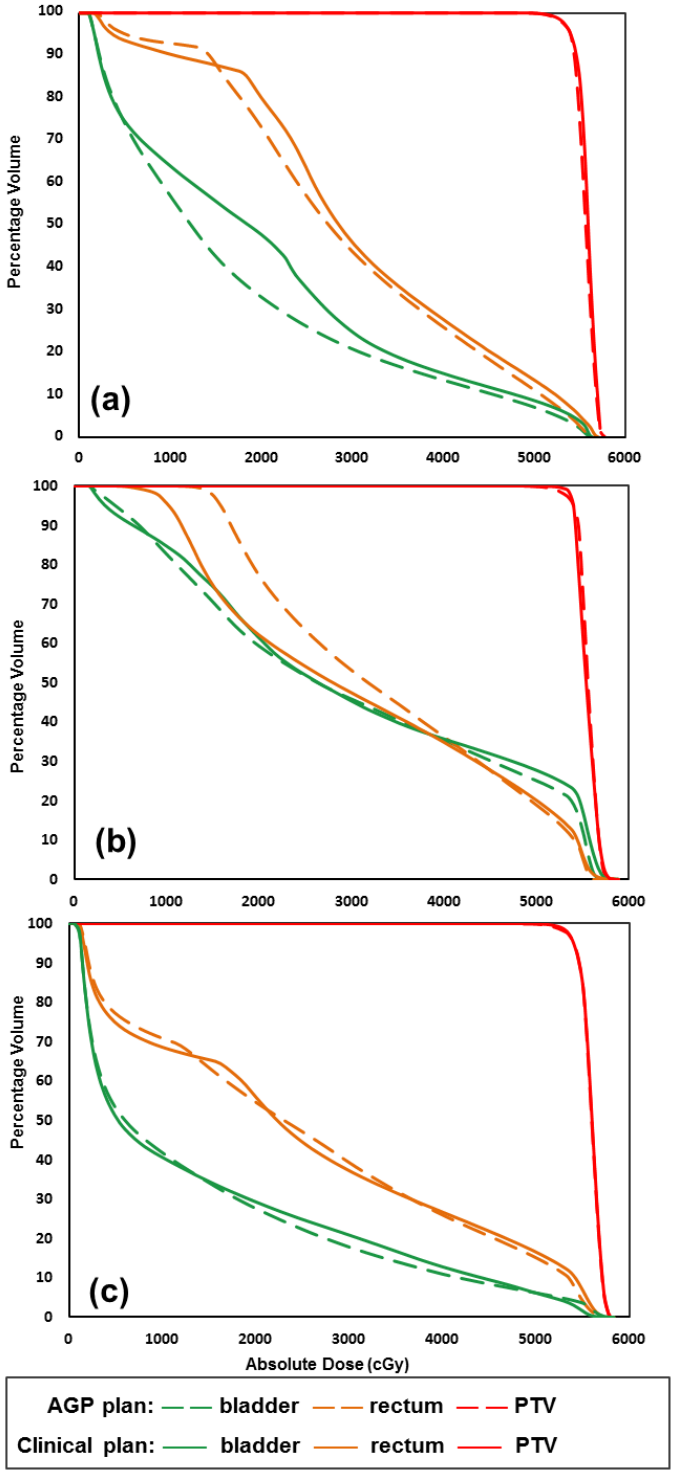

Figure 7: Example DVHs from the AGP and clinical plans.

Comparison of the dosimetric data and the corresponding *p*-values are shown in the boxplots in Figure 8. The homogeneity index and the conformity index of the AGP plan were not statistically significantly inferior to that of the clinical plan (p=0.21, p>0.99). The overall OAR sparing in the AGP plan, in terms of bladder gEUD and rectum gEUD, was not statistically significant: p>0.99 for bladder and p=0.99 for rectum. The 100% prescription dose volume for both the bladder and rectum (V100%) of the AGP plan were not significantly inferior to those of the clinical plan: p>0.99 for both the bladder and rectum. The V65% for the bladder and rectum also showed no significant difference (p=0.69, p=0.77). The two techniques produced plans that were clinically similar.

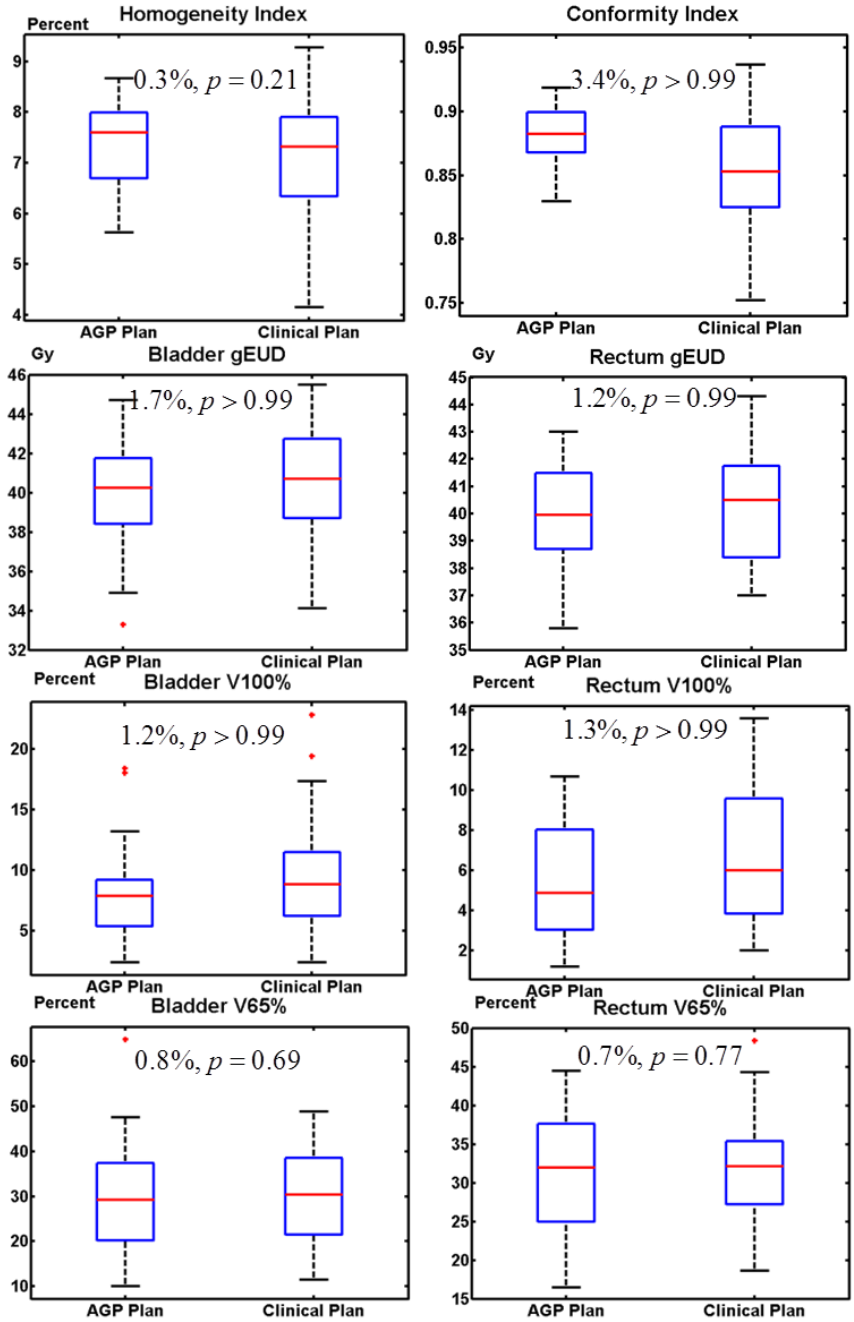

Figure 8: Boxplots of dosimetric parameters for AGP and clinical plans. The percent difference and the statistical significance between two plans were labeled.

3.3.4 Planning time and monitor unit

With the optimization objectives derived from the *goal dose*, the final AGP plan generation was automated with 70 continuous iterations taking about 1 minute to complete (Eclipse 11). The MUs for the AGP and clinical plans were similar: 116±24 MU and 124±30 MU (p=0.93), which corresponds to similar delivery time and leakage radiation.

3.4 Discussion

In this study, an atlas-guided planning technique was developed and investigated for its feasibility and benefit in prostate IMRT planning. Based on the classification results, a 5-case atlas was constructed and then used to guide treatment planning. The quality of plans generated with atlas guidance was compared to that of clinical plans.

The multi-dose atlas was used to predict *goal dose* to guide the inverse planning/optimization process. The choice of multi-dose atlas reduces the demand of transforming highly conformal dose distributions across the large anatomy variations in the patient population. Similar to image segmentation, the power of deformable image registration is limited because the assumption that the dose distribution follows the deformation of anatomy does not hold when anatomical variations become too large. For example, the deformation from a large to a small PTV_SV concaveness angle (i.e. from a round-shaped PTV to a concave-shaped PTV) will likely to introduce non-

realistic dose gradient around the PTV near the rectal wall. Therefore, the deformable registration with a single reference case could be less favorable when large anatomical variations exist especially across different patients.

The *goal dose* is a complete 3D dose distribution and can also be directly used to generate IMRT plans without extracting them into DVH objectives. Thongphiew et al. (Q. J. Wu et al., 2008) utilized a voxel-based optimization algorithm formulated as a linear goal programming (LGP) model (Chankong & Haimes, 1983), which was especially suitable for fast on-line plan re-optimization. Incorporating such algorithms in our technique may potentially further reduce overall treatment planning time, but it is currently not compatible with the clinical TPS that we are using. Further, IMRT was chosen as the treatment technique for prostate cancer in this study. However, the AGP technique is not just specific to IMRT and can also be directly applied to any treatment technique which calls for an inverse treatment planning algorithm such as volumetric modulated arc therapy (VMAT) and TomoTherapy® (Accuray, Inc., Sunnyvale, CA).

The anatomy-based multi-dose atlas and the deformable image registration driven dose transfer are the two essential components of this AGP technique. The shape pattern analysis considers the topology of different anatomies, and condenses them to a pattern map that summarizes the overall anatomy variations. The deformable image registration-driven dose transfer further fine tunes the local-regional anatomy-dose conformity match and eliminates the need to search for an exact match between atlas

and query anatomies. The combination of these two techniques in the AGP technique is able to provide accurate dose guidance with only a 5-case atlas and to guide the optimization process with full automation. Although some OAR sparing variations in the low dose region are observed, overall the results show comparable plan quality to that of the clinical plans, making the AGP technique a promising treatment planning tool with improved efficiency while maintaining similar plan quality.

This study used two anatomical features to classify the prostate anatomy into atlas. The two-feature descriptor simplifies the prostate target geometry to two dimensions. This simplification could potentially add to the error in the dose guidance. It is possible that different prostate anatomies result in the same 2D feature vectors. Higher dimension features could further regulate the anatomy similarity among cases. This could result in more than 5 clusters. In addition, current method did not consider the OAR information which could affect the dose gradient. The dose gradient would be steeper when the OAR is very close to the treatment volume. The OAR information could also be incorporated into the clustering process. The atlas may provide either "tight" dose contour or the "normal" dose contour for the new case. Future study is warranted to incorporate these features.

There are several confounding factors which affect the study. First, the proposed atlas-guided method is applicable to prostate IMRT only. The two-feature descriptor is specifically tailored towards the prostate anatomy. The atlas concept can be generalized

to other treatment sites but it requires hand crafting the site specific features. Secondly, although there was no statistical significance observed for the paired test of the dosimetry metrics, the test could be confounded by the planning parameters, such as the dose volume objective weights etc.

3.5 Conclusion

Atlas-guided treatment planning is feasible and efficient. Atlas-based patient-specific dose distribution objectives can effectively guide the optimizer to achieve similar quality when compared to the clinical plans.

4. Improving the treatment planning efficiency for the Whole Breast Radiation Therapy with a novel physics based rules driven technique

4.1 Introduction

Breast cancer has long been the most popular cancer among female. There were 231840 estimated new female breast cancer cases in 2015 (American Cancer Society, 2015). Depending on the staging of the cancer, lumpectomy, mastectomy, radiation therapy and chemotherapy may be operated concurrently. Lumpectomy and mastectomy removes palpable tumor cells and radiation therapy usually follows to control local recurrence. However, lumpectomy and mastectomy nowadays is less favorable due to cosmetic reasons. For early stage breast cancer, breast preservative treatment has been more and more popular. Under this scenario, whole breast radiation therapy (WBRT) is often adopted.

WBRT deploys multiple modalities including 3D, field-in-field, IMRT and VMAT. Conventionally, 3D treatment utilizing the physical wedge was the major treatment method. The planner determines which wedge to use and then adjust the weighting of both beams (medial and lateral beam) to achieve uniform dose distribution. Since the tunable parameter for 3D treatment is limited, inhomogeneous dose distribution is often observed. The maximum dose for 3D plan could be beyond 110% prescription dose. A large volume of hotspot may result in cosmetic issues, such as skin darkening. This is the intrinsic drawback of 3D treatment plan. In order to tackle the

inhomogeneity issue, field-in-field (FiF) treatment was developed. Similar as 3D treatment, FiF uses beam shaping devices (jaws, MLCs) to control the delivered radiation. FiF is advantageous over 3D treatment since it utilizes multiple segments to control the effective delivered fluence to the patient as compared to 3D treatment with only 2 segments. FiF substantially reduces the inhomogeneity in the irradiated volume. The combination of several segments is essentially equivalent to IMRT using step-and-shoot. However, FiF does not call fluence optimizer and is therefore a forward planning process.

IMRT and VMAT have also been adopted to treat breast cancer patients. Fluence modulation is favorable for breast irradiation due to the shape change in anatomy. However, the inverse planning process makes the task less intuitive, since setting the objectives in optimization space can hardly be directly linked to the dose to a certain volume. And sometimes, the optimization engine cannot meet certain objectives. The overall dose distribution is acceptable with some extra effort needed to remove hotspot volume. TPKDD has set foot on breast IMRT in recent years' research. Purdie et al. developed an automatic breast IMRT planning system which uses heuristic methods (Purdie et al., 2014). In general this method mimics the human planning process. The planner only needs to specify clinical requirements. VMAT is also applicable for breast treatment. However, due to the intrinsic gantry rotation nature, low dose spill to ipsilateral lung and heart is unavoidable. Limiting beam edge to control the dose spill to

the lung and heart is an easy manner to operate, which leads to the treatment using tangential field.

WBRT using tangential fields has been an increasing interest regarding the dose homogeneity within the irradiation volume and dose spill to OARs such as lungs and heart (Al-Rahbi et al., 2013; Caudell et al., 2007; Flejmer, Josefsson, Nilsson, Stenmarker, & Dasu, 2014). This method is also often referred as Electronic COMPensation (ECOMP) (Chui, Hong, Hunt, & McCormick, 2002; Evans et al., 1998). It calls MLCs to modulate the fluence therefore it is an IMRT process (Emmens & James, 2010). On the other hand, it does not call the inverse planning optimizer to modulate the fluence. The planner has to manually fine-tune the fluence for both beams (medial and lateral beams) to achieve uniform dose distribution. In Eclipse TPS, the planner can start with an initial fluence generated from "irregular shape compensator". Then the fluence fine-tuning process is iterative and time consuming, and sometimes the planner has to restart the planning process with different energy due to unachievable dosimetric constraints. High quality treatment planning for WBRT using tangential fields currently requires manual fluence editing that may take 1-4 hours, and the plan quality is highly dependent on the planner's experience. Given large volumes of breast cancer patients, plan turnaround time reduction is critical. This study addresses this issue by developing an ultra-fast high quality WBRT planning system which enables near-real-time and interactive planning while providing similar plan quality as human planners.

4.2 Materials and methods

A total of 40 IRB approved Whole breast RT plans were included in this study. 20 patients treated with single energy (SE, 6MV, 10 patients) or mixed energy (ME, 6/15MV, 10 patients) were randomly selected for model training. Additional 20 cases were used as validation cohort.

The WBRT planning system consists of three major steps: energy selection, fluence estimation and fluence fine-tuning. The proposed method tackles the most time-consuming component in manual planning which is fluence fine-tuning while providing an automated streamline for current task. It starts with an automated energy selection module, followed by an automated fluence map generation based on the anatomy of the patient. Then a sequential fluence tuning based on physics principals adjusts the beamlet fluence to achieve uniform dose distribution within the breast volume. The flowchart of the process is shown in Figure 9. Current clinical practice workflow is also shown in Figure 9 for comparison.

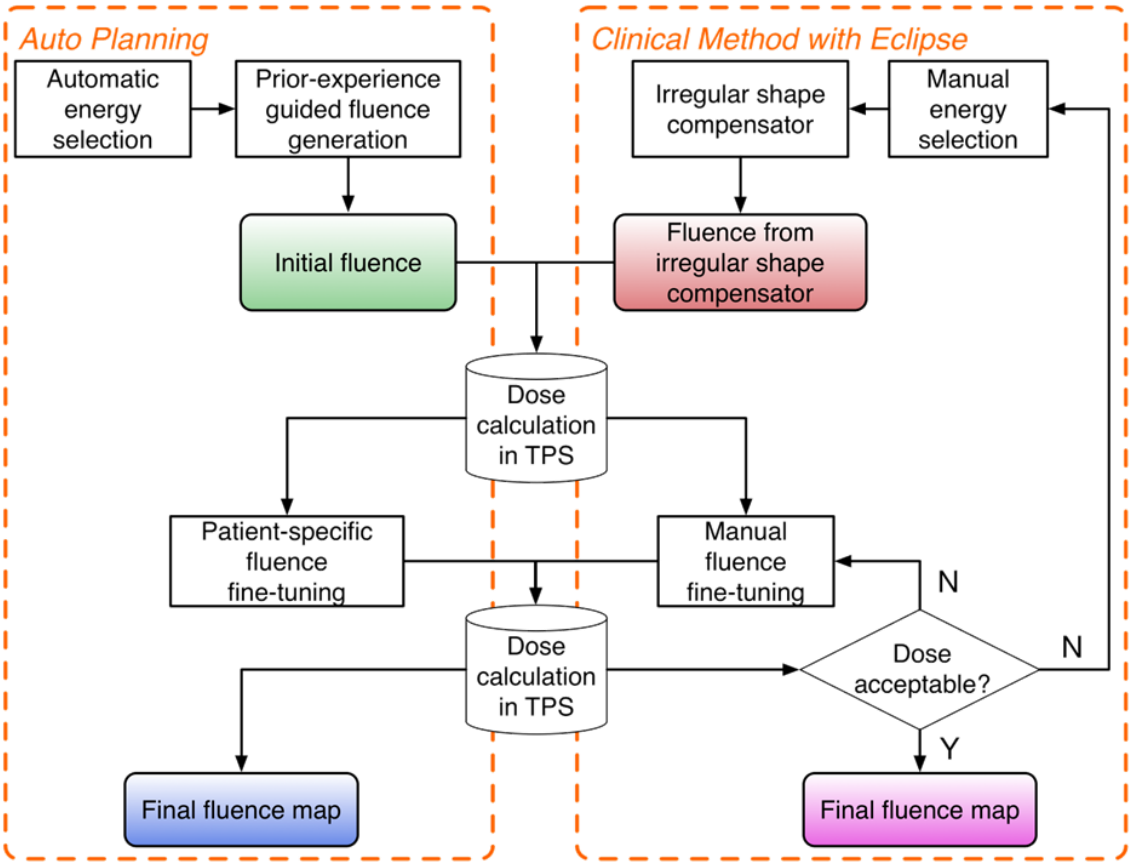

Figure 9: Flowchart of auto planning (left) and current clinical practice (right).

4.2.1 Energy selection

The automated process starts once the beam geometry is set up. Human planners may decide the beam energy for large or small breast patient but may lack confidence in making such decision for moderate-sized breast patients. In clinical practice, planner may provide both plans, i.e. single energy 6MV or mix energy 6/15MV, for physician to choose. Lack of energy selection guidance further adds workload to the planner and thus impedes the clinical workflow. To answer this question, prior plans were used to classify the query case for energy selection. First, the digital reconstructed radiograph

(DRR) was generated for each beam and the DRR intensity histogram within the irradiation volume was calculated for each patient. Second, principal component analysis (PCA) was performed and the first two component scores were used to represent each case.

4.2.2 Fluence estimation

The second step is generating fluence map to achieve optimal dose distribution. Human planners iteratively refine the fluence to achieve such goal. Alternatively, Eclipse TPS offers a module to generate the electronic compensator which is called "irregular shape compensator". It generates uniform dose distribution at specified depth (e.g. 40% depth). However, due to interplay effect in 2D which is primarily caused by beam divergence, the dose distribution is no longer homogenous in the entire volume. A machine learning algorithm was proposed to summarize and learn the correlation between anatomical features and the optimal fluence map. Several shape based features were identified as fluence map predictors including DRR intensity, penetration depth, penetration tissue inverse square factor, penetration depth in tissue excluding lung etc. Example feature map is shown in Figure 10. A Random Forest (RF) model was proposed to summarize the relationship between input variables (shape based features) and the output variable (fluence). The RF model assesses the effect of each predictor on predicting the output variable. A RF was trained using training cases with 150 trees. The RF first samples with replacement from all training data for each tree; then a tree is built

for each individual sample; finally the prediction result from each tree is aggregated and generates the output.

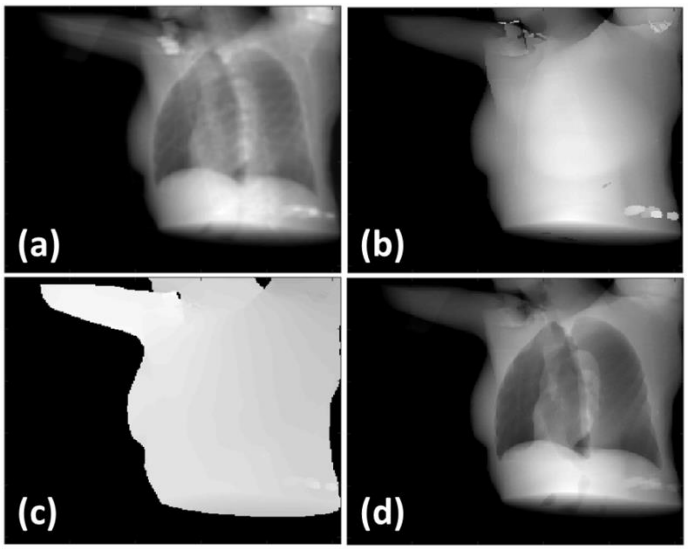

Figure 10: Example feature map included in RF: (a) DRR; (b) penetration depth; (c) inverse square factor; (d) penetration depth in tissue excluding lung.

4.2.3 Fluence fine-tuning

The fluence map generated from the RF model inherits the plan quality from the training cases. However, physician may have patient-specific requirement for target coverage or hotspot volume control. The proposed third step offers the physician to interactively fine-tune the 3D dose distribution. This step starts with finding reference points in the breast tissue, followed by reference point centrality correction and fluence fine-tuning. The flowchart is shown in Figure 11.

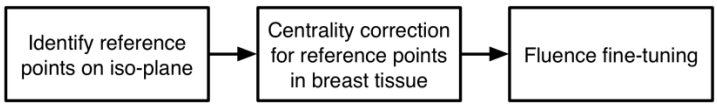

Figure 11: Flowchart for patient-specific dose fine-tuning.

Reference points were identified on the iso-plane in the irradiated volume (shown in Figure 12(a)). Centrality correction was then performed to balance the penetration tissue from either side for reference points in the breast tissue (shown in Figure 12(b)). Baseline dosimetric parameters to these points were learned from prior plans and these parameters can be tuned towards specific coverage requirement or hotspot control. The baseline parameters for reference points near skin and near chest wall are shown in Figure 12(c) and (d). These parameters were used in the following fluence fine-tuning.

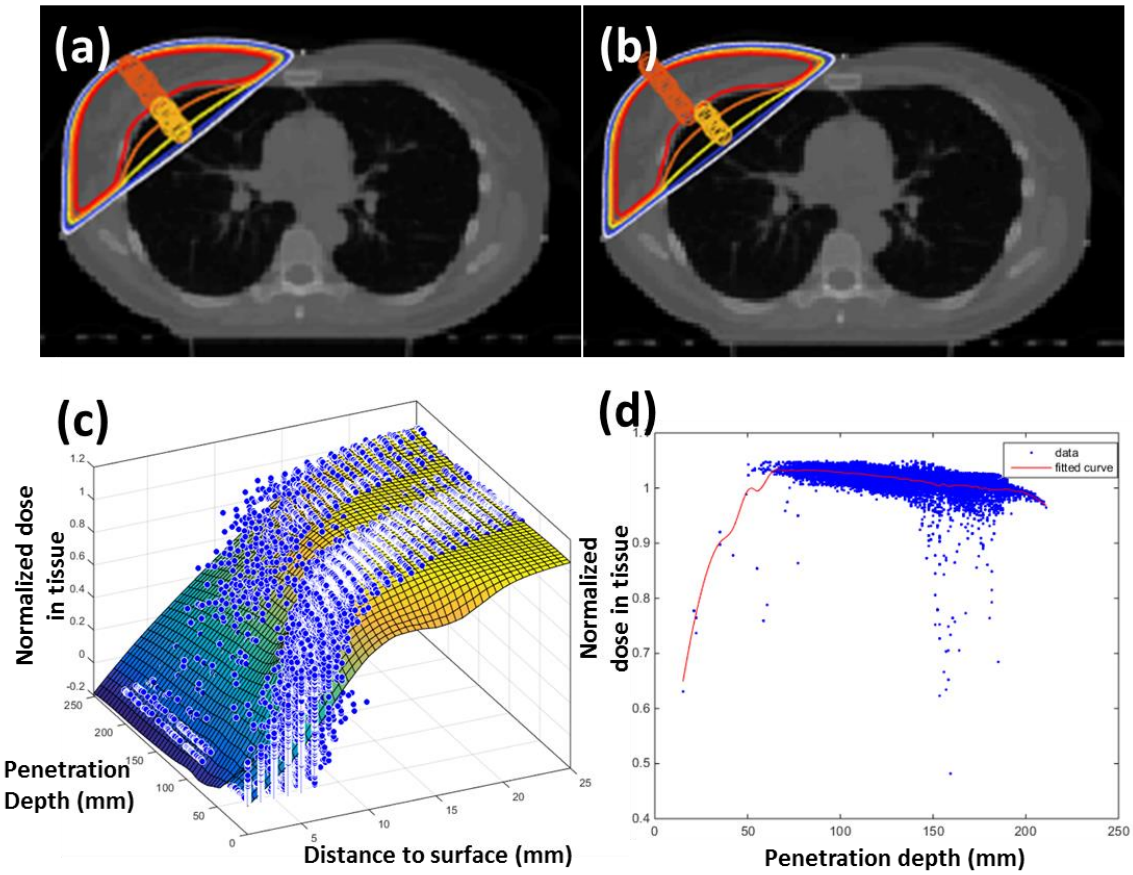

Figure 12: (a) example reference points located on the iso-plane; (b) reference points in the breast tissue after centrality correction; (c) dose in the tissue near skin extracted from prior plans; (d) dose in the tissue near chest wall extracted from prior plans.

The fluence fine-tuning started from the beamlet that connects the radiation source (shown as the start point in Figure 13). The beamlet was tuned radially from the start point to peripheral. The tuned fluence should satisfy the following condition:

$$f = \operatorname{argmin} \sum_i (D_i(f) - D_{i,model}) \qquad \text{Eq.(11)}$$

where i is reference point, $D_i(f)$ is the dose at reference point i after fluence tuning, $D_{i,model}$ is tunable dose parameter shown in Figure 12(c)(d). This condition only considers primary beam component and treats the scatter component as negligibly effective to the adjacent beamlet. The proposed method was validated for feasibility under the approximation assumption. The result dose distribution was analyzed for whether the preset patient-specific parameter had been effectively reflected.

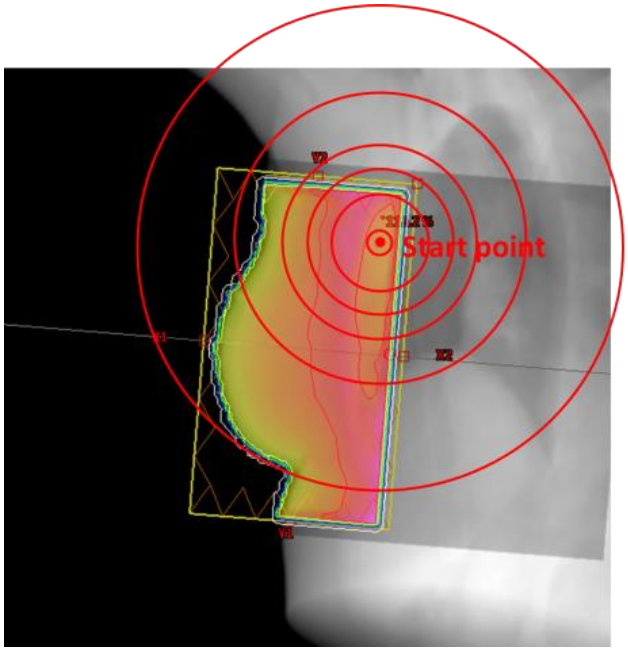

Figure 13: Illustration of fluence tuning path in the beam-eye-view. The fluence tuning starts from the center (start point) and spreads radially.

All optimization code has been implemented in C++ to reduce the calculation time. The proposed method was validated by comparing the auto plans with manually generated clinical-plans using Wilcoxon Signed-Rank test. The null hypothesis was that the auto plans were not inferior to the clinical plans in the dosimetric parameters.

Dosimetric comparison included PTV V100%, PTV V105%, V105cc, lung V10Gy, lung V20Gy and lung V95%.

4.3 Results

The PCA analysis result is shown in Figure 14. The DRR intensity histogram within the irradiation volume for each patient is shown in Figure 14(a). The first two principal components are shown in Figure 14(b). In Figure 14(c), blue dots represent single energy cases and red dots represent mix energy cases. PC1=0 served as a good classifier with an accuracy of 19/20 for the validation cohort, meaning the model suggested the same energy combination as the clinical plans.

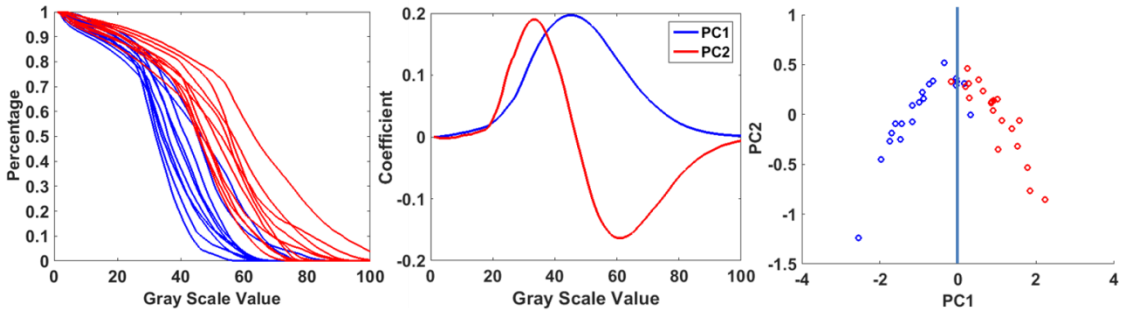

Figure 14: (a) DRR intensity histogram for single energy cases (blue) and mix energy cases (red); (b) PC coefficient for PC1 (blue) and PC2 (red); (c) PC1 and PC2 score for single energy cases (blue) and mix energy cases (red), PC1=0 is shown in blue.

For the fluence estimation RF model, the prediction error for the training data as a function of the number of trees is shown in Figure 15(a). The prediction error for all training data using 150 trees is shown in boxplot in Figure 15(b) and in scatter plot in Figure 15(c). The prediction error for validation data is shown in Figure 15(d). RF shows

promising potential in this task since aggregating results from multiple trees reduces the variability and is less susceptible to statistical noise.

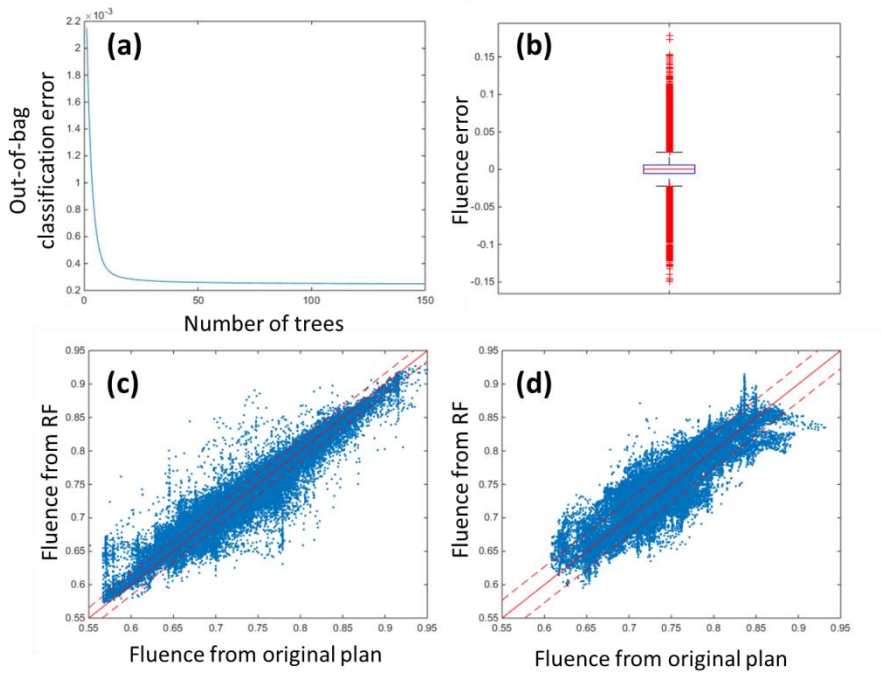

Figure 15: (a) Prediction error for the training data as a function of the number of trees; (b) boxplot of prediction error for all training data, outliers are shown in red; (c) scatter plot of prediction error for all training data, red solid line represents fluence from RF agrees with fluence from original plan, red dash lines represent one standard deviation of error; (d) scatter plot of prediction error for validation data, red solid line represents fluence from RF agrees with fluence from original plan, red dash lines represent one standard deviation of error.

The target volume coverage V100% was 78.1±4.7% for the auto plans, and 79.3±4.8% for the clinical plans (p=0.07). Volumes receiving 105% Rx were 69.2±78.0 cc for the auto plans compared to 83.9±87.2 cc for the clinical plans (p=0.91). The mean V10Gy, V20Gy of the ipsilateral lung was 24.4±6.7%, 18.6±6.0% for the auto plans and

24.6±6.7%, 18.9±6.1% for the clinical plans (p=0.99, >0.99). The statistics and boxplots are shown in Table 1 and Figure 16. Total computational time for the auto plans was < 20s.

Table 1: Dosimetric comparison between the auto plans and clinical plans. Statistical significance was tested via Wilcoxon Signed-Rank test. p<0.05 denotes statistical significance.

Plan metrics	Automated plan	Clinical plan	Wilcoxon Signed-Rank
PTV V100%	78.1±4.7%	79.3±4.8%	p=0.07
Lung V10Gy	24.4±6.7%	24.6±6.7%	p=0.99
Lung V20Gy	18.6±6.0%	18.9±6.1%	p>0.99
Lung V95%	2.1±2.3%	2.6±2.8%	p=0.96
PTV V105%	5.1±3.5%	6.4±4.8%	p=0.95
V105% (cc)	69.2±78.0	83.9±87.2	p=0.91

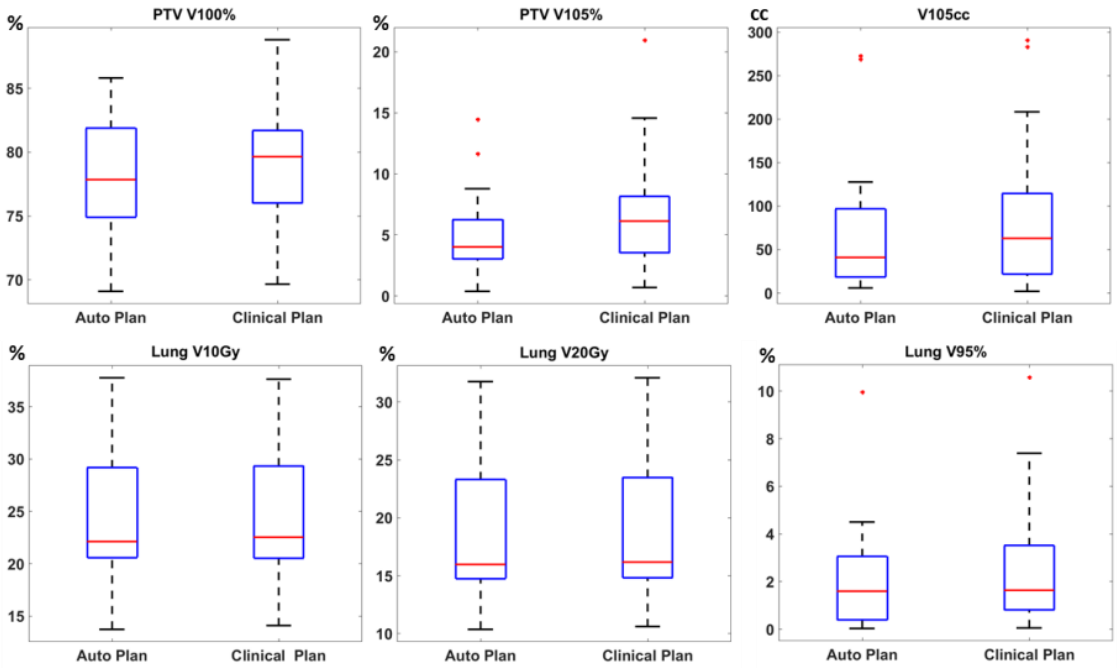

Figure 16: Boxplots of dosimetric comparison.

4.4 Discussion

We developed an ultra-fast fully automated system that can generate breast radiotherapy treatment plans with quality comparable to current clinical standard within seconds. This automated planning process can enable physicians and planners to very rapidly tailor/adapt breast radiotherapy for individual patient anatomical and tumor characteristics.

As reflected by the result, the dosimetric parameters between automated plan and manually generated clinical plan were overall comparable. No statistical significance was observed indicating non-inferiority of the proposed auto plans. More importantly, the automated method was able to achieve similar 105% Rx hotspot volume, which may take hours for human planner to reach such goal. The proposed method is able to prepare the tangential plan within very short period of time, and therefore multiple plans can be generated simultaneously to offer different clinical preference, such as improved coverage, reduced hotspot or reduced lung dose etc. Such flexible treatment frame is an important step towards precision medicine.

The proposed method bases on the physics principals, which is that the dose for each voxel is primarily control by the fluence intensity of beamlet from the medial and lateral beam. Correlating the voxel dose with the fluence intensity is a close approximation since we ignores the secondary radiation which is hard to model without considering patient geometry. Therefore, a good starting point (a good fluence

estimation) will guarantee the performance of the sequential fluence tuning. This study justified the feasibility of using physics principal to guide treatment planning, which is one step further than the anatomy-based method.

WBRT using tangential field enjoys the advantage of uniform dose distribution. Due to the redundant treatment planning process and logistic issue, such modality is currently not the most popular one for WBRT. Conventionally, 3D treatment using physical wedge was primarily used and was gradually replaced by FiF, IMRT and VMAT for dose homogeneity improvement. Better dose homogeneity has the benefit of lower rates of skin toxicity. Such advanced delivery techniques have seemingly inspiring advantages over conventional conformal techniques. However, American Society for Radiation Oncology (ASTRO) does not recommend routine use of IMRT to deliver WBRT although IMRT techniques have proven to reduce skin toxicity. This recommendation was provided considering value-based care where more sophisticated techniques such as IMRT has not been demonstrated to improve clinical advantages significantly while it is more expensive. However, WBRT using tangential field is billed as 3D technique although it is strictly an IMRT process. In the era when the society calls for value-based care, a more planning-efficient, low toxicity and inexpensive treatment technique could potentially gain preference for WBRT.

There are confounding factors for this application. The proposed technique works on the breast volume features to realize automatic planning. The current method

cannot generalize to other treatment sites using forward planning. The result may also be confounded by the fact that the automatic planning technique tries to minimize the 105% hotspot volume as much as possible. For the beamlets penetrating the lung, a reduction in 105% hotspot volume will result in the reduction in lung V10Gy and V20Gy. The clinical plans, however, may include the internal mammary nodes as the treatment volume. Extra fluence was usually applied from the medial beam which introduced extra dose to the lung.

4.5 Conclusion

The proposed RF model-based framework provides an efficient and automated mean to optimize breast planning. Automated plans provide similar target volume coverage and hotspot volume. The process reduces treatment-planning time from 1-4 hours to less than 10 min. Its high efficiency and near-real-time fine-tuning allows physician to spent less time waiting for plans, and to focus more on providing evidence-based, personalized care to breast cancer patients.

5. Development of a systematic workflow to improve the radiation therapy knowledge model quality

5.1 Introduction

5.1.1 TPKDD database and rapid learning

To extract useful knowledge from the dataset, appropriate treatment on the data is essential. The database we deal with in TPKDD is often massive, heterogeneous, dynamic and has various types of data. Therefore, several questions remain to be answered, e.g. how to treat noisy data (outlier), how to include multi-type data into the model, how to adapt the knowledge. Here, we are interested in extracting knowledge from rapidly evolving large volume database. One potential candidate tool is the rapid learning method.

A rapid learning health system model was proposed in Institute of Medicine to minimize the gap between clinical practice and research (Abernethy et al., 2010; Etheredge, 2007; Slutsky, 2007). The gist is to rapidly develop new evidence for clinical practice from information-rich database. A rapid learning system has the closed-loop format, which starts from collecting data, data interpretation for evidence generation, transforming treatment care, and finishes with outcome analysis which feeds back to the database. Such framework can be used in TPKDD to solidify the knowledge we extract. Under the TPKDD framework, a rapid learning approach can be pursued to address the nature of the database. First of all, outliers existing in the database need to be treated due to the massive and heterogeneous nature. Secondly, how to build a global model

which is generalizable to more heterogeneous data (different treatment sites for example) remains to be studied. Thirdly, how to adapt the knowledge with evolving new cases requires further analysis. In this part of the proposal, the issues discussed above will be addressed.

5.1.2 Novelty and outlier definition

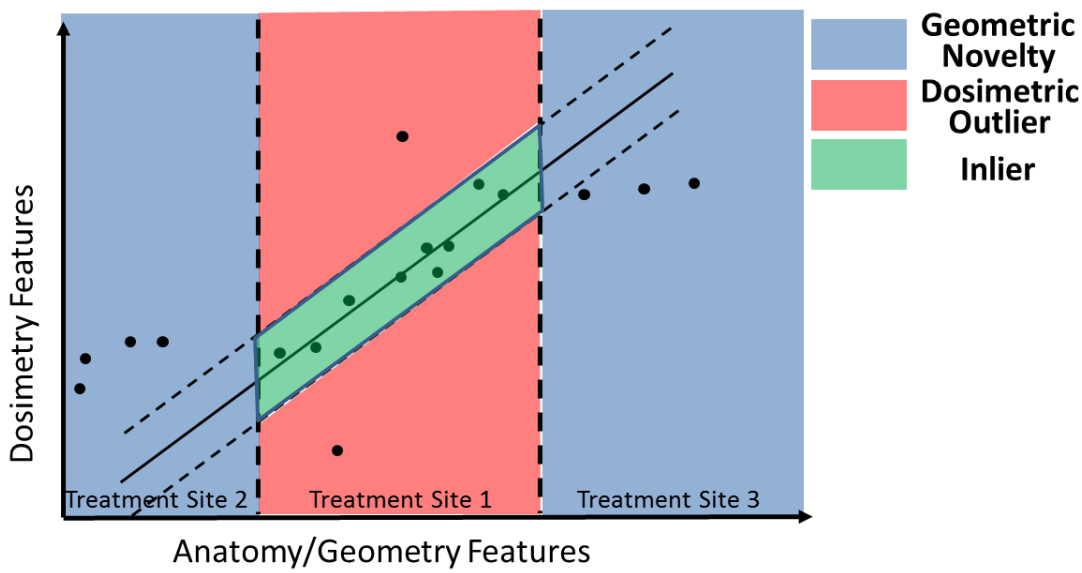

Figure 17: Illustration the geometric novelty and dosimetric outlier.

In the following Chapters, the concept of geometric novelty and dosimetric outlier will be studied. The multiple linear regression model proposed by Yuan et al. tries to fit the dosimetry features against the anatomy/geometry features as demonstrated in Figure 17. The geometric novelty case is defined as the case that displays different anatomy/geometry feature (blue shade area in Figure 17). The distinction in the anatomy/geometry feature does not rely on the dosimetry information,

i.e. the decision of signaling the geometric novelty case relies solely on the input anatomy feature. Often, the geometric novelty case is the one that is from a different treatment site. The following study will answer the question what the effect the geometric novelty case has on the model accuracy. The dosimetric outlier is defined as the case which shows similar anatomy/geometry feature while the dosimetry feature does not follow population mean (red shade area in Figure 17). The existence of the dosimetric outlier may potentially perturb the regression line and therefore degrade the prediction accuracy. Ideal scenario is that the model is trained with only inlier cases (green shade area in Figure 17). These questions will be answered in detail in the following Chapters.

5.1.3 Novelty and outlier in radiation therapy knowledge model

Knowledge-based planning aims to provide treatment planning guidance, such as dose-volume objectives and objective function weights. Recently, a commercial KBP software, RapidPlan (Varian Medical Systems, Palo Alto, USA), was developed and introduced to the Eclipse treatment planning system. Several pre-clinical studies have been performed to evaluate its ability in guiding treatment planning (Fogliata, Belosi, et al., 2014; Tol, Delaney, Dahele, Slotman, & Verbakel, 2015). In Tol et al.'s study, they found plans generated with RapidPlan were comparable to clinical plans if anatomy geometry was within the range of training cases (Tol et al., 2015). Fogliata et al. found improved dosimetric performance for the plans generated with the assistance of

RapidPlan as compared to the benchmark of clinically accepted plans (Fogliata, Belosi, et al., 2014). These studies not only prove the feasibility of implementing KBP in clinical environment, but also cautioned that in order to provide high quality model-based solution for clinical application, the model needs to be meticulously inspected.

In order to build a model that is generalizable to new cases, factors that need to be considered in the modeling and application process include the training data size, the existence of outliers, the range of the features it represents vs. the range of its potential clinical coverage etc. The range of the features is the distribution of the features of all cases. This is important in detecting the geometric novelty since the model may not be applicable to a new case if the new case's feature falls out of the range. The potential clinical coverage means the treatment site that the model can apply to. For example, the term "prostate model" usually means that the model is applicable to the prostate cases. Boutilier et al. analyzed the minimal required training sample size for predicting DVH points, DVH curves and objective function weight (Boutilier, Craig, Sharpe, & Chan, 2016). They performed the study using previously published methods (Boutilier, Lee, Craig, Sharpe, & Chan, 2015; B. Wu et al., 2011; Yuan et al., 2012b; Zhu et al., 2011). DVH point prediction using OVH curves proposed by Wu et al. (B. Wu et al., 2011) was analyzed for the minimal sample size required. DVH curve prediction based on PCA of dosimetric and geometric features proposed by Zhu et al. (Zhu et al., 2011) and Yuan et al. (Yuan et al., 2012b) was performed. For optimization objective function weight, they

applied logistic regression model using OVH based metrics and K-nearest neighbor model using 3-means clustering algorithm. Delaney et al. analyzed the effect of dosimetric outliers on the model using 70 HN cases (Delaney et al., 2016). They identified outlier cases from the model and composed a cleaned model. Then plans generated without the effort of sparing salivary glands were added to the cleaned model to assess the deterioration of model prediction caused by the outlier. They found that the existence of more than 20 dosimetric outliers can result in modest quality degradation as compared to the added outlier plan quality.

Outliers deviate from other observations and may be generated by a different mechanism (Hawkins, 1980). Due to the negative effect on the statistical analysis, such as increased error variance and reduced power of statistical tests, it is recommended to check the existence of the outliers.(Osborne & Overbay, 2004) Outlier detection has been heavily studied to identify anomaly among data (Aggarwal & Yu, 2001; Angiulli & Pizzuti, 2002; Arning, Agrawal, & Raghavan, 1996; Barnett & Lewis, 1994; Bay & Schwabacher, 2003; Breunig, Kriegel, Ng, & Sander, 2000; Fan, Zaiane, Foss, & Wu, 2006; Hodge & Austin, 2004; Knorr & Ng, 1998; Motulsky & Brown, 2006). There exist three categories of outlier detection techniques. The first category is the distance-based method. The outlier is identified if a fraction of β of all the observations in the dataset are further than r from it (Knorr & Ng, 1997). The second category is the cluster-based method. A cluster of smaller size, even including the cluster with only one observation,

is considered as clustered outliers. The algorithm measures the closeness of data and yields one cluster. The clustering result gives labels to all cases that are either "inlier" or "outlier". The third category is the spatial methods aimed to detect the spatial outliers. A spatial outlier is defined as the observation that is significantly different from its spatially related neighbors in the non-spatial attributes (Lu, Chen, & Kou, 2003). One example is the real time traffic map. One segment of the road with heavy traffic is considered as the spatial outlier if other segments of the road nearby are traffic free.

In radiation therapy KBP, the effect of the existence of the outlier needs to be studied. The model summarizes the dosimetry-anatomy relation and makes prediction based on the knowledge from the training data. High quality model relies on high quality plans. Delaney et al.'s study was trying to analyze the effect of the dosimetric outliers and demonstrated moderate degradation of the model quality with the existence of the dosimetric outliers. They proposed a two-step procedure to remove both the geometric and dosimetric outliers at the same time. The first step is identifying the observations which exceed the statistical metrics. The second step is confirming these outliers visually through regression, residual and geometric plot. Several questions and concerns still remain to be addressed. First, no simple and objective statistical tool exists to help identify the outliers. Second, the impact of the geometric novelty was not analyzed. In order to answer these questions, we provide a statistical tool to aid identifying the geometric novelty and dosimetric outliers. In addition, our study is

trying to evaluate the effect of the geometric novelty and dosimetric outliers respectively and to answer the question if cleaning the geometric novelty or dosimetric outliers is necessary. We used KBP for prostate cancer as an example application.

5.2 Materials and methods

Four groups of radiation therapy treatment plans in the prostate regions were included in this study: group 1 (G1) with 37 low-to-intermediate risk prostate cases; group 2 (G2) with 37 high risk prostate cases treated with lymph node (LN) irradiation; group 3 with 37 prostate bed irradiation cases; group 4 (G4) with 10 additional low-to-intermediate risk prostate cases other than those in G1. For G1-G3 cases, we used the IMRT plans designed for clinical treatment; and G4 cases were re-planned using the dynamic conformal arc technique (DARC). The two groups, G2 and G3, were used to represent the geometric variations relative to each other, i.e. the geometric/anatomic novelty. DARC plans in G4, which were not used to represent any clinical treatment techniques, were used to simulate dosimetric outliers to G1. In summary, G2 and G3 were used to analyze the geometric novelty, i.e. one group served as the inlier cases while the other group served as the geometric novelty cases. To study the dosimetric outlier cases, G1 cases were considered as the inlier cases, while G4 served as the dosimetric outlier cases. Figure 18 shows an example of the anatomy and dose distribution of four groups.

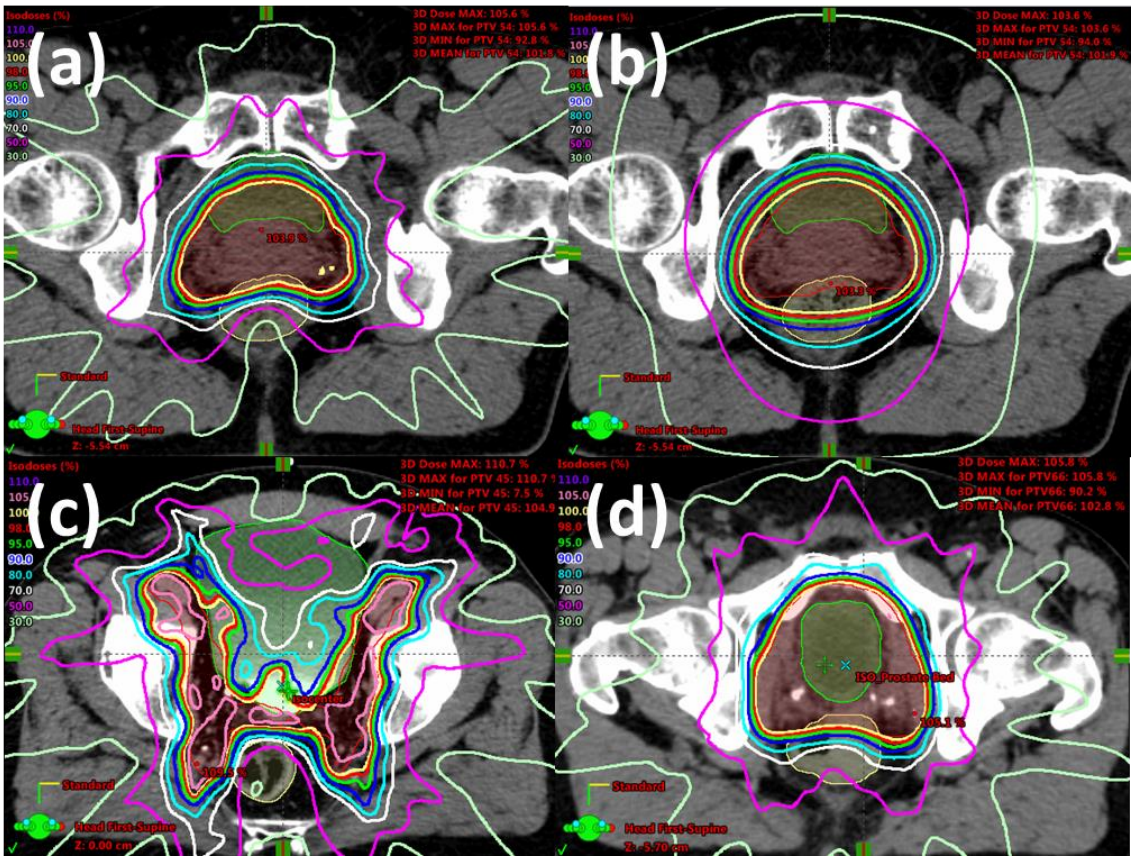

Figure 18: An example of anatomy and dose distribution of G1-G4 cases: (a) a prostate case (G1) shown with clinical IMRT dose distribution; (b) same prostate case from (a) shown with DARC dose distribution (G4); (c) a prostate plus LN case with clinical IMRT dose distribution (G2); (d) a prostate bed case (G3) with clinical IMRT dose distribution.

5.2.1 Model algorithm and study design

In this study, we used the KBP algorithm implemented by Yuan et al. (Yuan et al., 2012b) in earlier studies. This algorithm correlates the DVH (output) with the geometry features (input). The algorithm uses 22 geometry features including distance-to-target histogram (DTH) first three principal components (PCs), OARs overlap

portion, organ volume etc. A detailed list of the features is shown in Table 2. A stepwise multiple regression was performed to build the model.

Table 2: The OAR anatomical features analyzed in the algorithm. There are 11 features for each OAR relative to one PTV. One primary PTV and one boost PTV can be included. In this study, only the primary PTV was included. *PTV dose volume points are included to take into consideration of the OAR sparing variation among plans, since overly spared OAR can result in less homogenous PTV dose. This variation is adjusted by standardizing the dose volume point of training cases and set 0 for the new case.

Anatomical features
Distance to target histogram (DTH) principal component 1 (PC1)
DTH PC2
DTH PC3
Fraction of OAR volume overlapping with the PTV
Fraction of OAR volume outside the treatment field
OAR volume
PTV volume
OAR wrap angle around the PTV
DTH PC1 · DTH PC1
*PTV dose volume point 1 (PTV D2%)
*PTV dose volume point 2 (PTV D50%)

The first part of this study focused on the geometric/anatomic novelty. In particular, the statistical metric of *leverage* was studied for identifying the geometric novelty, and the mean Weighted Sum of Absolute Residuals (WSAR) was studied for assessing the impact of the existence of the geometric novelty. *Leverage* allows identifying the geometric novelty within a training dataset during modeling. It can also be used to assess whether a new case is a geometric novelty of an existing model. The Weighted Sum of Absolute Residual (WSAR) evaluates the effect of the geometric

novelty in a model and provides guidance whether cleaning up/excluding the novelty would be necessary to improve the modeling accuracy.

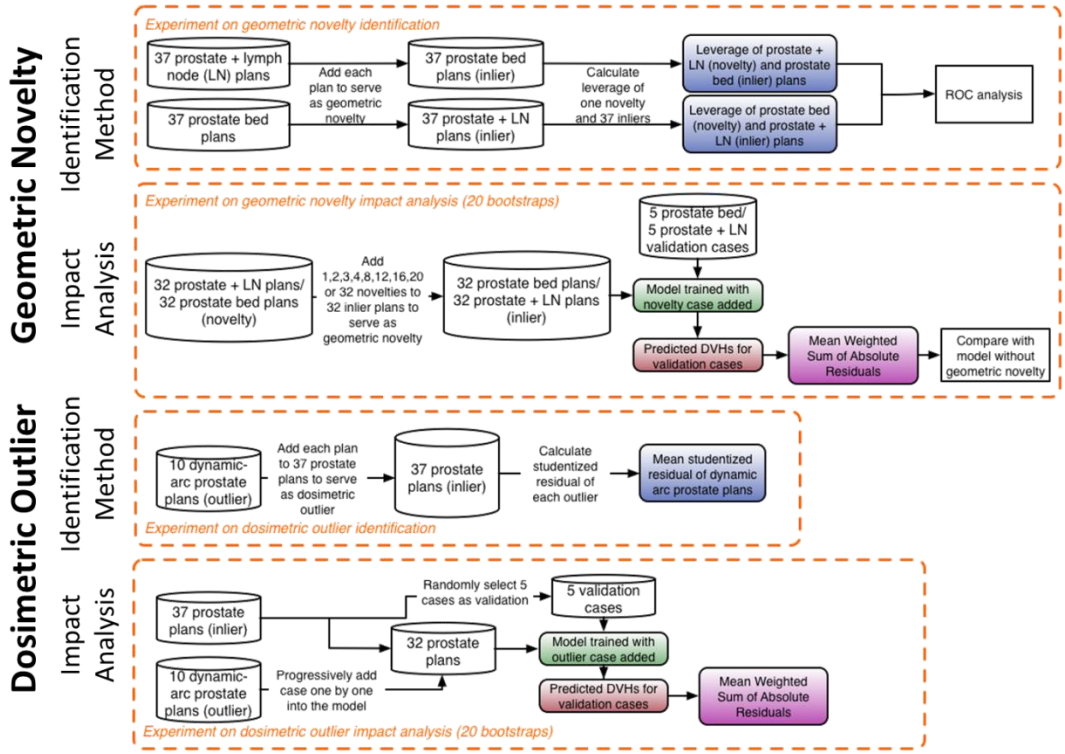

Figure 19: Flowchart of experiment on *geometric* novelty identification (top), *geometric* novelty impact analysis (second row), *dosimetric* outlier identification (third row) and *dosimetric* outlier impact analysis (bottom).

As shown in Figure 19, in the first experiment (top), each plan from the prostate plus LN group (G2) was individually added to the prostate bed model (G3) to serve as a geometric novelty. This process was repeated by individually adding each plan from the prostate bed model (G3) to the prostate plus LN group (G2). The anatomies of G2 and G3 cases were different, mimicking the process of introducing large geometric variation in the model. The *leverage* was calculated for the inliers and novelty, and receiver-

operating-characteristic (ROC) analysis was performed to determine the optimal threshold. In the second experiment, the geometric novelty were gradually added to the model to assess the impact, i.e. the prostate plus LN (G2) cases were gradually added to the prostate bed (G3) model and the prostate bed (G3) cases were gradually added to the prostate plus LN (G2) model. These models with the geometric novelty were then evaluated to assess the impact on prediction accuracy from the inclusion of the geometric novelty. In the third experiment, each dosimetric outlier from the DARC prostate group (G4) was added to the prostate (G1) model individually, and the model was trained with the corresponding outlier. The mean studentized residual of the dosimetric outlier from each experiment was recorded. In the fourth experiment, DARC prostate (G4) cases were gradually added to the prostate (G1) model to assess the model quality change.

5.2.2 Geometric novelty identification

First, a base model was trained with all G3 cases. Second, the geometric novelty, i.e. the prostate plus LN (G2) case, was individually added to the base model. Since the geometric novelty has certain feature that is far away from population mean, the *leverage* is able to reflect this characteristic. The *leverage* score of each training case is defined as

$$h_{ii} = (H)_{ii} \qquad \text{Eq.(12)}$$

i.e. the ith diagonal element of the hat matrix $H = X(X^T X)^{-1} X^T$, where X is the feature matrix. A feature matrix is an m-by-n matrix where m is the number of training cases

and n is the number of features. A stepwise regression was performed prior to the model training to select predictive features. The number of the features n selected varied between 1 and 5, and the selected feature subset was chosen as the feature matrix. Each element in the feature matrix is a scaler that quantifies a particular feature for a particular training case.

The *leverage* statistics of the inlier cases (G3) and novelty cases (G2) were recorded and the likelihood that the *leverage* of a randomly selected outlier is greater than that of a randomly selected inlier was assessed via Wilcoxon Rank-Sum test. The ROC analysis was performed to evaluate the performance of the *leverage* as a classifier to distinguish the geometric novelty. The inlier G3 cases were considered as condition **negative** and the novelty G2 cases were considered as condition **positive**. A *leverage* value larger than the threshold was considered as predicted condition **positive** while a *leverage* smaller than the threshold was considered as predicted condition **negative**. The sensitivity and specificity was calculated by varying the *leverage* threshold. The *leverage* of all inliers and novelty were pooled together and sorted ascendingly. The *leverage* threshold varied among the mean of two adjacent *leverage* values. The Youden's J index (Youden, 1950) was calculated to find the optimal threshold for differentiating the geometric inliers and novelty. The optimal threshold has the largest difference between the true positive rate and the false negative rate. This workflow was repeated by adding

the prostate bed (G3) cases to the base model trained with the prostate plus LN (G2) cases. The flowchart is shown on the top row in Figure 19.

To validate the effectiveness of using the *leverage* as a geometric novelty identification tool, leave-one-out cross validation was performed. For each of 37 geometric novelty cases, the optimal threshold was calculated using the other 36 cases. This threshold was then applied on this left-out case. If the *leverage* of this case is larger than the calculated threshold, it is marked as "detected". The detection rate of all 37 geometric novelty cases was reported for both the bladder and the rectum using G2 and G3 as the geometric novelty.

5.2.3 Impact on model accuracy and necessity of cleaning geometric novelty

In this part of the study, 32 cases were randomly selected from G3 to train a base model and then 1, 2, 3, 4, 8, 12, 16, 20 and 32 G2 cases were added to the base model to mimic different percentage of the geometric novelty (3, 6, 9, 13, 25, 38, 50, 63 and 100%) in the modeling process and assess the impact of modeling accuracy. This resulted into 9 knowledge models, in addition to the base model. Finally, five G3 cases other than the 32 cases used for training formed the validation cohort. This workflow was repeated by adding G3 cases to the base model trained with G2 cases.

The mean WSAR of the validation cases was calculated for the base model and the 9 models trained with different numbers of geometric novelty. The WSAR is given as

$$WSAR = \sum_{D=1}^{100} w_D \cdot \left|V_{c,D} - V_{p,D}\right| \cdot \Delta D \qquad \text{Eq.(13)}$$

where $V_{c,D}$ is the dose volume point for the clinical DVH at bin D; $V_{p,D}$ is the dose volume point for the predicted DVH at bin D. w_D is the weight for each bin, which varies from 50 (for 1st bin) to 90 (for 100th bin) divided by the sum of weights of all bins (normalization). ΔD is the bin width. This set of weighting penalizes more towards high dose region which is in correspondence with the clinical focus placed on the OAR dose.

The experiment was repeated 20 times with randomly selected training and validation cases. Statistical significance of the difference between the model with and without geometric novelty was calculated using Wilcoxon Rank-Sum test. Bonferroni correction was applied for multiple comparisons. The significance level was adjusted as $\alpha = 0.5/m$, where m is the number of hypothesis. Since there were 9 hypotheses, the significance level was set as 0.0056. The flowchart is shown on the second row in Figure 19.

5.2.4 Dosimetric outlier identification

The presence of the dosimetric outliers alters the correlation between the geometry and dose distribution. The studentized residual can be used to aid the identification of the dosimetric outliers. The studentized residual r_i is defined as:

$$r_i = \frac{e_i}{s(e_i)} \qquad \text{Eq.(14)}$$

where $e_i = y_i - \tilde{y}_i$, y_i is the response variable for i and $\tilde{y}_i$ is the regression prediction for i. $s(e_i)$ is the standard deviation of the prediction error. A studentized residual of 3 was chosen as outlier threshold. For the scenario when the model is trained but no cleaning has been performed, the studentized residual larger than 3 can signal the existence of the dosimetric outliers. For the scenario where there comes a new case, this new case can be added to the training cohort to train a new model and the studentized residual will be calculated to identify the outlier. The current algorithm decomposed the DVH curve into PCs and the first four PCs were used to build the model. Since the first PC of the DVH accounts for most of the variation in the DVH curve, we focus on the regression of the first PC of the DVH for the outlier analysis in this study.

In this study, we used the prostate cases planned with DARC (G4) which did not aim to spare the OAR to simulate the dosimetric outliers. DARC plans are good simulation of the dosimetric outliers because they do not strive to spare the dose to the OARs and often result in higher doses. From the DVH perspective, the DVH of DARC is higher than that of IMRT plans for all dose regions, and therefore results in higher score in the DVH first PC (Yuan et al., 2012b). Thus, G4 cases simulate *negative outliers* (positive studentized residual) once added to the model. Each outlier case was individually added to the prostate case (G1) dataset to train the model and obtain the studentized residual. The mean studentized residual of the outlier cases under this

simulation scenario was reported for both the bladder and rectum. The flowchart is shown on the third row in Figure 19.

5.2.5 Impact of dosimetric outliers on model accuracy

The impact of the dosimetric outliers was simulated by gradually adding multiple DARC prostate (G4) cases into the clinical prostate IMRT cohort (G1).

Similarly to the geometrical outlier analysis, the base model was trained with 32 cases from G1 with the remaining 5 cases from G1 reserved as the validation cases. Then each dosimetric outlier case in G4 was progressively added to the new base model until all 10 cases were added. Since the outliers introduced were all *negative outliers*, a monotonic degradation of model quality could be anticipated as the number of outlier increased. The performance of the model was evaluated by the WSAR.

The experiment was repeated 20 times via bootstrapping. The WSAR of the models with dosimetric outliers was compared to that of the base model (i.e. without outlier) via Wilcoxon Rank-Sum test. Bonferroni correction was applied for multiple comparisons. Since there were 10 hypotheses, the significance level was set as 0.005. The flowchart is shown on the bottom row in Figure 19.

5.3 Results

5.3.1 Leverage of geometric novelty

The mean and standard deviation of the *leverage* of the inlier and novelty cases are shown in Table 3. The mean of the *leverage* of the inlier cases was smaller than that of the corresponding novelty cases. The *leverage* of the inlier and novelty cases was significantly different (p<0.0001). Boxplots of the *leverage* are shown in Figure 20 for better visualization of the distribution. The distributions were most separated in the bladder when the prostate plus LN cases were added to the prostate bed model. This is in good agreement with the anatomical difference of the prostate plus LN plans and the prostate bed plans. When adding the prostate bed cases to the prostate plus LN cases, the *leverage* distribution was less separated than adding the prostate plus LN cases to the prostate bed cases. The area-under-curve (AUC) for differentiating the prostate plus LN case (G2) from the prostate bed cases (G3) for the bladder was 0.98 (threshold: 0.27); for the rectum, the AUC was 0.81 (threshold: 0.11). For differentiating the prostate bed case (G3) from the prostate plus LN cases (G2), the AUC was 0.86 (threshold: 0.11) for the bladder and 0.71 for the rectum (threshold: 0.11). The *leverage* could be used as a metric for identifying the geometric novelty as reflected by the AUC value.

The usage of the *leverage* as a geometric novelty identification tool was validated using leave-one-out cross validation. For the bladder, the sensitivity (predicted novelty cases divided by total novelty cases) of the prostate plus LN (G2) cases from the prostate

bed (G3) cases was 92% (34/37), and that of the prostate bed (G3) cases from the prostate plus LN (G2) cases was 76% (28/37). For the rectum, the sensitivity was 76% (28/37) and 73% (27/37), respectively.

Table 3: Mean and standard deviation (SD) of the *leverage* of the inlier and novelty group for various models trained with the geometric novelty in the first experiment. The *leverage* of both the inlier and novelty under the scenario of adding the prostate plus LN (G2) case to the prostate bed (G3) model is shown in the first row and adding the prostate bed (G3) case to the prostate plus LN (G2) model is shown in the second row. The statistical significance was calculated via Wilcoxon Rank-Sum test.

	(Mean, SD)		*p* value		(Mean, SD)		*p* value
	Inlier	Novelty			Inlier	Novelty	
G2+G3 Bladder	(0.11, 0.09)	(0.62, 0.22)	<0.001	G2+G3 Rectum	(0.09, 0.09)	(0.25, 0.26)	<0.001
G3+G2 Bladder	(0.08, 0.11)	(0.24, 0.19)	<0.001	G3+G2 Rectum	(0.10, 0.08)	(0.15, 0.10)	<0.001

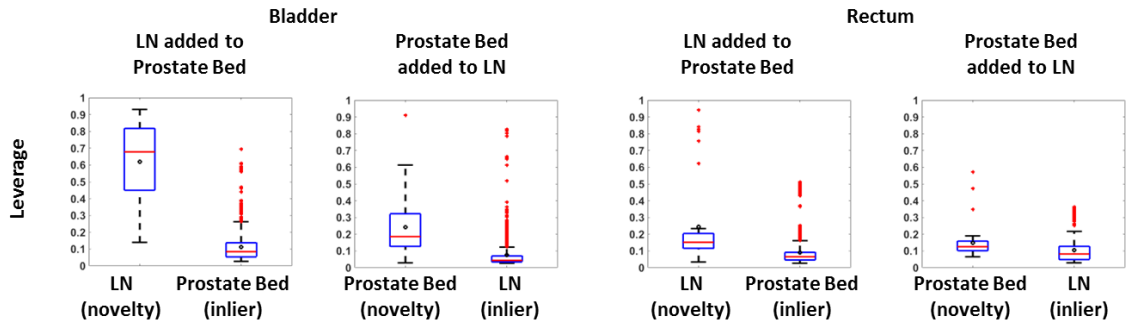

Figure 20: Boxplots of the *leverage* distribution of the novelty and inliers in the first experiment. The *leverage* distribution of the bladder model is shown on the left two subplots and the rectum model is shown on the right two subplots. Each geometric novelty case was added to the model and the *leverage* of the one geometric novelty LN/prostate bed case and the other 37 inlier prostate bed/LN cases was recorded. After adding all geometric novelty cases, the *leverage* statistics of the inliers and outliers were pooled to compose the boxplot. The *leverage* characterizes the distance of the data from the population mean and it has the range between 0 and 1. The edges of the box denote the interquartile range. The red bar denotes the median. The mean is represented as the black circle. The whiskers extend to the extreme data point within 1.5 times the interquartile range from the 25th/75th percentile. Data points beyond the whiskers are denoted as red "+".

5.3.2 Impact of geometric novelty on model accuracy

The mean WSAR for the base model and the 9 models trained with 1/2/3/4/8/12/16/20/32 geometric novelty cases is shown in Figure 21. For the bladder, significant degradation in model accuracy was observed after adding 16 G2 cases into the G3 model (p=0.0080, 0.0010 for adding 12 and 16 G2 cases into the G3 model; significance level at 0.0056). Adding G3 cases into the G2 model did not degrade the model quality at significance levels (p>0.0056 for all models). For the rectum, adding 32 G2 cases into the G3 model degraded the model quality (p<0.0001 for adding 32 G2 cases into the G3 model). Adding G3 cases into the G2 model did not degrade the model quality. The result showed the negative impact of the geometric novelty on the bladder and suggested the need to identify them to improve the model quality.

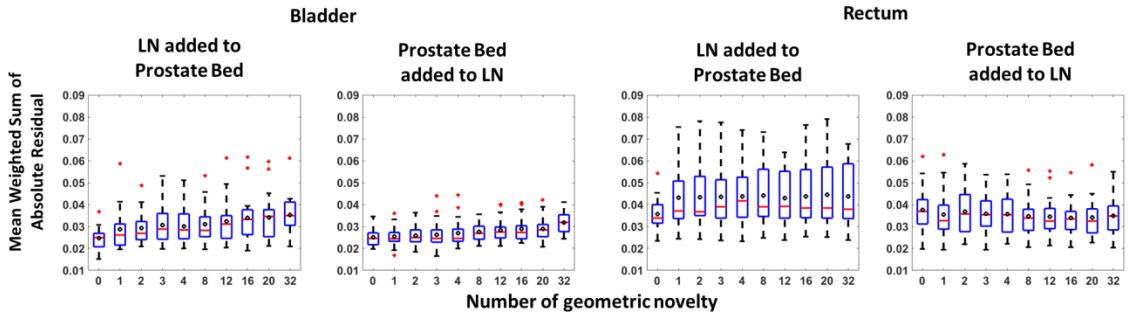

Figure 21: Mean Weighted Sum of Absolute Residuals distributions for the models with various numbers of geometric novelty (LN/prostate bed) added. The WSAR distribution of the bladder model is shown on the left two subplots and the rectum model is shown on the right two subplots. The edges of the box denote the interquartile range. The red bar denotes the median. The mean is represented as the black circle. The whiskers extend to the extreme data point within 1.5 times the interquartile range from the 25th/75th percentile. Data points beyond the whiskers are denoted as red "+". 1/2/3/4/8/12/16/20/32 geometric novelty (LN/prostate bed) were progressively added to the prostate bed/LN model with 32 cases and the model quality change was reflected by the WSAR. The WSAR was recorded for each bootstrap and the experiment was repeated 20 times. After adding 16 prostate plus LN cases into the prostate bed cases, the bladder model observed significant model quality change. Adding 32 prostate bed cases into the prostate plus LN cases degraded the model quality (p<0.0001). Adding the prostate plus LN cases into the prostate bed model or adding the prostate bed cases into the prostate plus LN model did not change the rectum model quality at p=0.0056.

5.3.3 Studentized residual of dosimetric outliers

Each of ten dosimetric outliers was added to the prostate model. The mean studentized residual of the dosimetric outlier cases was 10.06 for the bladder model and 9.87 for the rectum model. And the corresponding mean studentized residual of the inlier cases was -0.12 for the bladder model and -0.12 for the rectum model. The positive studentized residual signals *negative* dosimetric outliers where the original response variable (DVH PC1) in the model is higher than the model prediction. The *negative*

outliers are associated with less optimal OAR sparing while in the contrary the *positive* outliers are related to better OAR sparing than the model prediction. *Positive* outliers have less negative clinical impact as compared to the *negative* outliers, and they were kept in the model (Delaney et al., 2016).

5.3.4 Impact of dosimetric outliers on model accuracy

The WSAR for the validation cases is shown in Figure 22 with the increasing number of dosimetric outlier cases introduced into the model. The increasing number of the dosimetric outlier cases increases the mean WSAR for both the bladder and rectum. For the bladder, significant difference in the WSAR was observed for the models trained with 3 outliers (p=0.0038 for 3 outliers; significance level of 0.005); and for the rectum, only 1 outlier is needed in the model to create significant difference (p=0.0003 for 1 outlier).

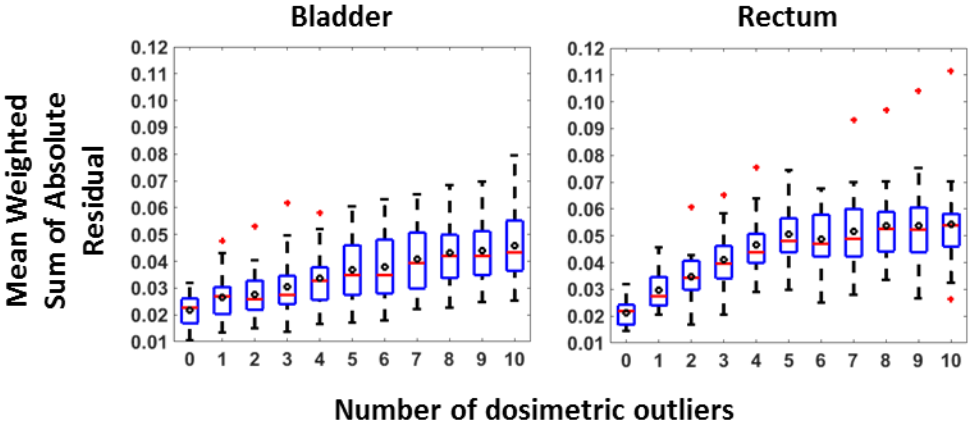

Figure 22: Mean Weighted Sum of Absolute Residuals distribution of the prediction from the models trained with different numbers of outliers. The edges of the box denote the interquartile range. The red bar denotes the median. The mean is represented as the black circle. The whiskers extend to the extreme data point within 1.5 times the interquartile range from the 25th/75th percentile. Data points beyond the whiskers are denoted as red "+". The dosimetric outlier cases were progressively added to the model until all 10 outlier cases were added. There were a total of 11 models with varying dosimetric outliers existing in the model from 0 to 10. Each model predicted the DVH curve for 5 validation prostate cases not used in the model training. The experiment was bootstrapped 20 times. At each bootstrap, the mean Weighted Sum of Absolute Residuals of the 5 validation prostate cases was recorded and all 20 bootstraps were plotted in the figure. Adding 3 dosimetric outlier cases affected the bladder model quality while adding only 1 dosimetric outlier case affected the rectum model quality.

5.4 Discussion

In this study, a systematic workflow for identifying and analyzing the geometric novelty and dosimetric outliers was established. Our study indicates that the *leverage* can be an effective metric for identifying the geometric novelty in a radiation therapy knowledge model. The simulation performed in this study showed the necessity of

cleaning both the geometric novelty and dosimetric outliers to maintain and improve the model quality.

Previous studies investigated how to train a good model from various perspectives (Boutilier et al., 2016; Delaney et al., 2016). Boutilier et al. implemented previously published methods of model training to investigate the required sample size for a model from statistical point of view. Delaney et al. analyzed the model quality of commercial software with different training datasets with or without dosimetric outliers. Delaney et al. found 3.9 Gy mean dose prediction deterioration while our study found that the existence of dosimetric outliers degraded the model quality at statistical significance level. Our study agreed with Delaney et al.'s study on the deterioration effect of the dosimetric outliers. Current study investigated both the geometric novelty and dosimetric outliers using a statistical method. The systematic workflow we established started from identifying and cleaning the geometric novelty, which could be a necessary extra step prior to treating the dosimetric outliers. We used the novelty and outliers that are likely to occur and are well understood in clinical practice and studied their identification using simulation. This is in contrast to previous studies that identify outliers using software recommendation.

An outlier-free model is able to predict the achievable dose-volume goal with high accuracy if the query case is not a geometric novelty. Extreme caution is recommended when predicting the dose-volume endpoint for a geometric novelty case,

since the dosimetry-anatomy relation for such geometric novelty case may not be fully captured and represented by the model. Therefore, when implementing the model to make predictions, it is necessary to compare the feature of the query case with that of the training cases.

Removing novelty/outliers have several clinical impacts. A new case can either be the geometric inlier or the geometric novelty. If this case is the geometric inlier, the current model is able to make prediction for this patient. If this case is indeed a geometric novelty, a different model needs to be applied. Removing geometric novelty will reduce the variation of the anatomy within a model and thus result in more models necessary to cover all cases. Building a model that can predict equally well for all cases is one possible solution and requires further study. For the dosimetric outliers, the plan quality deviates from the majority of other cases. The *negative* outlier is the case that the OAR is not sufficiently spared. It is more likely to be dosimetrically inferior. We need high quality plan to feed into the model so that the model can predict well for the new cases. Removing the dosimetric outliers is able to reduce the variation of the prediction and improves accuracy.

The effect of the model quality degradation by the dosimetric outliers was demonstrated with one example case illustrated below. A prostate model without dosimetric outlier and a prostate model with 10 dosimetric outliers were validated on one extra prostate case which was not used to train the model. The predicted DVH

curves from both models were used to extract the dose-volume objectives for both the bladder and the rectum to guide the treatment planning. The dose-volume objectives were the same as the ones used by Yang et al. (Y. Yang et al., 2015). As shown in Figure 23, the prediction from the outlier added model was less favorable than that of the outlier free model which agreed better with the clinical plan DVH. The prediction guided plan DVH agreed well with the prediction for both models. The two prediction guided plan DVHs differed in the medium-to-high dose region. This example demonstrated that the model quality change could have resulted in the final plan quality variation.

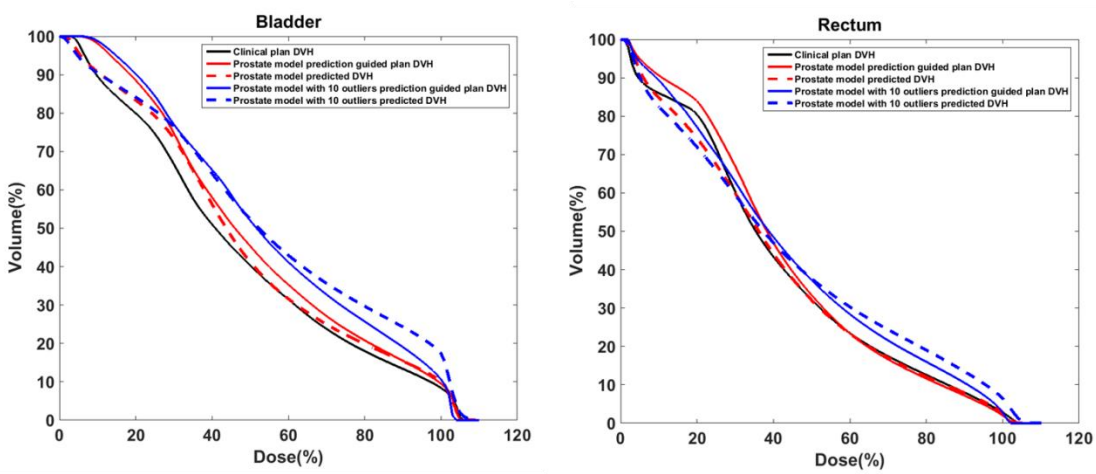

Figure 23: Clinical plan DVH, model predicted DVHs and prediction guided plan DVHs comparisons for the bladder (left) and the rectum (right) for one example prostate case. The black solid line is the clinical plan DVH. The red dash line is the predicted DVH from the prostate model without the dosimetric outlier. The red solid line is the outlier free prostate model prediction guided plan DVH. The blue dash line is the predicted DVH from the prostate model with 10 dosimetric outliers. The blue solid line is the 10 outliers added prostate model prediction guided plan DVH.

There are several limitations in this study. First of all, the baseline KBP model is currently trained with the cases from the same treatment site. In order to analyze the impact of the geometric novelty, radiation therapy plans from abdominal sites other than baseline model treatment site were treated as the geometric novelty. For example, a prostate bed case is a geometric novelty for a prostate plus LN model. However, it is possible that a hybrid model could be trained using cases from multiple treatment sites. For example, the prostate bed case no longer serves as a geometric novelty for a pelvic model which includes the prostate, the prostate bed and the prostate plus LN cases. In this scenario, the geometric novelty may become the inliers and additional metrics or tools are needed to inspect the quality of the cases. This is beyond the scope of this study and requires further investigation. Radiation therapy plans generated using DARC were treated as the dosimetric outliers when added to the single-site (prostate) model. The existence of the dosimetric outliers did degrade the model quality and the cleaning process is recommended. An outlier often arises from abnormal mechanism, e.g. treatment modality in this study. These outlier cases cannot represent the core of regular observation of the current plan population. However, the outlier cases may accumulate due to treatment protocol change at the institution and therefore updating the model is necessary. Further investigation about updating the model will be discussed in Chapter 9. Secondly, the dosimetric outliers introduced were all *negative outliers* with positive studentized residual in regression. *Negative outliers* often exist in the plans where the

dose to the OAR is not minimized with full effort, such as in the DARC technology. We designed this experiment setup to answer the question whether and how insufficiently spared dosimetric outliers affect the model quality. On the other hand, the *positive outliers*, where the OAR dose was overly minimized, could also exist in the clinical cases. The *positive* outliers are the cases where the OAR is better spared than the model prediction. OAR over-sparing is often related to tradeoff between multiple OARs. Although over-sparing for such organ does not degrade the quality for this organ, the tradeoff choice may make other OAR's sparing objectives unachievable. Thus, the analysis of the *positive* outliers requires modeling multiple OARs and further investigation is warranted to deal with this scenario. Thirdly, the *leverage* was used as the metric to identify the geometric novelty. Among several types of algorithms, the *leverage* falls into the category of the distance-based method of outlier detection. It measures the distance of each datum point to the mean of the data population. The *leverage* statistic is able to reflect the distance of each datum point to the mean of the population so that the cases can be inspected according to the sequence of *leverage* statistics. The leverage is able to identify the geometric novelty case one by one. The clustering-based method could also be employed in KBP novelty/outlier detection which will be discussed in Chapter 6. One cluster can be generated around the bulk of data while the observations outside the cluster frontier will be identified as the

novelty/outliers. This method is capable of identifying multiple novelty/outliers at the same time.

5.5 Conclusion

A systematic workflow for identifying and analyzing the geometric novelty and dosimetric outliers has been established. The *leverage* and studentized residual have demonstrated effectiveness in identifying the geometric novelty and dosimetric outliers respectively in the training datasets. Results in this study clearly illustrated that the existence of both the geometric novelty and dosimetric outliers degraded the model prediction accuracy and the process of identifying and cleaning them is necessary. The recommended workflow provides a solution to generate high quality knowledge models for clinical use to improve patient care in radiation therapy.

6. Development of a novel geometric novelty and dosimetric outlier detection system using a clustering-based method

6.1 Introduction

In the previous Chapter, a systematic workflow to identify and clean up the geometric novelty and dosimetric outliers was developed. The process starts with identifying and cleaning geometric novelty, followed by identifying and cleaning dosimetric outliers. The term "geometric novelty" and "dosimetric outlier" arises from clinical interpretation while the former considers the anatomy variation (input/features) and the latter considers the plan quality (output/response). A knowledge-based model performs regression task to learn the mapping from input to output.

$$\text{Input: } \{(x_i, y_i)\}_{i=1}^{n}, x_i \in \boldsymbol{\chi}, y_i \in \mathfrak{R} \qquad \text{Eq.(15)}$$

$$\text{Output: } f: \boldsymbol{\chi} \rightarrow \mathfrak{R} \qquad \text{Eq.(16)}$$

Leverage was used to identify "geometric novelty" while the studentized residual was used to identify "dosimetric outlier". These metrics are somewhat sensitive to the distribution of the dataset and cannot provide full around information about the outlierness. Two data with similar *leverage* could be totally different in anatomy. Studentized residual is calculated based on least-squared regression, which is sensitive

to the outlier cases in the model. Aforementioned challenges could be improved using more sophisticated outlier detection system.

From the statistics point of view, if the training dataset is not pulled by outliers, we are interested in detecting anomalies in new observations. This is often referred as "novelty detection". This is an unsupervised learning process since the dataset contains no outlier and we are more interested in the feature space. Usually clustering is performed for such task. On the other hand, if the training data contain outliers, we need to fit the central mode of the training data while ignoring the deviant observations. This process is called "outlier detection". The label of the data is required for such task. To look at the knowledge-based model from statistics point of view, we are interested in knowing the novelty plan and the outlier plan. Novelty plan could potentially affect the model quality as demonstrated in the previous Chapters. In the clinical settings, previously treated IMRT plans were all treatment approved which signified clinical acceptance. However, it does not necessarily mean that the quality of a plan follows the central mode of the database. Such dataset misses the label of outlierness. The outlierness could be identified by reviewing all IMRT plans within the cohort but the task is almost impossible to perform due to limited resources. Therefore a statistical model is warranted to provide such label before performing the task of improving the knowledge-based model quality. In this study, a fully automated systematic workflow to identify and clean up novelty and outlier from a knowledge-based model was

developed. The model was trained with a large cohort of pelvic IMRT plans with varying anatomical features.

6.2 Materials and methods

6.2.1 Materials

A total of 365 pelvic IMRT plans from four groups were included: group 1 (G1) with 126 low-to-intermediate risk prostate cases; group 2 (G2) with 126 high risk prostate cases treated with lymph node irradiation (LN); group 3 with 103 prostate bed irradiation cases; group 4 (G4) with 10 additional low-to-intermediate risk prostate cases other than those in G1. G1-G3 cases were included as inlier cases with potential novelty which we have no prior information of. G4 served as outliers remaining to be detected. The hypothesis is that the proposed workflow will be able to identify novelty and outlier cases.

6.2.2 Workflow

The whole process started with novelty detection utilizing one-class support vector machine (OCSVM), followed by outlier detection using robust regression followed by outlier identification (ROUT) method (Motulsky & Brown, 2006). The flowchart is shown in Figure 24.

Figure 24: Flowchart of novelty and outlier identification using one-class SVM followed by ROUT method.

6.2.3 One class support vector machine (OCSVM) for identifying geometric novelty

In order to detect novelty, a clustering task was performed. We are more interested in the novelty in the feature space χ. The philosophy behind the novelty detection is to use statistical methods such as clustering to identify novel anatomical features which may not fit into the model. Such identified case shall have specific clinical interpretation, such as large organ volume, irregular organ shape etc. Since for most scenarios, resources are limited to visually identify novel cases, no prior label exists to perform the classification task. Instead, the clustering method using one-class SVM is proposed to identify novel cases. The proposed method is composed of two major steps: 1. isometric mapping is applied to the feature matrix (22-dimensional) to reduce the dimension to 2D (manifold learning). Isometric mapping preserves the topological relation between data points. Since in novelty detection, the similarity is of major interest, the topological relation shall be preserved to effectively identify the novelty. 2. One-class SVM is trained using all data as input using the radial basis function (RBF) kernel. The penalty function was preset to 5% (percentage of novelty). The RBF kernel is written as:

$$k(\boldsymbol{x}, \boldsymbol{y}) = exp(-\gamma \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_2^2) \qquad \text{Eq.(17)}$$

where $\boldsymbol{x}, \boldsymbol{y}$ is the feature of two data, γ is a tunable parameter. The value of γ affects the kernel size and also affects the tightness of the cluster. A very tight cluster is suboptimal since it will lose generalization power. A loose cluster may not conform well to the frontier. An optimization process to select appropriate γ was previously proposed (Xiao, Wang, & Xu, 2015). The difference between the distance to the frontier surface of the "edge point" and "interior point" was maximized.

$$f_O(s) = \max_{\boldsymbol{x}_i \in \Omega_{IN}} d_N(\boldsymbol{x}_i) - \max_{\boldsymbol{x}_j \in \Omega_{ED}} d_N(\boldsymbol{x}_j) \qquad \text{Eq.(18)}$$

The first term is the distance of "interior point" to the frontier surface, and the second term is the distance of "edge point" to the surface. d_N is the normalized distance in the mapping space (Figure 25).

$$d_N(\boldsymbol{x}) = \frac{d(\boldsymbol{x})}{1 - d_\Pi} \qquad \text{Eq.(19)}$$

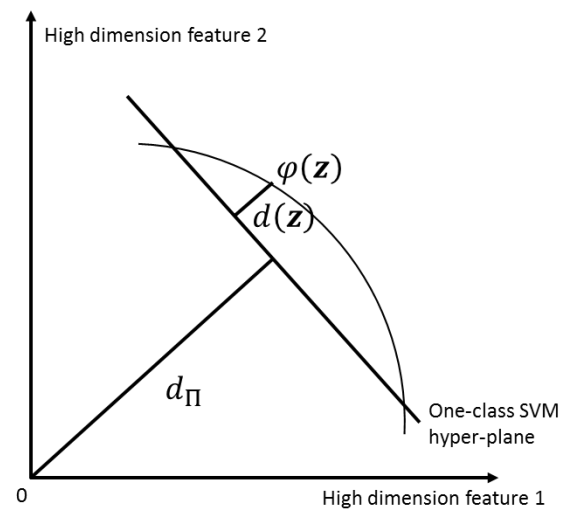

Figure 25: OCSVM projection in high dimensional space.

The "edge point" and "interior point" was determined using the edge detection algorithm developed by Li and Maguire (Y. Li & Maguire, 2011). For an "edge point", most of its neighboring points should locate on one side of the tangent plane; while for an "interior point", its neighboring points shall locate in all 2π directions. First, k vectors originating from the query point were generated for k nearest neighbors. Then all k vectors were normalized and the sum of all k vectors was calculated. Finally the dot product between each vector and the sum vector was calculated.

$$\boldsymbol{v}_i^n = \sum_{j=1}^{k} \boldsymbol{v}_{ij}^u \qquad \text{Eq.(20)}$$

$$\theta_{ij} = \boldsymbol{v}_{ij}^u \cdot \boldsymbol{v}_i^n \qquad \text{Eq.(21)}$$

$$l_i = \frac{1}{k}\sum_{j=1}^{k}\left(\theta_{ij} \geq 0\right) \qquad \text{Eq.(22)}$$

$\boldsymbol{v}_{ij}^{u}$ is the normalized vector for the neighbor j. k is preset to 30. l_i belongs to (0,1]. A point was determined as an "edge point" if $l_i > 0.9$. It was determined as an "interior point" if $l_i \in [0.4,0.6]$. f_O was then optimized against γ and the optimal γ was used to perform one-class SVM. Novelty cases shall be identified which locate beyond the frontier.

6.2.4 Robust regression followed by outlier identification (ROUT) method for identifying dosimetric outliers

Once novelty cases were identified, the ROUT method was performed to identify outlier cases. The ROUT method started with the robust regression followed by outlier detection. The robust regression assumes Lorentzian distribution rather than Gaussian distribution which is the underlying assumption of the least-squared regression. The robust regression is less sensitive to the outlying points compared to least-squared regression. After the robust regression was done, the Robust Standard Deviation of the Residuals (RSDR) was calculated using the equation below. $P68$ is 68% percentile of residuals from the robust regression. N is the number of data, and K is the number of features/attributes used in the regression.

$$RSDR = P68\frac{N}{N-K} \qquad \text{Eq.(23)}$$

$RSDR$ is an analogy of the standard deviation while $RSDR$ is less sensitive to the outlying data. Based on $RSDR$, the two-sided p value for each data was calculated and compared to the α value.

$$\alpha = \frac{Q(N-(i-1))}{N} \qquad \text{Eq.(24)}$$

Q is false discovery rate (FDR) and i is the index from $0.7N$ to N (data is ranked based on ascending residual, we are interested in 30% most outlying data). The Q value was set to 1%. If p value was less than α, then the current data and the data beyond (larger residual) were identified as the outliers.

The effectiveness of identifying the novelty and outlier cases was verified using 10-fold cross validation by implementing the model developed by Yuan et al. Both the bladder and the rectum were used in this study for analysis since the geometry varies among different groups. All inlier cases (novelty and outlier free) were divided into 10 folds. A clean model was trained using 9 folds cases. The clean model was validated on the remaining 1 fold cases as well as the novelty and outlier cases. The mean WSAR for all validation inlier cases was compared against that of the novelty and outlier cases. Wilcoxon Rank-Sum test was performed to compare the prediction accuracy between two groups of cases. Some example novelty and outlier cases were inspected for the clinical interpretation of being the novelty/outlier cases.

6.3 Results

The isometric mapping of all data into 2D for both the bladder and rectum is shown in Figure 26. Data aggregation was observed for some data while some other data points were detached from the cluster, reflecting dissimilarities between the cohort and the data. The frontier determined by the one-class SVM using various γ is shown in Figure 27. A small γ generated a relatively loose frontier while a large γ generates a tight frontier. An appropriate frontier was able to conform nicely on the dataset.

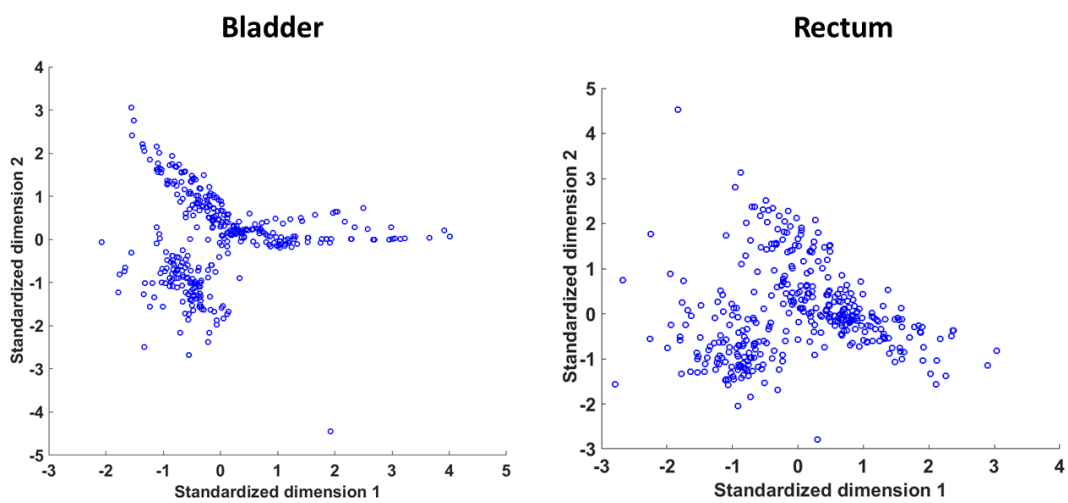

Figure 26: Isometric mapping of all data in 2D. The bladder feature mapping is shown on the left and the rectum feature mapping is shown on the right. Horizontal and vertical axes are standardized units.

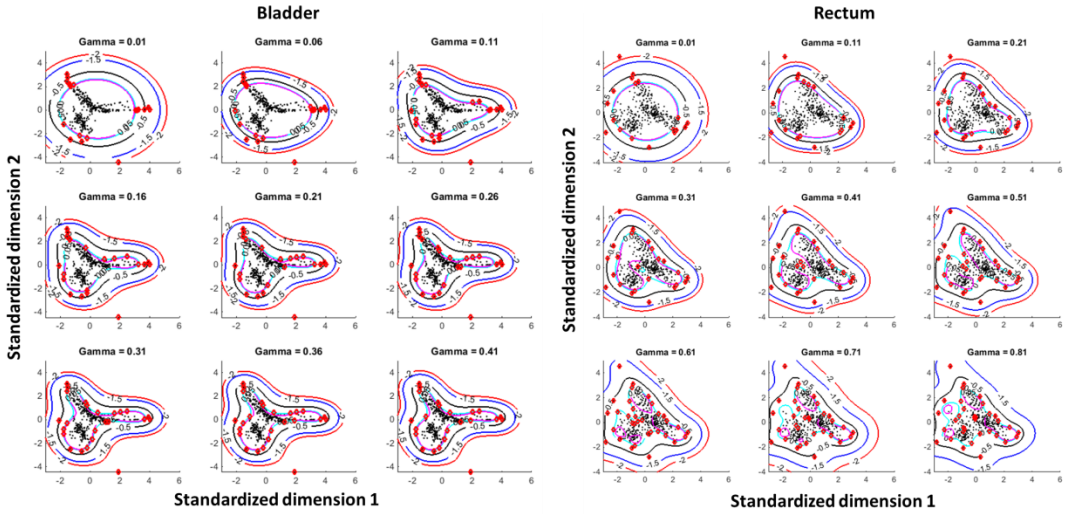

Figure 27: Frontiers generated by OCSVM using varios γ (0.01-0.41 for the bladder and 0.01-0.81 for the rectum). The bladder frontiers are shown on the left and rectum frontiers are shown on the right. A larger γ has tighter frontier. Red, blue, black, cyan and pink contour denote the isocontour of -2, -1.5, -0.5, 0 and 0.05 decision boundary. A value larger than 0 is classified as positive/inlier class. The black dot denotes inlier case and the red circle denotes geometric novelty.

Using the method proposed by Li and Maguire, the edge points and interior points were identified as demonstrated in Figure 28. Red star denotes edge point and the black square denotes edge point. The plot of f_O against γ is shown in Figure 29. The optimal γ chosen was 0.23 for the bladder and 0.44 for the rectum in this dataset. The corresponding frontier and the identified novelty cases are shown in Figure 30. Green dots represent cases within the frontier. Red crosses represent detected novelty cases. Black squares represent the support vectors. For the bladder model, a total of 18 cases were identified as the novelty cases and they were subsequently cleaned. For the outlier detection part, a total of 23 cases were identified. All 10 DARC cases were identified. The remaining 324 cases were used to train the cleaned bladder model. For the rectum

model, a total of 20 cases were identified as the novelty cases and they were subsequently cleaned. For the outlier detection part, a total of 18 cases were identified as the outlier cases. All 10 DARC cases were identified. A total of 327 cases were used to train the cleaned rectum model. The comparison of the model prediction accuracy between the inlier case group and novelty/outlier case group is shown in Figure 31. The error of the inlier case group was significantly different from that of the novelty/outlier group. Novelty and outlier cases presented different dosimetry-anatomy relation compared to the core observation in the model.

Example novelty and outlier cases for the bladder model are shown in Figure 32. Figure 32(a) and 32(b) are two novelty cases. Figure 32(a) displays an extremely large bladder volume while (b) has large overlap between the PTV and the bladder. Figure 32(c) and 32(d) are two outlier cases. For both cases, the bladder sparing is minimal which deviates from the mean sparing performance within the data cohort although the plan quality is clinically acceptable.

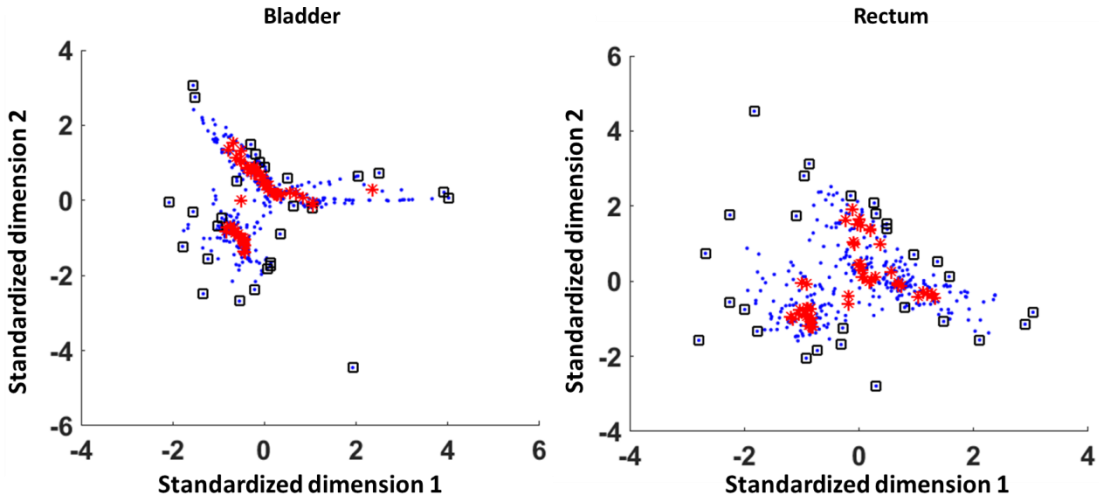

Figure 28: Detected edge point (black box) and interior points (red star).

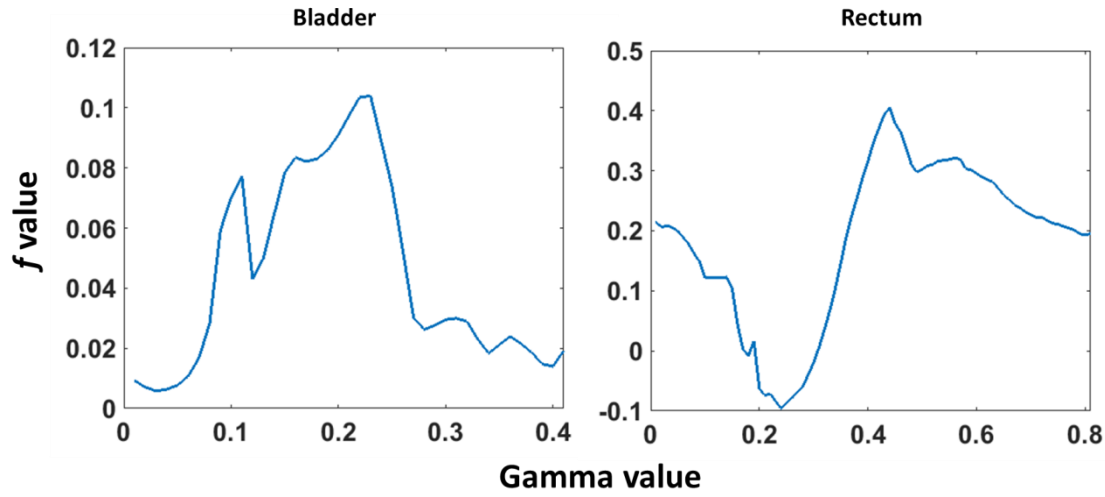

Figure 29: f_0 changes versus various γ. The optimal γ was 0.23 for the bladder model shown on the left and 0.44 for the rectum model shown on the right.

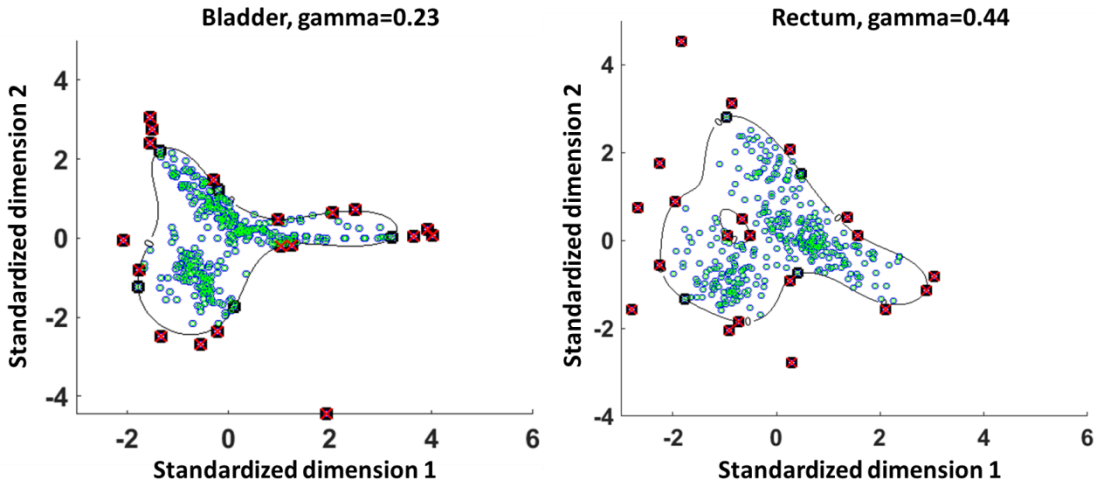

Figure 30: Frontier generated from OCSVM is shown in black curve. Green dots represent cases within the frontier. Red crosses represent detected novelty cases. Black squares represent support vectors. The bladder frontier is shown on the left and the rectum frontier is shown on the right.

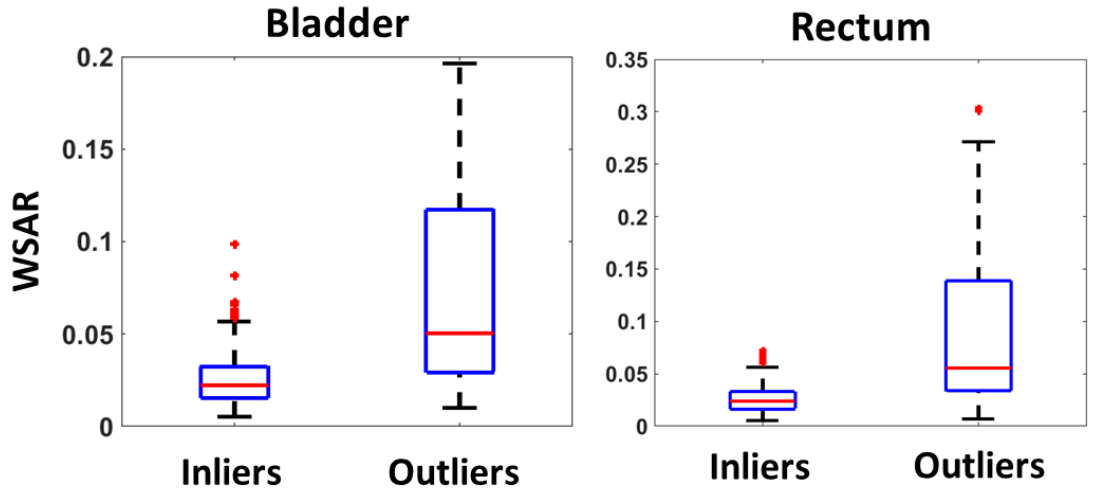

Figure 31: The WSAR comparison between the inlier case group and the novelty/outlier case group. The bladder model accuracy is shown on the left and the rectum model accuracy is shown on the right. For the bladder, the mean and standard deviation of the WSAR was 0.026, 0.014 for the inliers and 0.073, 0.054 for the novelty/outliers. For the rectum, the mean and standard deviation of the WSAR was 0.027, 0.013 for the inliers and 0.094, 0.087 for the novelty/outliers. The inlier and novelty/outlier WSAR was significantly different for both the bladder and the rectum.

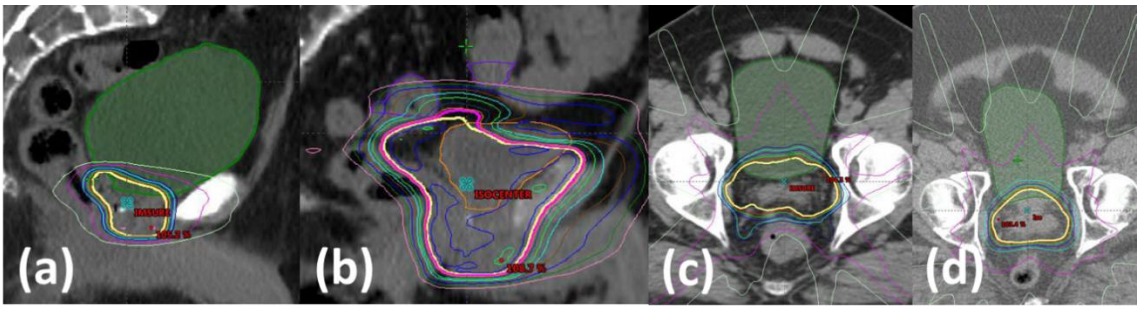

Figure 32: Example detected novelty cases ((a) and (b)) and outlier cases ((c) and (d)). (a) has larger bladder (green) volume than normal cases. (b) have larger PTV-OAR overlap than normal cases. (c) and (d) did not spare adequately for the bladder (pink isodose line denotes 50% Rx).

6.4 Discussion

This study proposed a hybrid approach of the OCSVM and the ROUT method to identify multiple geometric novelty and dosimetric outlier cases. The entire pipeline workflow was automatic without human intervention. The OCSVM step served as the geometric novelty identification tool while the ROUT method identified the negative dosimetric outliers. The result showed that all intentionally added dosimetric outliers were identified. The cleaned model trained with only inlier cases showed different prediction accuracy for the inlier and the novelty/outlier cases, denoting distinctive geometry-dosimetry relation between these two groups. Removing the novelty and outlier cases improved the training case plan quality and therefore the model quality.

The OCSVM step generated the frontier which distinguishes the inlier and novelty/outlier group. The OCSVM is essentially a clustering task (unsupervised

learning) rather than the classification task (supervised learning) such as the traditional SVM classification. The classification task tries to correctly yield the predicted class label for the training dataset while maintaining decent generalization ability. This is especially important for the SVM classification. The SVM using the Gaussian kernel is superior to other kernels such as linear and quadratic kernels in providing the non-linear classification boundary. However, the balance between the classification accuracy and the smoothness of the boundary needs to be found to improve the generalization ability. This balance also holds true for the OCSVM although it is a clustering task. For the OCSVM, 5% was the preset percentage of the novelty cases. This percentage is tunable according to the user's setting. Since there is no ground truth class label for clustering evaluation, external evaluation is needed to identify the correct parameter setting for the optimal frontier. In this study, the optimization of γ was performed to fine tune the parameter.

Isometric mapping was used in this study to reduce the dimension of the feature vectors while maintaining relative topological relation among all data points. The proposed framework of the OCSVM and ROUT also worked with high dimensional data when the isometric mapping was not used. Converting high dimension data to 2D served the purpose of better visualization and inspection of the data distribution.

The frontier generated from the OCSVM was unique for the training dataset, meaning that the inclusion of new data may perturb the frontier. It makes sense from the

modeling perspective, since the growth of the training dataset increase the geometry variation and the frontier may expand. It is possible that the geometric novelty case is no longer novelty when a batch of new cases is included in the model. This would suggest re-training the frontier when new cases are added to the model.

6.5 Conclusion

A one-class SVM approach followed by the ROUT method is implemented to identify the geometric novelty and dosimetric outliers in the radiation therapy knowledge model. The proposed method is effective in identifying both the geometric novelty and dosimetric outliers. The model trained with only inlier cases shows different prediction accuracy for the inlier group and the novelty/outlier group, denoting the difference in the geometry-dosimetry relation between two groups. The proposed method has the advantages of identifying multiple geometric novelty and dosimetric outliers simultaneously.

7. Building a global knowledge model to include plans with different clinical conditions and indications using model tree

7.1 Introduction

In this Chapter, a global model is developed to include cases with different clinical conditions and indications. Radiation therapy plans within one treatment site usually exhibits similar anatomy-dosimetry relation. This also implicates that such relation is often different between various treatment sites. Using a model trained with prostate plus LN cases may not necessarily predict well for the prostate cases. Previous Chapters also confirmed that the existence of cases that do not anatomically belong to the training data group may degrade the model quality. This conclusion holds true for the model trained with a linear model. In the clinical settings, the cases with different anatomical features and plan quality will continue to emerge. For example, it remains questionable if we can build a model that is trained with the cases from various pelvic treatment sites while the model predicts equally well as compared to the model trained with the cases from single treatment site. For some complicated treatment sites, the clinical decision is made upon the patient's anatomy, which then affects the plan quality. For example, tradeoff is often made between two parotid glands depending on the anatomy for the HN IMRT patients. One parotid may be sacrificed if it locates close to the tumor while the other parotid is spared with maximum effort to preserve patient's

life quality. Therefore we are interested in building a model which is able to differentiate the anatomy and then makes prediction using prior cases with similar anatomy.

M5 model tree is a good candidate for such task. It is developed based on the decision tree, which bifurcates based on the certain feature to improve the node purity. The decision tree continues to grow until the stopping criteria is reached and the cases locate in the same sub-branch form a leaf. The decision tree can perform both the regression and classification task through majority voting using cases within the same leaf. M5 model tree improves the decision tree by performing regression using cases within the same leaf rather than directly using the mean value of the training cases. It is a piecewise linear model. Therefore in this part of the study, we are trying to build a model tree using plans with different clinical conditions. Specifically, pelvic IMRT plans were used as the training cohort and its performance was compared with the single-site linear model which was treated as the best model available. In addition, a model tree was built for the HN parotid trade-off cases to see if the prediction is comparable with the model trained with either bi-lateral sparing or single-side sparing parotid cases.

7.2 Materials and methods

7.2.1 Pelvic cases

In this study, a total of 355 pelvic IMRT plans from three groups were included: group 1 (G1) with 126 low-to-intermediate risk prostate cases; group 2 (G2) with 126

high risk prostate cases treated with lymph node irradiation (LN); group 3 with 103 prostate bed irradiation cases.

The model tree was built and validated using 10-fold cross validation. The prostate cases were divided into 10 folds, and 9 folds together with the prostate plus LN and prostate bed cases were used to train the model tree. The other one fold was reserved for validation. The model was trained repeatedly leaving all 10 folds out for training. A baseline single-site model was trained using 9 folds of the prostate cases. The model tree was trained with varying minimal leaf size of 5, 6, 7, 8, 9, 10, 20, 30, 40 and 50.

The WSAR of the validation cases was compared between the model tree and the single-site linear model. The experiment was repeated for the prostate plus LN single-site model and the prostate bed single-site model with the difference in 10-folding the prostate plus LN cases and the prostate bed cases, respectively. The proposed experiment was performed in both the bladder and the rectum model. One example flowchart for the comparison of the model tree and the single-site prostate model is shown in Figure 33.

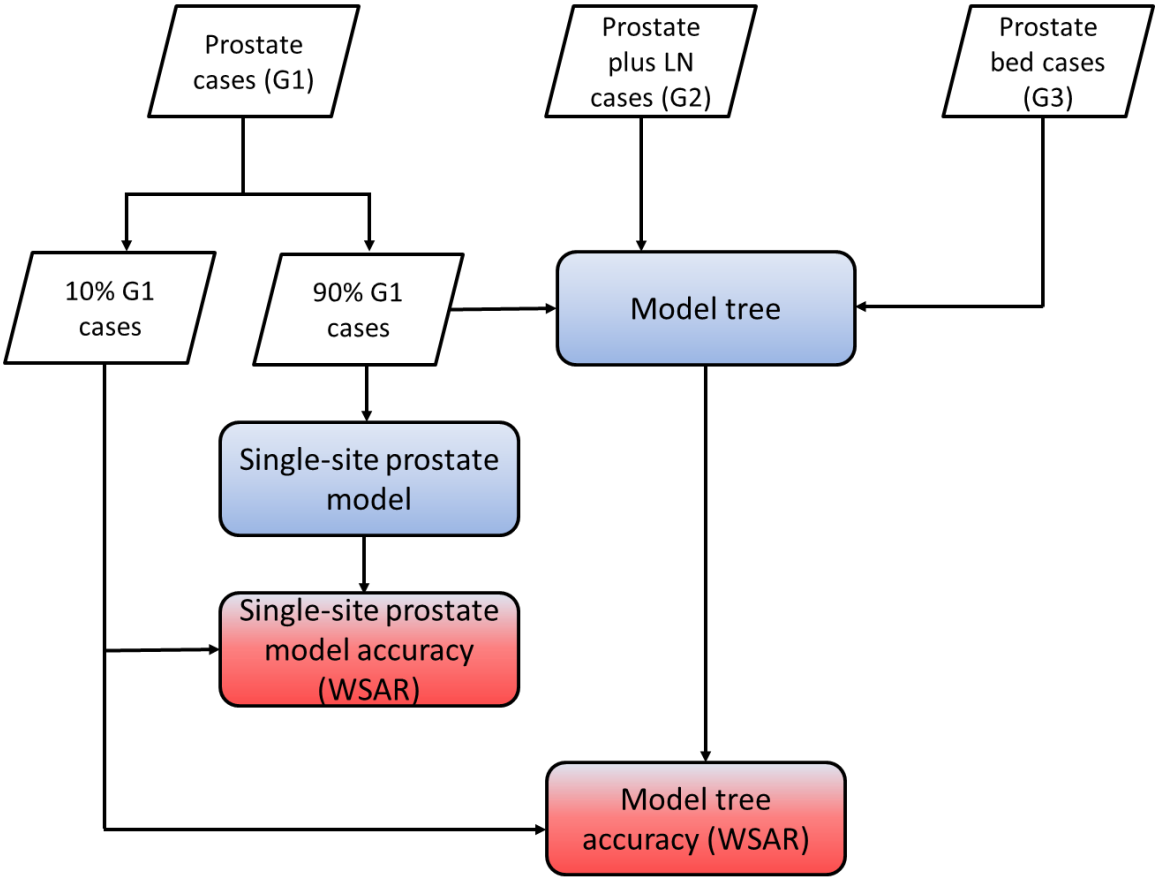

Figure 33: Flowchart of the model tree training and the single-site prostate model training. The validation was performed by applying the trained model to the extra one fold of prostate cases and calculating the WSAR.

7.2.2 HN cases

A total of 73 HN IMRT cases were used. Among all 73 cases, 14 cases' right parotid was given up and the other 14 cases' left parotid was given up. Both parotids from all 73 cases excluding those given up ones were used to train the model tree. Since the prediction was applied to each parotid, the bi-lateral cases can be used twice as the training data with each parotid as an individual entry. There were a total of $73 \times 2 - 14 - 14 = 118$ inputs. The features and the output is formalized as below.

$$f\left(\begin{bmatrix} R-parotid\ feature\ (59\times 22) & L-parotid\ feature\ (59\times 22) \\ L-parotid\ feature\ (59\times 22) & R-parotid\ feature\ (59\times 22) \end{bmatrix}\right) = \begin{bmatrix} R-parotid\ prediction\ (59\times 1) \\ L-parotid\ prediction\ (59\times 1) \end{bmatrix} \quad \text{Eq.(25)}$$

The model tree was trained with the minimal leaf size of 5. Among all 118 inputs, 103 entries were used to build the model tree for the HN cases. The other 15 entries were reserved for validation purpose. Among the 15 validation entries, 5 entries were single-side sparing parotids and 10 entries were bi-lateral parotids (5 patients in total; each patient has a pair of parotids). A baseline model was trained using the method proposed by Yuan et al. (Yuan et al., 2014). Two models were built to account for the different sparing performance. The first model was the standard model which was trained using all training bi-lateral cases. The second model was the single-side (single or single-lateral) model which was trained using all training single-side cases. Each individual model was able to capture the geometry-dosimetry relation within their training cohort. This dual model paradigm served as the baseline model for this study.

The bi-lateral sparing and single-side sparing parotids vary in the effort of reducing parotid median dose. Therefore it is important to know if the model over predicts the DVH or under predicts the DVH. In order to better compare the model quality, a modified version of the WSAR was proposed. The Weighted Sum of Residuals

(WSR) of the validation cases was proposed and adopted in this study. The WSR is given as

$$WSR = \sum_{D=1}^{100} w_D \cdot (V_{c,D} - V_{p,D}) \cdot \Delta D \qquad \text{Eq.(26)}$$

where $V_{c,D}$ is the dose volume point for the clinical DVH at bin D; $V_{p,D}$ is the dose volume point for the predicted DVH at bin D. w_D is the weight for each bin, which varies from 50 (for 1st bin) to 90 (for 100th bin) divided by the sum of weights of all bins (normalization). ΔD is the bin width. This set of weighting penalizes more towards high dose region which is in correspondence with the clinical focus placed on the OAR dose. This is a modified version of WSAR to better reflect if the model over predicts or under predicts the DVH. It preserves the sign of difference since it is important in this task to know if the model over predicts or under predicts.

The WSR for all validation cases was compared between the model tree and the standard/single model. In addition, the D50 of the predicted parotid gland was compared among all trained models since it is the clinical evaluation metric for the parotid sparing performance. The flowchart is shown in Figure 34.

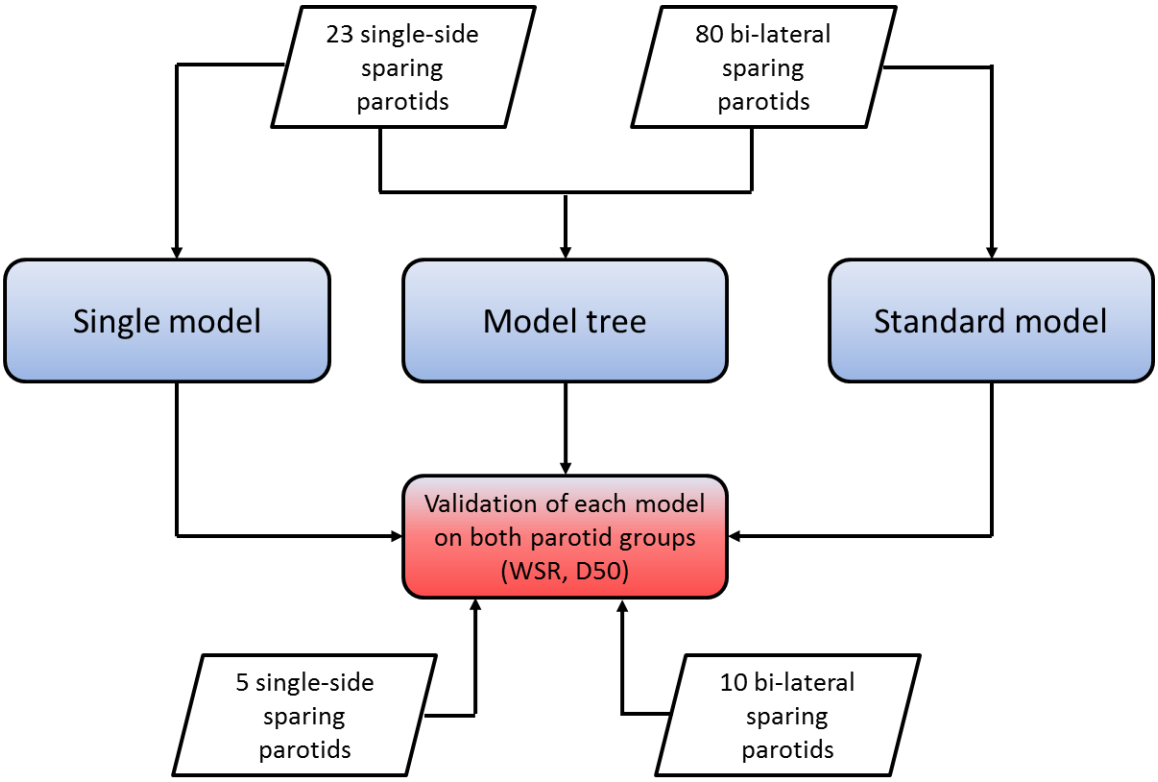

Figure 34: Flowchart of the model tree training as well as the single/standard model training. The single model was trained using 23 single-side parotids. The standard model was trained using 80 bi-lateral parotids. The model tree was trained using 23 single-side parotids and 80 bi-lateral parotids. Each model was validated on 5 single-side and 10 bi-lateral validation parotids. The WSAR and the D50 of the model prediction was evaluated.

7.3 Results

7.3.1 Pelvic cases

Figure 35 shows the WSAR of the prostate cases between the model tree and the single-site prostate model using the bladder as the OAR. The validation case WSAR is shown on the left and the training case WSAR is shown on the right. The model tree was trained with varying minimal leaf sizes shown on the horizontal axis. The validation error increased as the minimal leaf size increased. Statistical significance between the

model tree with minimal leaf size of 20 and the single-site prostate model was observed ($p<0.005$). Figure 36 shows the comparison for the rectum model. No statistical significance was observed. Figure 37 and Figure 38 shows similar comparison when the model tree was applied on the prostate plus LN cases for the bladder and rectum, respectively. Significant difference between the model tree with the minimal leaf size of 40 and the baseline model was observed for the prostate plus LN cases ($p<0.005$) for the bladder. When applied for the prostate bed cases (shown in Figure 39 and 40), the model tree with minimal leaf size of 30 was significantly different from the baseline model ($p<0.005$) for the bladder. The model tree was able to provide similar prediction accuracy as compared to the baseline model if a small minimal leaf size was used to train the model tree.

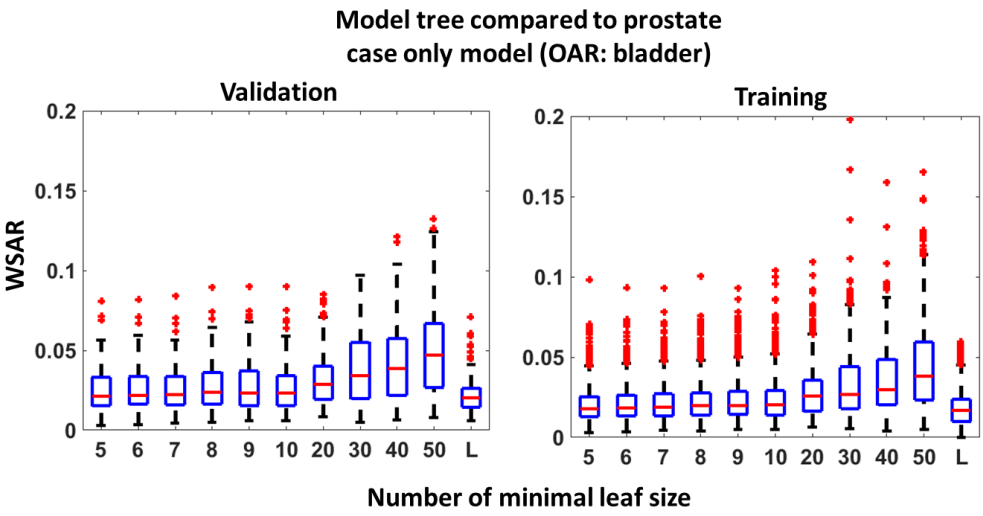

Figure 35: The comparison of the model tree and the single-site prostate model prediction on *prostate* cases for both the training and validation cohort. The OAR is the bladder. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. Statistical significance was observed for the model tree using 20 minimal leaf size and the baseline model.

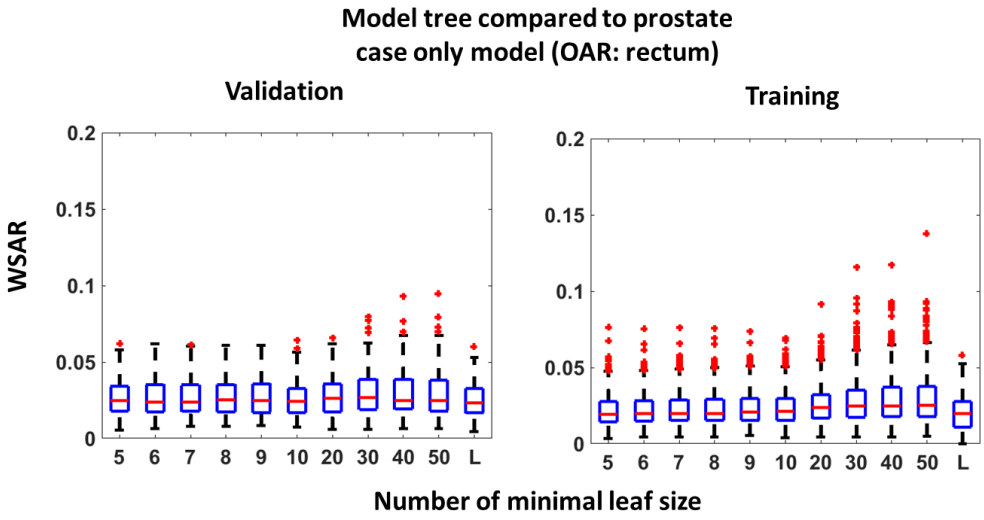

Figure 36: The comparison of the model tree and the single-site prostate model prediction on *prostate* cases for both the training and validation cohort. The OAR is the rectum. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. No statistical significance was observed for the model tree and the baseline model.

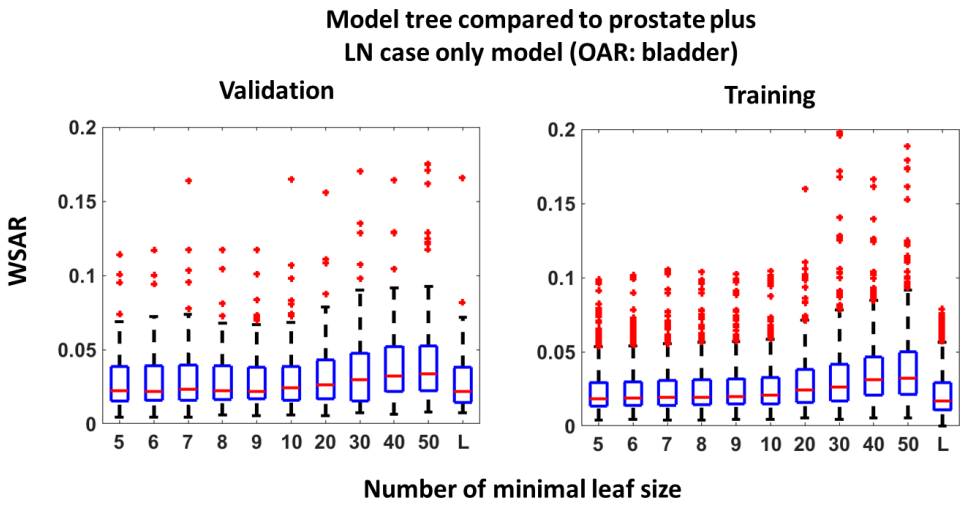

Figure 37: The comparison of the model tree and the single-site prostate plus LN model prediction on *prostate plus LN* **cases for both the training and validation cohort. The OAR is the bladder. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. Statistical significance was observed for the model tree using 40 minimal leaf size and the baseline model.**

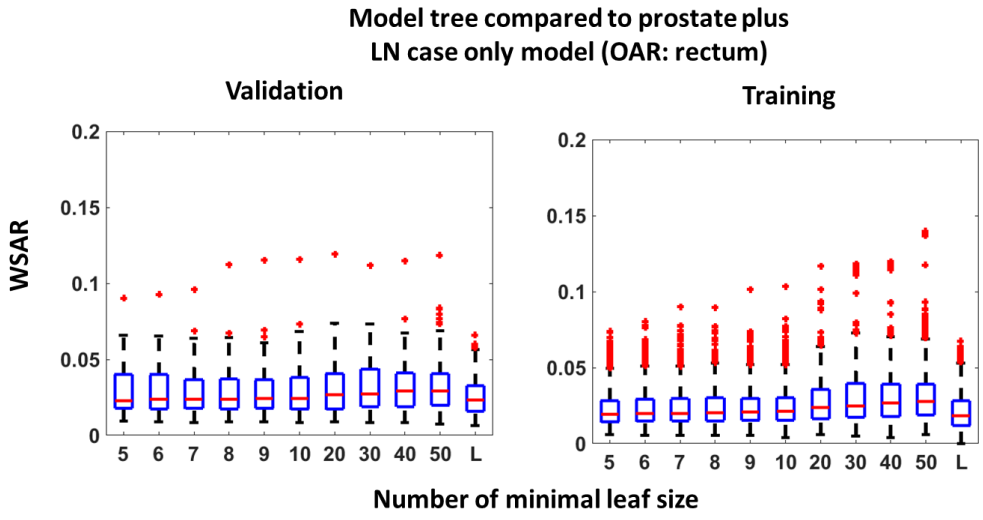

Figure 38: The comparison of the model tree and the single-site prostate plus LN model prediction on *prostate plus LN* **cases for both the training and validation cohort. The OAR is the rectum. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. No statistical significance was observed for the model tree and the baseline model.**

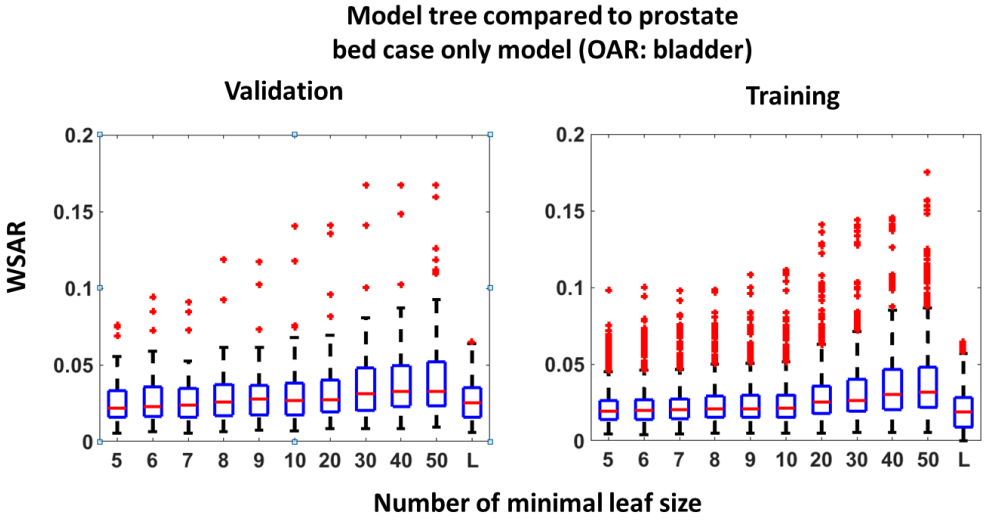

Figure 39: The comparison of the model tree and the single-site prostate bed model prediction on *prostate bed* cases for both the training and validation cohort. The OAR is the bladder. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. Statistical significance was observed for the model tree using 30 minimal leaf size and the baseline model.

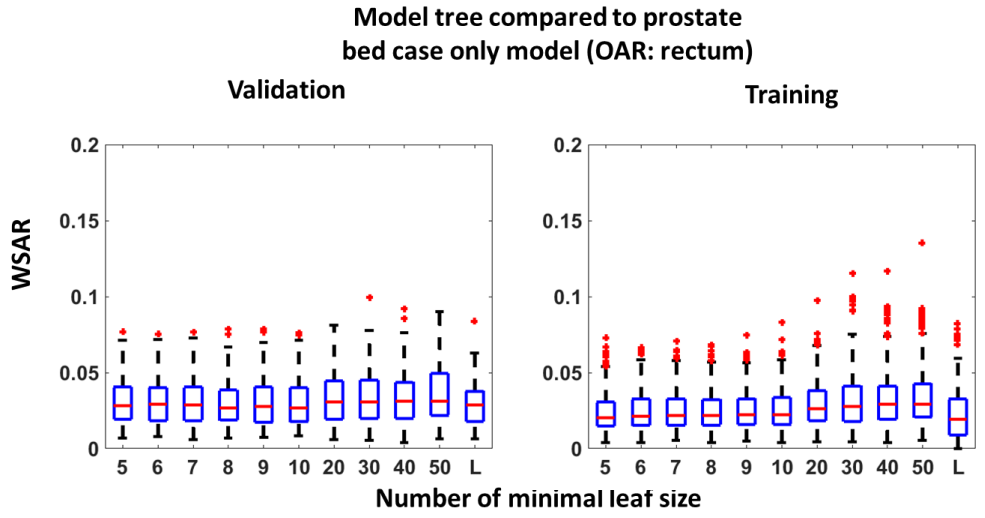

Figure 40: The comparison of the model tree and the single-site prostate bed model prediction on *prostate bed* cases for both the training and validation cohort. The OAR is the rectum. The model tree was trained using various leaf sizes. "L" denotes the baseline linear model. No statistical significance was observed for the model tree and the baseline model.

7.3.2 HN cases

The model tree was trained with the minimal leaf size of 5 for the HN cases. The prediction WSR of the model tree, the standard model and the single model on both the bi-lateral and single-side validation parotids is shown in Figure 41. The left two columns are the prediction WSR of the standard model predicting the bi-lateral cases and the single model predicting single-side cases. The middle two columns are the standard model predicting single-side cases and the single model predicting the bi-lateral cases. The right three columns are the model tree predicting the bi-lateral cases, single-side cases and all validation cases combined. The mean and standard deviation of the WSR for different models is shown in Table 4. The standard model predicting the bi-lateral cases and the single-side model predicting the single-side cases served as the baseline. As shown in the table, the standard model had mean WSR of -0.016 for the bi-lateral validation cases while the single-side model had mean WSR of 0.016 for the single-side validation cases, meaning there was minimal over/under prediction. In contrast, the standard model over predicted the single-lateral cases as shown by the mean WSR. The model tree in general had similar mean WSR as the baseline model prediction with the mean WSR of 0.007 for the bi-lateral cases and -0.014 for the single-side cases.

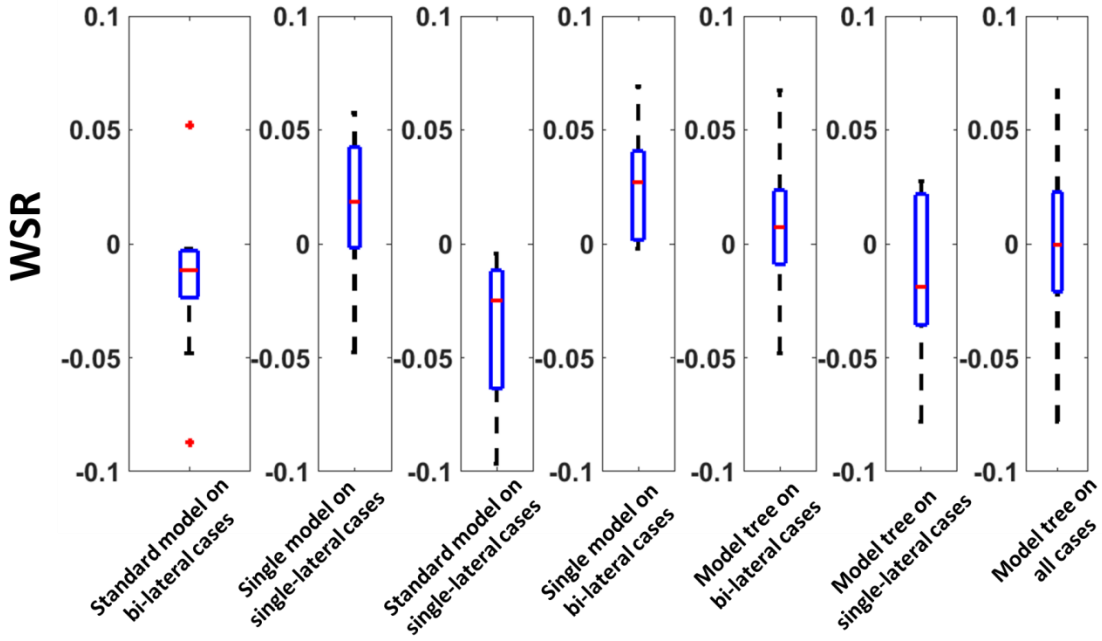

Figure 41: The WSR of the standard model, the single-side model and the model tree on the validation bi-lateral and single-lateral parotids.

Table 4: Mean and standard deviation of the WSR for different models on both the validation bi-lateral and single-lateral cases.

	Standard model on bi-lateral cases	Single model on single-lateral cases	Standard model on single-lateral cases	Single model on bi-lateral cases	Model tree on bi-lateral cases	Model tree on single-lateral cases	Model tree on all cases
Mean	-0.016	0.016	-0.039	0.010	0.007	-0.014	0.001
SD	0.036	0.040	0.037	0.070	0.034	0.042	0.037

The model prediction of the parotid D50 is shown in Figure 42. The difference of the D50 model prediction and the clinical plan D50 is plotted. The mean and standard deviation of the D50 difference for all models are summarized in Table 5. The baseline prediction, i.e. the standard model predicting the bi-lateral cases and the single-side model predicting the single-side cases, had the mean D50 difference below 2 Gy. The standard model over predicted for the single-side cases. The model tree had similar prediction accuracy for the bi-lateral cases. However, a relatively large discrepancy was observed for the model tree predicting the single-side cases possibly due to the mixture of the bi-lateral and single-side cases at certain leaves.

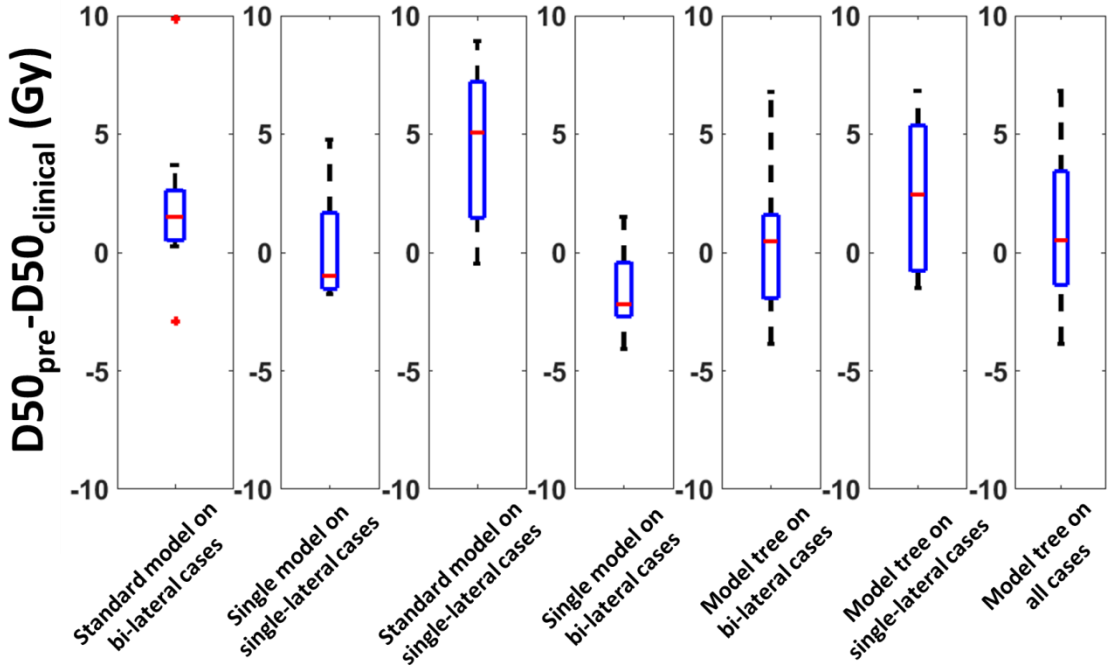

Figure 42: D50 difference of the standard model, the single-side model and the model tree on the validation bi-lateral and single-lateral parotids. The difference of the predicted D50 and the clinical D50 is shown in the boxplot.

Table 5: Mean and standard deviation of the D50 for different models on both the validation bi-lateral and single-lateral cases.

	Standard model on bi-lateral cases	Single model on single-lateral cases	Standard model on single-lateral cases	Single model on bi-lateral cases	Model tree on bi-lateral cases	Model tree on single-lateral cases	Model tree on all cases
Mean(Gy)	2.01	0.24	4.44	-0.68	0.48	2.42	1.13
SD(Gy)	3.33	2.68	3.71	4.17	3.07	3.52	3.24

7.4 Discussion

The model tree was implemented to train a global model to include cases with different clinical conditions and indications. The model tree was able to give similar prediction accuracy as compared to the current baseline single-site model when appropriate minimal leaf size was selected for the pelvic cases. As the number of minimal leaf size increased, the prediction error increased probably due to the fact that cases with different geometry were included in the same leaf. This observation agreed with the result shown in Chapter 5, where the inclusion of geometric novelty cases added to the error of prediction.

As shown by the results, the bladder model gained more benefit from the model tree than the rectum model. It agreed with the observations that the bladder anatomy varies more among three treatment sites than that of the rectum anatomy. It also agreed with the observation from Chapter 5 about the geometric novelty introduction into the model where the existence of the geometric novelty had marginal effect on the model

quality. The model tree could be beneficial for cases that show substantial anatomical variation.

For the HN cases, the model tree had similar prediction accuracy of the WSR for both groups of the validation cases as compared to the baseline model. However, the model tree had difficulty in predicting the D50 for the single-lateral parotids. One possible reason was that for certain leaves, both the bi-lateral and the single-side parotids were used to train the local regression model. The mixture of the cases with different clinical indications may introduce large variation in the model prediction, which is especially detrimental for the single-lateral parotid prediction. In addition, the clinical plan was generated using dose volume constraints at 50% volume with a relatively large weighting. This configuration may result in the local curvature of the DVH line around 50% volume which cannot be reflected in the WSR.

The model tree adopts the philosophy that all cases are segmented based on the anatomical features and then the model is trained locally with cases confined in the same leaf. It reduces the anatomical variations and improves the accuracy accordingly. Compared to the global model trained directly using all cases, the model tree formed a piecewise regression model which characterized the local anatomy-dosimetry features better. A good model tree relies on the number of minimal leaf size, which was showed to be optimal between 5 and 10. The minimal leaf size is sensitive to the number of all cases available for training, since a large leaf size may reduce the depth of tree and limits

the number of features that can be used for segmenting the cases. The depth of the tree needs to be sufficiently large to fully utilize the predictive features to group cases with similar feature pattern. As the number of cases increases, it is possible to increase the minimal leaf size to improve the goodness-of-fit for the local regression.

There are several limitations for the model tree. First of all, only limited subset of the features was used for building the decision tree component. It did not fully utilize all feature information available especially for small training samples. Secondly, limited control of the branch splitting criteria is available. The branch splitting is determined by the single feature that has largest variations, where this feature may not be predictive for the dosimetry information (low correlation). Finally, the model tree may not be suitable for extrapolation. For example, if there is a new case that is an extreme geometric novelty, the prediction based on a certain leaf regression model may not be accurate for that case. Possible alternative methods for building a model using cases with different clinical conditions can be further explored. Such candidates include the clustering-based method. All training cases are clustered based on the feature and then the regression is performed. In the next Chapter, a clustering-based method for building the model tree will be introduced.

7.5 Conclusion

In this Chapter, a novel global model using the model tree is developed. The model tree segments all cases based on the geometry features and performs regression

on the local cases. The model tree shows similar model prediction accuracy as compared to the baseline single-site model when a small minimal leaf size is selected. The model tree is effective in building a global model for cases with different clinical conditions and indications as shown in the example pelvic cases and HN cases.

8. Implementing a clustering-based method to build multiple radiation therapy knowledge models for heterogeneous plan dataset

8.1 Introduction

In the previous Chapter, a global model using the model tree was developed. The model tree is effective in grouping cases based on certain features and performs regression on local aggregated cases. As mentioned in previous Chapter, there are several limitations of the model tree which makes it hard to interpret and regulate the similarity among cases. There are several alternative methods for the model tree.

One competing technique is the clustering-based method. It adopts the philosophy of segmenting all cases into small aggregated clusters and performs the regression locally. It utilizes the similar philosophy as compared to the model tree while the clustering method is superior in the ability of tuning and regulating the segmentation process. Clustering falls into the category of unsupervised learning, where all data in the clustering process has no label to differentiate them. An ideal clustering process shall give aggregated clusters, where all data within the same cluster have close proximity and data from different clusters are well separated. The similarity metric needs to be specified, while usually the distance based metric is used. For example, the Euclidian distance is one of the commonly used similarity measure. In Chapter 3, the *k*-medoids algorithm utilized the Euclidian distance as the similarity measure.

There are many clustering algorithms available which are suitable for specific tasks. Depending on whether the clustering result is deterministic or probabilistic, the clustering algorithms fall into two categories. The first category is the hard clustering. It means that the generated cluster frontier will give specific cluster label for all data. The possible labels are mutually exclusive. Typical hard clustering algorithms include *k*-medoids algorithm and *k*-means algorithm. In contrast to the hard clustering, the other category is the soft clustering. It is a density based estimation algorithm. The clustering algorithm gives the probability of each data falling into each cluster. One typical example is the Gaussian Mixture Model (GMM). It starts with the pre-determined number of Gaussian kernels and tries to fit the data density using the expectation maximization (EM).

Since there is no true label to evaluate the clustering result, external evaluation metric is usually necessary to verify the result. In Chapter 3, the average silhouette width was used to choose the optimal number of clusters. This is based on the goodness of cluster which relies on the similarity measures of the data. An alternative approach is to use the external validation. Once all data are clustered, extra modeling processing can be applied to evaluate the similarity of the cases within a specific cluster. In this Chapter, this method will be used to evaluate the clustering process. The optimal number of the clusters will be determined by the prediction accuracy of the model built with the cases within a cluster. In this part of the study, the same dataset from the previous Chapter

was be used. The clustering-based global model was compared against the baseline model which is the single-site model.

8.2 Materials and methods

In this study, a total of 355 pelvic IMRT plans from three groups were included: group 1 (G1) with 126 low-to-intermediate risk prostate cases; group 2 (G2) with 126 high risk prostate cases treated with lymph node irradiation (LN); group 3 with 103 prostate bed irradiation cases. The OARs included the bladder and rectum.

The clustering-based method combined with the multiple linear regression was proposed to improve model accuracy for a large training dataset with notable geometric variation. The current baseline model was the single-site model, with which the proposed method was compared against. The clustering-based model was trained and compared against the baseline prostate model, prostate plus LN model and prostate bed model respectively.

The clustering-based method started with clustering all cases based on the geometric features. In this study, each case had 22 features which were the same as the previous study. The *k*-means algorithm was used to cluster all cases into designated number of clusters. Then cases from the same cluster were used to train a local model using multiple linear regression model proposed by Yuan et al. For the new case, it was assigned to a specific cluster based on the Euclidian distance and the model trained with the cases from the same cluster was used to predict for this new case. The number of

clusters was initiated prior to the clustering task. In this study, the number of clusters varied between 2 and 10.

To compare the clustering-based model with the prostate model, all G1 cases were divided into 10 folds, and 9 folds together with all G2 and G3 cases were used to train the clustering-based model. The same 9 folds were used to train the baseline single-site prostate model. The remaining 1 fold was reserved for validation purpose. 10-fold cross validation was performed and the WSAR for all validation cases was recorded. The mean WSAR of all validation cases was calculated and compared for clustering-based method and the baseline model. The comparison was done to compare the clustering-based model with the prostate plus LN and prostate bed model as well. The flowchart is shown in Figure 43.

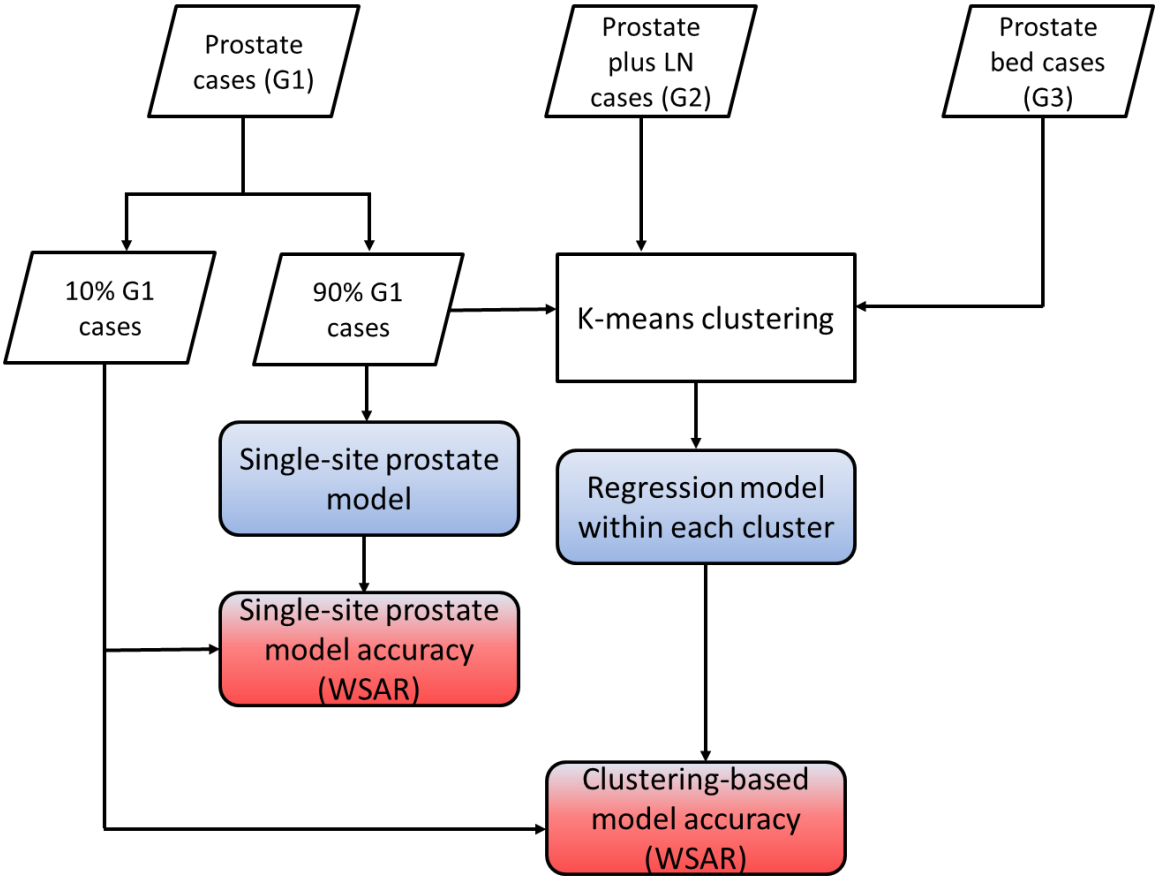

Figure 43: Flowchart of the clustering-based model training using *k*-means algorithm and the single-site prostate model training. The validation was performed by applying the trained model to the extra one fold of prostate cases and calculating the WSAR.

8.3 Results

The *k*-means clustering was performed to segment all cases into specific number of clusters. One example clustering result is shown in Figure 44. It shows the clustering result using 90% prostate bed cases, all prostate and prostate plus LN cases. The number of cases within each cluster varied from the minimum of 21 to the maximum of 114, showing large variation in the number of cases in each cluster.

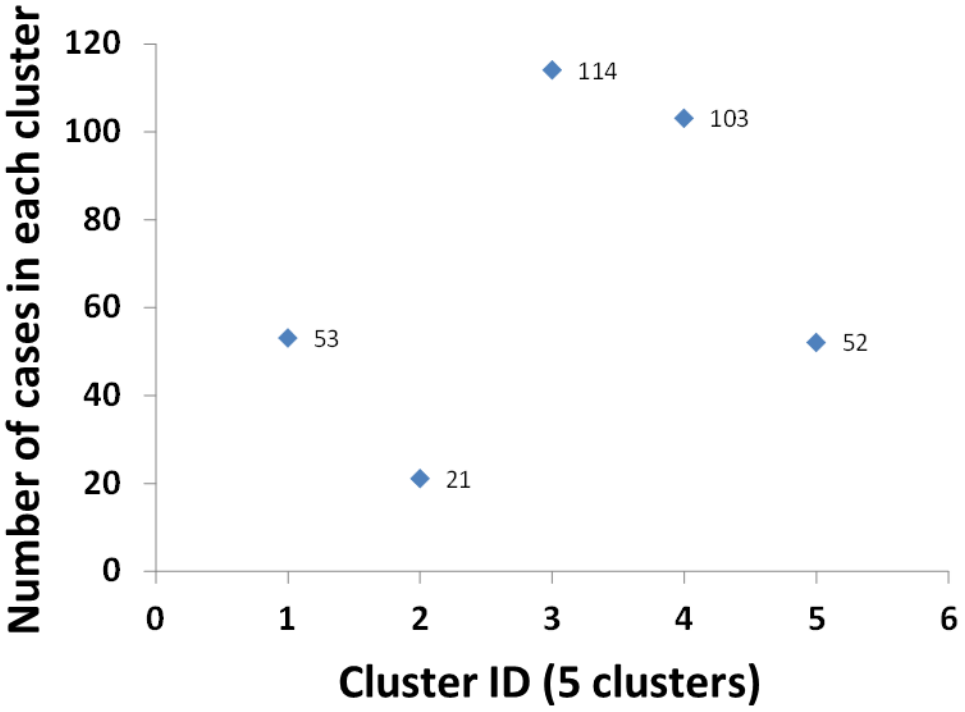

Figure 44: The number of cases within each of the five clusters for the *k*-means algorithm for the model trained with 90% prostate bed cases, 100% prostate cases and 100% prostate plus LN cases.

The WSAR of all validation cases were recorded for the clustering-based model and the baseline single-site model. The distribution of WSAR is shown in Figure 45. The horizontal axis shows different numbers of clusters and "S" denotes the single-site model. The comparison between the clustering-based model and the prostate model is shown on the first row, while the comparison between the clustering-based model and the prostate plus LN model is shown on the middle row. The comparison between the clustering-based model and the prostate bed model is shown on the bottom row. The bladder model is shown on the left column and the rectum model is shown on the right

column. For all comparisons, the clustering-based model showed similar prediction accuracy for the validation cases as compared to the baseline model. The mean WSAR of the clustering-based model and the baseline model are shown in Table 6. The mean WSAR of different numbers of the clusters in general agreed with that of the baseline model with the accuracy of 0.01 WSAR.

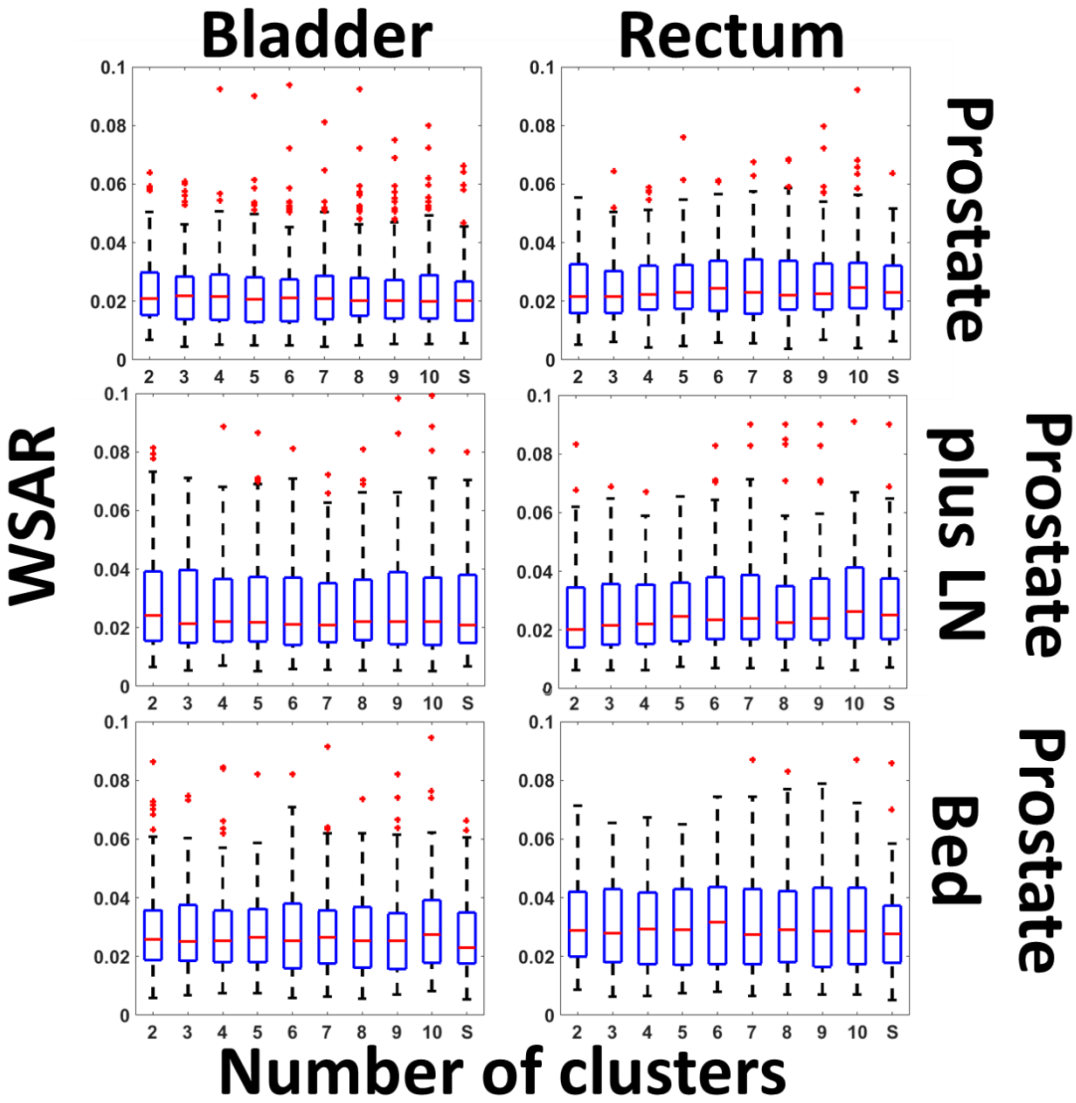

Figure 45: Boxplots of the WSAR of the validation cases using the clustering-based model and the single-site model. The first row is the clustering-based model compared with the single-site prostate model. The second row is clustering-based model compared with the single-site prostate plus LN model. The third row is clustering-based model compared with the single-site prostate bed model. The bladder prediction is shown on the left and the rectum prediction is shown on the right. The horizontal axis is the number of clusters. "S" denotes the single-site model. The clustering-based model agreed with the single-site model despite of the number of clusters specified.

Table 6: The mean WSAR of the validation cases for the clustering-based model with varying numbers of clusters (2, 3, 4, 5, 6, 7, 8, 9, 10) and the baseline single-site model. The clustering-based method agreed with the single-site model with the accuracy of 0.01 mean WSAR.

Number of clusters	2	3	4	5	6	7	8	9	10	Single-site model
G1 bladder model	0.0239	0.0239	0.0236	0.0232	0.0236	0.0235	0.0237	0.0233	0.0245	0.0224
G1 rectum model	0.0245	0.0244	0.0251	0.0259	0.0258	0.0259	0.0263	0.0263	0.0268	0.0252
G2 bladder model	0.0291	0.0278	0.0277	0.0287	0.0278	0.0272	0.0285	0.0278	0.0283	0.0278
G2 rectum model	0.0260	0.0262	0.0260	0.0276	0.0288	0.0295	0.0280	0.0287	0.0309	0.0286
G3 bladder model	0.0299	0.0285	0.0288	0.0288	0.0276	0.0284	0.0276	0.0282	0.0304	0.0270
G3 rectum model	0.0317	0.0304	0.0303	0.0307	0.0321	0.0316	0.0312	0.0314	0.0323	0.0290

8.4 Discussion

This study analyzed the clustering-based method to accommodate large variation of the geometry features in the training data cohort. *K*-means clustering algorithm was used to first segment all cases into multiple clusters. Then cases from the same cluster were used to train the local regression model. The clustering-based method served as an alternative for the model tree which is discussed in the previous Chapter. The *k*-means clustering method used Euclidian distance as the similarity metric, where all input features were used to calculate the Euclidian distance. The *k*-means algorithm started with a random subset of cases as the cluster mean. All cases identified their cluster assignment based on the minimum Euclidian distance to the cluster mean. Once the cluster assignment was settled, the cluster mean will be updated based on all cases belonging to the same cluster. The cluster assignment process then repeated. The clustering process stopped when the cluster assignment did not change. The *k*-means algorithm is computationally expensive if exhaustive search is required for the global optimal solution. Therefore the initial cluster mean is usually randomly generated. The *k*-means algorithm utilizes all features which is hard for the model tree to accomplish. The model tree bifurcates at each node using only one feature, while the number of features used is usually limited by the depth of the tree and the total number of cases available. The number of clusters for the *k*-means algorithm is determined by the user,

which can incorporate the clinical interpretation. For example, depending on the various treatment sites included in the training dataset, the number of clusters can be determined *in priori*. In contrast, the model tree needs to adapt the minimal leaf size depending on the number of features included as well as the number of cases available.

The result from this study showed that the clustering-based model had similar prediction accuracy as the baseline single-site model regardless of the number of clusters. The result demonstrated that segmenting all cases into two clusters was sufficient to improve the model accuracy for the current dataset. The benefit for the clustering-based method would be more prominent for other dataset, for example, a mixture of pelvic cases and HN cases.

This study used the *k*-means algorithm to perform the clustering task. There are many other alternative methods, among which the *k*-medoids algorithm can also help segment the cases. The *k*-medoids algorithm was used in Chapter 3 to aid cluster all prostate cases into multiple clusters to form the prostate atlas. The difference of the *k*-means and the *k*-medoids algorithm is that the *k*-medoids algorithm asks for a specific case to serve as the cluster mean while the *k*-means algorithm calculates the mean value of all cases in the same cluster to be the cluster center. The *k*-medoids algorithm was used since a specific case needed to be the center of the cluster and served as the atlas case. In Chapter 3, the average silhouette width was used to evaluate the cluster

tightness. In this Chapter, the optimal number of clusters can be determined by the accuracy of the proposed model although all models showed similar accuracy.

There are several limitations for this study. First, as reflected in Figure 41, the clustering-based method may result in clusters of different sizes. Unlike the atlas-based method, the multiple linear regression method relies on a sufficient number of training cases to improve the statistical power. A small cluster is more prone to be overfitting. Since the *k*-means algorithm does not try to balance the number of cases within each cluster, other clustering algorithms can be explored in the future to guarantee the clusters with similar size. Secondly, the *k*-means algorithm takes in all features to calculate the Euclidian distance. It is possible that some features do not differentiate cases from different treatment sites. One another step of feature selection could be implemented prior to calculating the Euclidian distance. Since the current method of calculating the Euclidian distance is capable of matching the model accuracy with the baseline model, it could be beneficial to implement the feature selection for more complicated tasks. At last, the clustering-based method only used the input geometry features. The output feature was not used to aid the clustering process. Therefore, the bladder model and the rectum model had the same clustering result. Future study could incorporate the output label/value to improve the performance of the clustering-based method.

8.5 Conclusion

In this Chapter, a clustering-based model is proposed to train a global model for cases with different clinical conditions. It serves as an alternative method for the model tree proposed in Chapter 7. The *k*-means algorithm is used to segment all cases into clusters and the regression model is trained within each cluster. The proposed clustering-based model has similar prediction accuracy as compared to the baseline single-site model. The clustering-based model makes use of all geometry features to calculate the Euclidian distance for the clustering task. It contrasts the model tree which uses single feature to split the cases into branches.

9. Facilitating the incremental learning of radiation therapy knowledge model

9.1 Introduction

The KBP approach for providing guidance for radiation therapy treatment planning relies on the quality of training data. Unlike the MCO or heuristic optimization approach which does not relies on previous knowledge, the KBP approach asks for meticulous inspection of the quality of the training data before implementing the KBP model into the clinic. Several issues have been addressed in previous Chapters, including the outlier identifications and building a global model. These topics fall under the umbrella of rapid learning, which tries to mine the knowledge in a rapid fashion and adapts the model as the data accumulate. In real clinical settings, the training data for the model is usually small at the beginning, and increases over time. Questions remain about how to deal with the incoming new data. Possible solutions include re-training the model, or retaining the new data and retraining the model until certain amount of new cases accumulate. This Chapter will provide the answer for whether re-training the model is necessary for various initial data size when there are new incoming cases.

The concept of incremental learning started in the industry (Geng & Smith-Miles, 2009). For some machine learning schemes, it is sometime impractical to re-train the model whenever there is a new case. It may be time-consuming to re-train the model using the whole dataset. For example, the face recognition system is built upon an initial dataset. And if there is new subset of face patterns available for the inclusion into the

system, usually re-training the model is unnecessary and the extra knowledge or pattern is added into the model. Similar as the tasks in the industry, the radiation therapy knowledge modeling requires effort for the data inspection and model training. Therefore sometimes it is not necessary to re-train the model each time there is a new case available. This part of the study will analyze when re-training the model is suitable in order to improve the efficiency while avoiding unnecessary model training.

Incremental learning for radiation therapy treatment planning knowledge model is important in the sense that re-training and inspecting the model is time consuming for the current clinical practice. The KBP model has not yet been widely implemented in the clinic and the concept of KBP is relatively new for most physics staffs. For the clinic and institution where the KBP model is being used, the physics staff with expertise in model training is responsible for managing the modeling process, including collecting the previous treatment plans for training, training the model based on the collected plans and inspecting/validating the model quality. It is time consuming to go through the entire pipeline to publish a model for clinical use. Re-training the model based upon each new treatment plan is often unrealistic in terms of staff's time and effort. And in addition it may be unnecessary or unbeneficial to re-train the model frequently. It would be beneficial to know what the frequency of re-training the model is based on the new observation. A balance between the clinical workflow efficiency and the model accuracy needs to be identified.

This study used the pelvic cases as an example. The initial model was trained with various training sample size (namely 20, 40, 60). Then new cases emerged and were added to the model to initiate the model re-training. The model prediction accuracy was compared among the initial model and the new models trained with the inclusion of new cases.

9.2 Materials and methods

9.2.1 Materials

In this study, a total of 355 pelvic IMRT plans from three groups were included: group 1 (G1) with 126 low-to-intermediate risk prostate cases; group 2 (G2) with 126 high risk prostate cases treated with lymph node irradiation (LN); group 3 with 103 prostate bed irradiation cases. The OARs included both the bladder and the rectum.

9.2.2 Single-site incremental learning

A baseline model was trained using a certain training sample size (20, 40 or 60) from a specific treatment group. Then new cases from the same group were progressively added to the model to mimic the scenario of emerging new cases. The number of new cases simulated included 1, 2, 3, 4, 5, 10, 15, 20, 25 and 30. The original training dataset was combined with the emerging new cases to train the updated model. Each model, including the baseline model, the updated models with the inclusion of 1, 2, 3, 4, 5, 10, 15, 20, 25 and 30 new cases, was validated on an extra subset of 10 cases from the same treatment group. All the training cases, the emerging new cases and the

validation cases were randomly selected from each treatment group. For validation purpose, the experiment was repeated 20 times using bootstrap. The proposed experiment was performed for G1, G2 and G3 group respectively. Both the bladder and the rectum model were simulated for each group.

The WSAR for all validation cases were recorded to denote the prediction accuracy as previous Chapters. The mean WSAR of all 10 validation cases was recorded for each bootstrap (a total of 20 bootstraps). The mean WSAR of all bootstraps was recorded and plotted to compare against different models. The flowchart is shown in Figure 46.

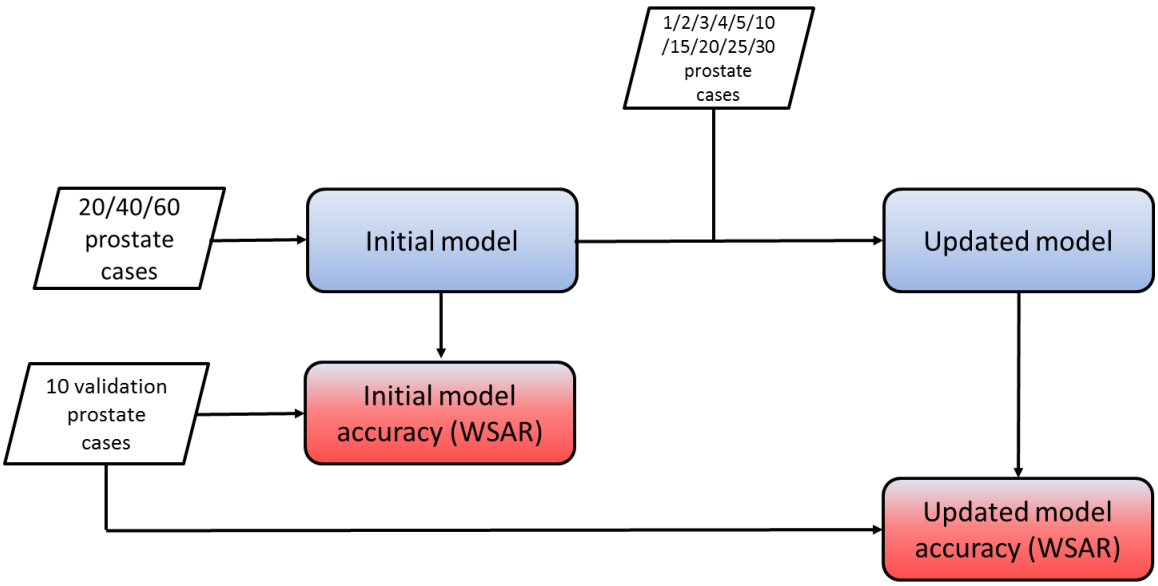

Figure 46: Flowchart of the incremental learning for the single-site prostate model. An initial model was trained with 20/40/60 cases. Additional 1/2/3/4/5/10/15/20/25/30 cases were progressively added to the model to generate the updated model. An extra subset of 10 prostate cases was used to evaluate the initial model and the updated model based on the WSAR.

9.2.3 Multiple-site incremental learning

The model was trained incrementally using cases from multiple treatment sites. In this part of the study, a baseline model was trained using 10 G1 cases and 10 G2 cases. Additional G1 and G2 cases were added to the model to mimic the scenario of emerging new cases. 2, 4, 6, 8, 10, 20, 30, 40, 50 and 60 G1 and G2 cases (G1 and G2 cases added in pairs) were progressively added to the baseline model trained with 20 cases. All models were validated on the extra 10 G1 and 10 G2 cases. The experiment was repeated 20 times using bootstrap. Both the bladder and the rectum model were simulated. The mean WSAR of all 20 validation cases was recorded for each bootstrap. The flowchart for the model trained with G1 and G2 cases is shown in Figure 47.

In addition to the model trained with cases from two treatment groups (G1 and G2), all three groups of cases were included to train the model in an incremental fashion. A baseline model was trained using 10 G1, 10 G2 and 10 G3 cases. 3, 6, 9, 12, 15, 30, 45, 60, 75 and 90 G1, G2 and G3 cases (equal number of G1, G2 and G3 cases) were progressively added to the baseline model trained with 30 cases. An extra subset of 10 G1, 10 G2 and 10 G3 cases served as the validation dataset. The experiment was again repeated for 20 times using the bootstrap. Both the bladder and the rectum model were simulated. The mean WSAR of all 30 validation cases was recorded for comparison among all models.

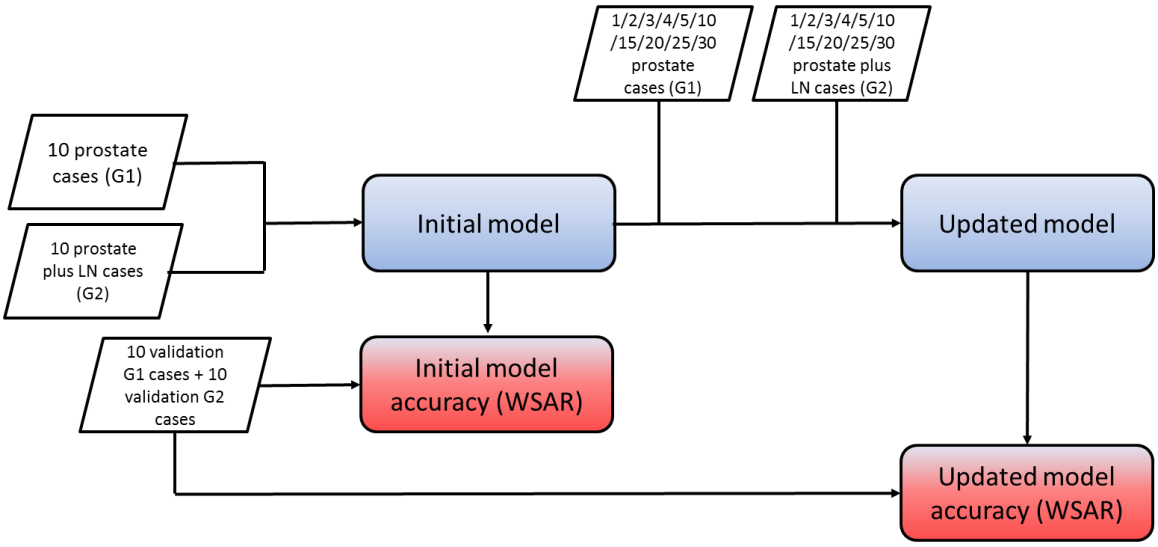

Figure 47: Flowchart of the incremental learning for the multiple-site model trained with G1 and G2 cases. An initial model was trained with 10 G1 and 10 G2 cases. Additional 1/2/3/4/5/10/15/20/25/30 G1 and 1/2/3/4/5/10/15/20/25/30 G2 cases were progressively added to the model in pairs to generate the updated model. An extra subset of 10 G1 and 10 G2 cases was used to evaluate the initial model and the updated model based on the WSAR.

9.3 Results

9.3.1 Single-site incremental learning

The baseline model and the updated model were validated on the extra set of cases. The mean WSAR of the validation cases was recorded for both the bladder model and the rectum model. The mean WSAR distribution of all 20 bootstraps for the bladder model is shown in Figure 48, and the mean WSAR distribution for the rectum model is shown in Figure 49. Each row represents the prostate model, the prostate plus LN model and the prostate bed model respectively from the top to the bottom. The initial sample size varied among 20 cases (left column), 40 cases (middle column) and 60 cases (right column). As observed from Figure 48 and 49, the inclusion of new cases showed

marginal improvement over the original 20-case model. The 40-case and 60-case model showed relatively stable model accuracy despite of the inclusion of new cases. Comparing the figures on the same row, the model trained with large dataset tended to have better accuracy than the small training dataset, indicating the benefit of including more cases in the model. However, this improvement of the accuracy relied on sufficiently large new training cases rather than an individual one. If the criterion of re-training the model was set as 5% improvement of WSAR, 10 cases were needed to invoke re-training the model for the 20-case prostate case bladder model. For the 20-case prostate plus LN and prostate bed case bladder model, the re-training criterion was 10 and 10, respectively as well. For the rectum model, the re-training criterion for the 20-case prostate case model, prostate plus LN case model and the prostate bed case model was 4, 10 and 10, respectively. The detailed mean WSAR of all models for the bladder and the rectum is shown in Table 7 and 8, respectively. The model quality was in general consistent with the variation of 0.003 mean WSAR among all models with different training sizes.

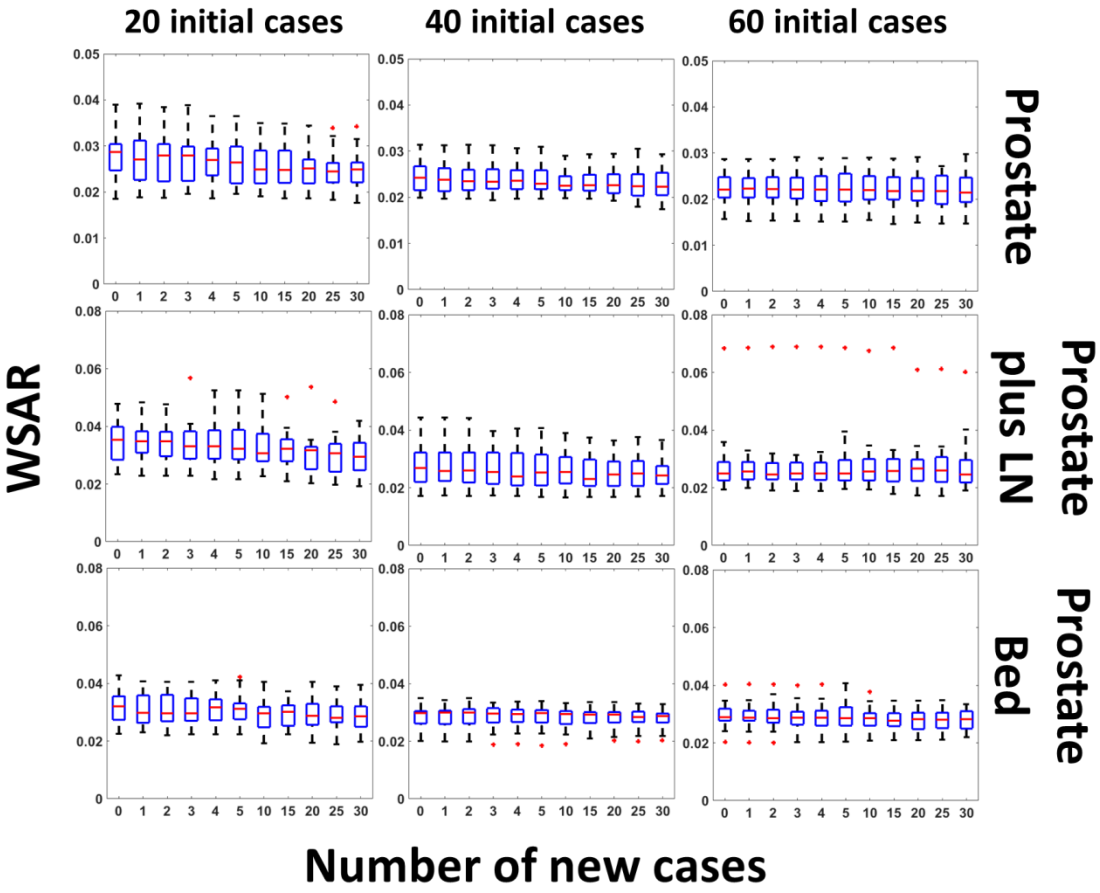

Figure 48: Boxplots of the WSAR of the validation cases using the initial model and the updated model. The first row is the prostate model. The second row is the prostate plus LN model. The third row is the prostate bed model. Initial models trained with 20, 40 and 60 cases are shown in the left, middle and right column respectively. The horizontal axis denotes the number of new cases added to the updated model where "0" denotes the initial model. These boxplots show the prediction for the *bladder*.

Table 7: Mean WSAR of the validation cases from the initial model and the updated model with different training settings. "0" denotes the initial model. The number of initial training cases is indicated in the bracket. The first row presents the number of added new cases. This table summarizes the statistics for the *bladder*.

Number of new cases	0	1	2	3	4	5	10	15	20	25	30
Prostate (20)	0.0276	0.0271	0.0270	0.0272	0.0269	0.0265	0.0257	0.0255	0.0251	0.0246	0.0248
Prostate (40)	0.0244	0.0241	0.0240	0.0241	0.0240	0.0238	0.0232	0.0234	0.0234	0.0232	0.0229
Prostate (60)	0.0221	0.0220	0.0220	0.0221	0.0222	0.0224	0.0223	0.0221	0.0218	0.0218	0.0217
Prostate plus LN (20)	0.0346	0.0343	0.0340	0.0337	0.0338	0.0333	0.0323	0.0319	0.0303	0.0300	0.0297
Prostate plus LN (40)	0.0278	0.0277	0.0275	0.0271	0.0264	0.0265	0.0263	0.0255	0.0256	0.0255	0.0252
Prostate plus LN (60)	0.0300	0.0302	0.0298	0.0299	0.0300	0.0283	0.0280	0.0276	0.0275	0.0274	0.0273
Prostate bed (20)	0.0320	0.0310	0.0308	0.0307	0.0311	0.0306	0.0291	0.0294	0.0294	0.0289	0.0290
Prostate bed (40)	0.0290	0.0290	0.0289	0.0287	0.0287	0.0287	0.0281	0.0279	0.0284	0.0282	0.0278
Prostate bed (60)	0.0294	0.0293	0.0292	0.0290	0.0290	0.0293	0.0286	0.0281	0.0277	0.0278	0.0279

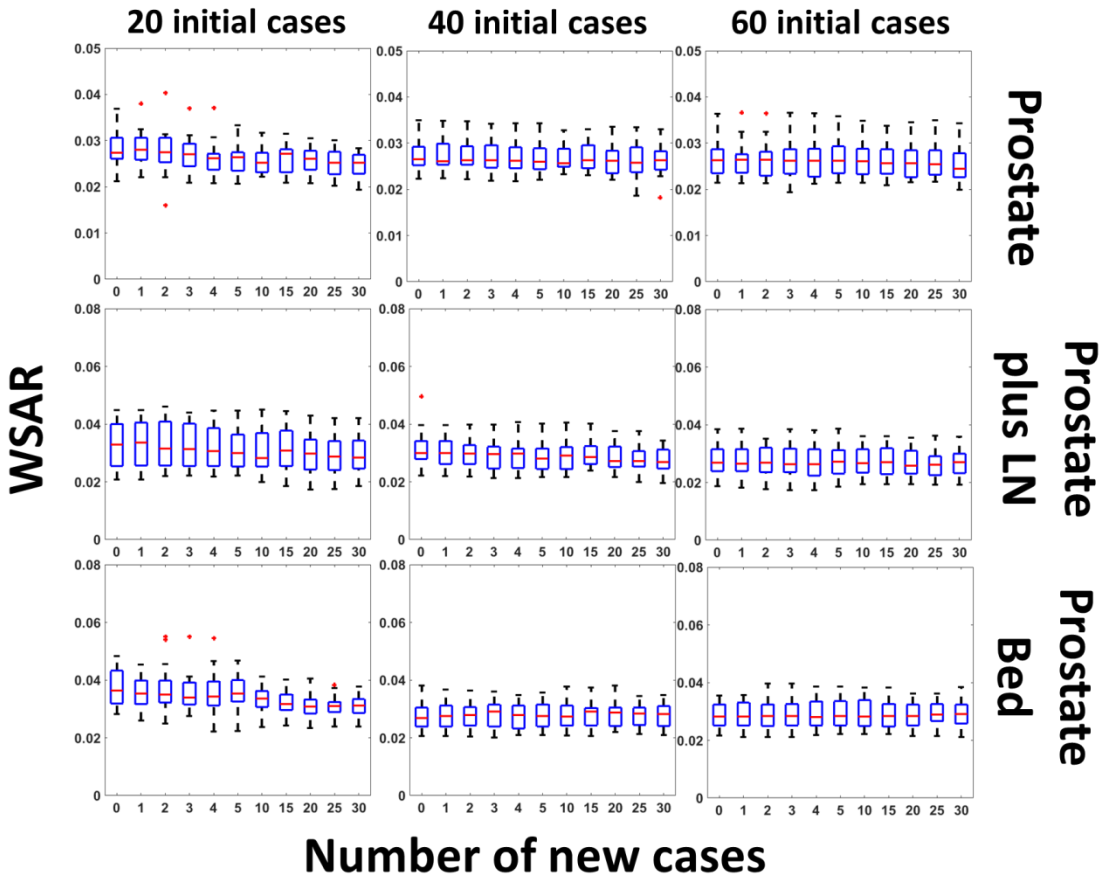

Figure 49: Boxplots of the WSAR of the validation cases using the initial model and the updated model. The first row is the prostate model. The second row is the prostate plus LN model. The third row is the prostate bed model. Initial models trained with 20, 40 and 60 cases are shown in the left, middle and right column respectively. The horizontal axis denotes the number of new cases added to the updated model where "0" denotes the initial model. These boxplots show the prediction for the *rectum*.

Table 8: Mean WSAR of the validation cases from the initial model and the updated model with different training settings. "0" denotes the initial model. The number of initial training cases is indicated in the bracket. The first row presents the number of added new cases. This table summarizes the statistics for the *rectum*.

Number of new cases	0	1	2	3	4	5	10	15	20	25	30
Prostate (20)	0.0280	0.0282	0.0276	0.0273	0.0260	0.0259	0.0255	0.0261	0.0257	0.0252	0.0249
Prostate (40)	0.0273	0.0273	0.0272	0.0271	0.0268	0.0267	0.0268	0.0269	0.0264	0.0264	0.0265
Prostate (60)	0.0266	0.0263	0.0262	0.0262	0.0262	0.0265	0.0263	0.0263	0.0262	0.0263	0.0257
Prostate plus LN (20)	0.0325	0.0329	0.0326	0.0322	0.0317	0.0312	0.0307	0.0314	0.0298	0.0294	0.0294
Prostate plus LN (40)	0.0310	0.0300	0.0298	0.0290	0.0291	0.0288	0.0290	0.0293	0.0286	0.0278	0.0277
Prostate plus LN (60)	0.0276	0.0274	0.0274	0.0274	0.0272	0.0278	0.0275	0.0274	0.0268	0.0262	0.0266
Prostate bed (20)	0.0374	0.0358	0.0367	0.0356	0.0356	0.0358	0.0333	0.0320	0.0310	0.0308	0.0310
Prostate bed (40)	0.0275	0.0277	0.0278	0.0275	0.0279	0.0280	0.0280	0.0280	0.0282	0.0277	0.0279
Prostate bed (60)	0.0286	0.0288	0.0291	0.0294	0.0295	0.0295	0.0294	0.0291	0.0287	0.0291	0.0290

9.3.2 Multiple-site incremental learning

Figure 50 shows the boxplots of model prediction accuracy for different numbers of new cases. The first row shows the models trained with G1 and G2 cases and the second row shows the models trained with all three groups of cases. The bladder model is shown on the left column while the rectum model is shown on the right column. There was a prominent trend of improved model accuracy as the number of cases increased. For the action threshold of 5% mean WSAR improvement, 6 cases were necessary to invoke model re-training for the G1 + G2 bladder and rectum model. For the model trained with all three treatment groups, 9 cases were necessary for the bladder model and 30 cases were necessary for the rectum model. The mean WSAR of each model is shown in Table 9. The models with different numbers of training cases agreed with each other in general with the variation of 0.01 mean WSAR.

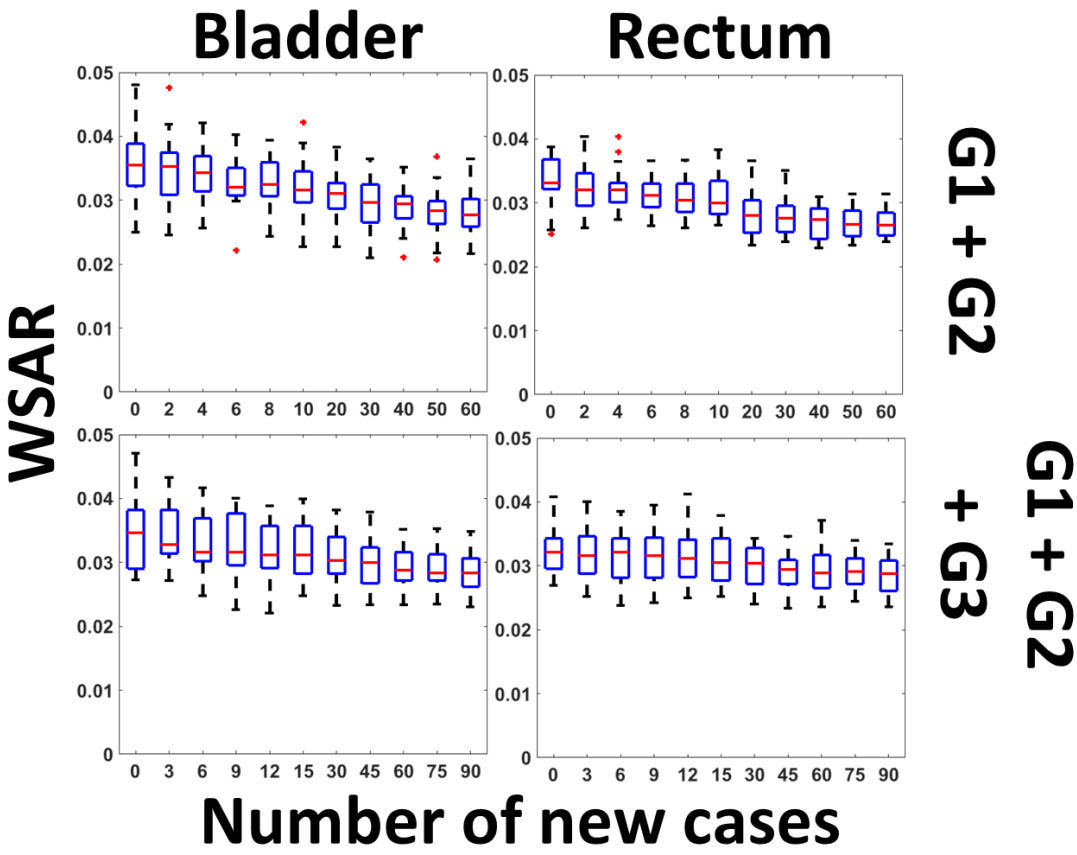

Figure 50: Boxplots of the WSAR of the validation cases using the initial model and the updated model. The model was trained with G1 + G2 cases (first row, 20 cases for the initial model) or the G1 + G2 + G3 cases (second row, 30 cases for the initial model). The initial model was trained with 10 cases from each treatment group. The bladder prediction is shown on the left and the rectum prediction is shown on the right. The horizontal axis denotes the number of new cases added to the updated model where "0" denotes the initial model. Equal number of cases from each group was added to the updated model.

Table 9: Mean WSAR of the validation cases from the initial model and the updated model with different training settings. "0" denotes the initial model. The top three rows show the model trained with G1 + G2 cases. The bottom three rows show the model trained with G1 + G2 + G3 cases.

Number of new cases	0	2	4	6	8	10	20	30	40	50	60
G1 + G2 bladder	0.0361	0.0346	0.0345	0.0330	0.0326	0.0319	0.0306	0.0293	0.0290	0.0281	0.0279
G1 + G2 rectum	0.0334	0.0326	0.0322	0.0315	0.0311	0.0308	0.0284	0.0280	0.0269	0.0268	0.0268
Number of new cases	0	3	6	9	12	15	30	45	60	75	90
G1 + G2 + G3 bladder	0.0346	0.0343	0.0330	0.0327	0.0316	0.0318	0.0308	0.0298	0.0292	0.0291	0.0286
G1 + G2 + G3 rectum	0.0323	0.0319	0.0315	0.0318	0.0315	0.0310	0.0300	0.0295	0.0292	0.0292	0.0286

9.4 Discussion

This Chapter analyzed the increment learning of radiation therapy knowledge model. It answers the question when re-training the model is necessary. Based on the result of this study, re-training is only necessary for small training sample (~20) at the frequency between every 4 to 10 cases. As the number of training cases accumulated, re-training became less frequent and less necessary.

The proposed method can also be extended to other treatment sites to identify the frequency of re-training the model, for example the HN and lung cases. The pelvic case geometry is relatively consistent, which may ask for less number of cases to reach the best achievable model accuracy. For more complicated treatment geometry, more cases may be necessary to saturate the prediction accuracy as indicated by Boutilier et al. Larger required sample size may be correlated to more frequent model update and it remains to be confirmed in the future study.

This study analyzed the single-site and multiple-site knowledge model. The concept of incremental learning can also be generalized to the global model introduced in Chapter 7 and 8. The decision of re-training the model needs to take into consideration of the concentration of cases from different treatment sites, as it may affect the decision tree growth and the clustering result.

There are several limitations in this study. First, the proposed method recommends re-training the model with every 4-10 new cases. This treatment does not

take into consideration of the each case's specific dosimetry-geometry relation with respect to the population mean. Instead, a general solution was provided for how frequent re-training the model is necessary. It is possible that a certain case represents a unique dosimetry-geometry relation and should be reserved for case based reasoning. Ideal scenario is that a well-trained model knows whether the new case is suitable for population based regression prediction or specific case based reasoning. This framework needs to be constructed in the future study. In addition, the proposed incremental learning did not take into consideration of geometric novelty and dosimetric outliers. This is more prominent for small training sample since the geometric variation is small for such data size. As indicated by previous Chapters, it is always recommended to clean the geometric novelty and dosimetric outliers to improve the model accuracy. However for small training sample (~20), the inclusion of geometric novelty could potentially increase the model generalization capability. And the dosimetric outliers can become inliers as the training sample size increases. Future study is warranted to establish an automatic model training system which self-adapts the knowledge based on the new observation.

9.5 Conclusion

The knowledge-based model for pelvic cases is trained in an incremental fashion. It is found that re-training the model is necessary for small training sample for every 4-10 new cases. Re-training is less frequent as the training sample size reaches 40 for the

pelvic cases. Such recommendation can avoid unnecessary repetitive model re-training which does not provide substantial improvement of the model quality.

10. Summary

This study focused on the current challenges and limitations of TPKDD and provided tools to aid clinical implementation of the knowledge-based models. There were three challenges addressed in this study. The first part tried to improve the knowledge-based model training efficiency. An atlas-guided treatment planning framework was proposed for prostate IMRT treatment planning. Unlike traditional regression based model training, the atlas-based method relied on only 5 atlas cases to provide treatment planning guidance. The dose contour of the atlas case was deformed towards the query case's anatomy using the deformable image registration. The deformed dose, the *goal dose*, was conformal to the target shape and served as the 3D dose guidance. The 3D dose guidance can be converted to 1D dose-volume constraints to guide inverse IMRT optimization, or the 3D dose can directly guide the fluence optimization using linear programming. The proposed atlas-guided treatment planning framework was able to provide similar plan quality as compared to the clinically treated plan.

The second part of the study tried to provide solution to unsolved clinical tasks to improve the clinical efficiency. Due to the wide spectrum of delivery technique and planning technique, the knowledge-based model is usually limited to a specific delivery technique and/or planning technique. The widely available knowledge-based model for inverse IMRT planning is not applicable for the manual forward planning. One example

is the whole breast ECOMP planning. The human planner takes between one and four hours to achieve a clinically acceptable plan. This part of study proposed an automatic treatment planning tool to accelerate the treatment planning process for WBRT. A three-step process was proposed to mimic the clinical workflow. First, a PCA based energy selection tool was developed to identify single energy or mixed energy for the patient. Secondly, a random forest model was developed to predict the fluence intensity for each pixel on the fluence map. At last, the proposed workflow concluded with one time fluence fine tuning to further enhance the dose homogeneity within the treatment volume. The energy selection tool was able to provide identical energy choice for 19 out of 20 retrospective validation cases. The automatically generated plan was comparable with the clinically treated plan in terms of the treatment volume coverage as well as the hotspot volume. The proposed automatic planning system achieved substantial reduction in the treatment planning time and realized near-real-time planning.

The third part of the study worked on rapid learning of radiation therapy knowledge-based model. It is an important step towards the clinical implementation of the knowledge-based model. Many topics were addressed in this part of the study. First of all, a systematic workflow was proposed to improve the model quality. Statistical tools, the *leverage* and the studentized residual, were used to identify geometric novelty and dosimetric outlier cases. The proposed workflow demonstrated the feasibility of identifying the geometric novelty and dosimetric outlier cases, and the benefit of

excluding them from the model training. Secondly, a one class SVM followed by ROUT method was proposed to identify a batch of geometric novelty and dosimetric outlier cases. The proposed method was validated on four groups of pelvic cases including deliberately added dosimetric outlier cases. This method was able to identify the geometric novelty cases as well as all intentionally added dosimetric outlier cases. The model trained with the cleaned subset of training cases showed improved model quality. Thirdly, this part of the study also addressed building a global model to include training cases with different clinical indications and conditions. The model tree was implemented to train a decision tree based regression model for the pelvic cases. In addition, the model tree was also implemented on the HN tradeoff model. The result showed that the model tree was able to provide similar prediction accuracy as compared to the baseline model trained with cases of same clinical indication and condition. In addition, a clustering-based method was proposed to build a global model. It served as an alternative method to the model tree while the clustering-based method made full use of all geometry features available. The *k*-means algorithm was called upon to perform the clustering task. The result showed that the two-cluster based model was able to predict similarly as the baseline single-site model. This part of the study was concluded with the implementation of incremental learning. This study answered the question when re-training the model is necessary. A model was trained with small training samples and new cases emerged to perturb the model. The model trained with

original dataset was compared against the models with the inclusion of new cases. The result showed that, for the pelvic cases, re-training the model was necessary for every 4-10 new cases when about 20 cases were used to train the original model. As the number of total cases accumulated, the re-training became less frequent and less necessary.

The first two parts of the study focused on the data mining component in the KDD workflow, while the third part of the study focused on the data selection and data preprocessing. All three parts of this study provided solution and suggestion for the wide clinical implementation of TPKDD. The proposed methods and techniques, together with the results, provided detailed information of the knowledge-based approach to accelerating the clinical workflow and improving the patient care.

The original impetus of introducing knowledge discovery in databases of radiation therapy treatment planning is to mine the hidden knowledge in clinical practice and use it as future guidance. It aims to accelerate the clinical workflow and guarantee the consistency of the patient care. By extracting the patient and machine features and implementing the machine learning algorithms to learn the relation in between, we are able to summarize and store the knowledge which would be very difficult or infeasible for the human being to explicitly write down. This approach could accurately make predictions based on previous practice and has shown great potential in many areas in medicine. Esteva et al. developed deep convolutional neural networks (CNNs) to predict and classify the skin cancer type (Esteva et al., 2017). The CNN was

trained using a large dataset (~10^5) of clinical images with disease labels. This study showed very promising result as it achieved similar performance with board-certified dermatologists.

The knowledge learning framework provides the guidance of what is achievable in the clinical setting based on the previous practice. The machine is not yet able to make predictions for novel conditions or diseases. For radiation therapy treatment planning specifically, the prediction given by the machine is similar as the training data fed into the system in terms of the plan quality. Therefore, special caution is recommended when training and using the model. On one hand, the training data imported into the model needs to be meticulously inspected for plan quality assurance. On the other hand, the model prediction serves as the reference for "what-to-expect" for the new patient's anatomy when the radiation therapy plan is generated. The model prediction may not necessarily be the ultimate dosimetric goal for the new patient, since the model's prediction is based on the mean behavior of previous patient population and patient specific decision could be made towards this new patient. The model merely serves as a start point for the planner to generate a clinically acceptable plan to avoid lengthy iterative trial-and-error process, rather than assuming all responsibility of the planner or the clinician. In addition, the knowledge model could serve as an education tool for new planners. The model can guide the learner through the practice process by indicating what dosimetric goal can be achieved for a specific patient anatomy. This application

could potentially impact the way we train next generation of radiation therapy treatment planners.

As indicated in previous Chapters, the proposed knowledge models are very sensitive to the treatment site, clinical conditions and indications, patient population and treatment technology etc. For example, the breast ECOMP planning tool utilizes the specific anatomical features of the breast anatomy, and therefore the tool currently is only applicable to the breast radiation therapy treatment planning. In addition, the random forest model utilizes the training patient cohort from Duke University Medical Center. It is possible that if the patient population changes, the anatomical features may change accordingly. This is especially true for breast cancer patients across the world. Special caution is recommended when the model is transferred to apply on another patient population. In addition to the patient population variations, clinical practice may be different among different clinics. For example, the maximum dose tolerance of the treatment target could be different across institutions. This preference could potentially affect the OAR sparing performance, making the plan quality slightly different from the model's perspective. Therefore, the prediction generated from institution-A model may not be favored at institution-B if some of the dosimetric endpoint tolerances are different. It is necessary to cross compare and validate the plan quality variation and model prediction if the model is about to be implemented to a different clinic.

The knowledge discovery in databases framework of radiation therapy treatment planning is not the terminal pursuit in regard to improving the clinical efficiency and patient care. The knowledge learning concept shows the advantage of fast and accurate prediction although the solution may not necessarily global optimal. The knowledge learning could be combined with the MCO approach to generate the global optimal solution in a fast fashion. The implementation of the knowledge model confines the searching space around the Pareto front. Then a local MCO can be further invoked to identify the right combination of the dosimetric parameters that best fits the clinical preference. This approach starts off with the patient population mean based prediction, followed by patient specific radiation therapy parameter tailoring. A successful implementation of this framework is an important step forward towards precision medicine in radiation therapy.

References

Abernethy, A. P., Etheredge, L. M., Ganz, P. A., Wallace, P., German, R. R., Neti, C., . . . Murphy, S. B. (2010). Rapid-learning system for cancer care. *J Clin Oncol, 28*(27), 4268-4274. doi:10.1200/JCO.2010.28.5478

Aggarwal, C. C., & Yu, P. S. (2001). Outlier detection for high dimensional data. *in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 2001*, 37-46.

Ahunbay, E. E., Peng, C., Chen, G. P., Narayanan, S., Yu, C., Lawton, C., & Li, X. A. (2008). An on-line replanning scheme for interfractional variations. *Med Phys, 35*(8), 3607-3615.

Ahunbay, E. E., Peng, C., Godley, A., Schultz, C., & Li, X. A. (2009). An on-line replanning method for head and neck adaptive radiotherapy. *Med Phys, 36*(10), 4776-4790.

Ahunbay, E. E., Peng, C., Holmes, S., Godley, A., Lawton, C., & Li, X. A. (2010). Online adaptive replanning method for prostate radiotherapy. *Int J Radiat Oncol Biol Phys, 77*(5), 1561-1572. doi:10.1016/j.ijrobp.2009.10.013

Al-Rahbi, Z. S., Ravichandran, R., Binukumar, J. P., Davis, C. A., Satyapal, N., & Al-Mandhari, Z. (2013). A Dosimetric Comparison of Radiotherapy Techniques in the Treatment of Carcinoma of Breast. *Journal of Cancer Therapy, 04*(11), 10-17. doi:10.4236/jct.2013.411A002

Albano, C., Dunn III, W., Edlund, U., Johansson, E., Norden, B., Sjostrom, M., & Wold, S. (1978). Four Levels of Pattern Recognition. *Computer Technique and Optimization, 103*, 429-443.

American Cancer Society. (2015). *Breast Cancer Facts & Figures 2015-2016*. Retrieved from

Angiulli, F., & Pizzuti, C. (2002). Fast outlier detection in high dimensional spaces. *In Proc. PKDD*.

Appenzoller, L., Michalski, J., Thorstad, W., Mutic, S., & Moore, K. (2012). Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys, 39*. doi:10.1118/1.4761864

Arning, A., Agrawal, R., & Raghavan, P. (1996). A linear method for deviation detection in large databases. *In Proc. KDD*.

Barnett, V., & Lewis, T. (1994). *Outlier in Statistical Data* (3rd ed.): John Wiley&Sons.

Bay, S., & Schwabacher, M. (2003). Mining distance-based outliers in near linear time with randomization and a simple pruning rule. *In Proc. KDD*.

Boutilier, J. J., Craig, T., Sharpe, M. B., & Chan, T. C. (2016). Sample size requirements for knowledge-based treatment planning. *Med Phys, 43*(3), 1212. doi:10.1118/1.4941363

Boutilier, J. J., Lee, T., Craig, T., Sharpe, M. B., & Chan, T. C. (2015). Models for predicting objective function weights in prostate cancer IMRT. *Med Phys, 42*(4), 1586-1595. doi:10.1118/1.4914140

Breunig, M. M., Kriegel, H., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. *In Proceedings of the ACM SIGMOD Conference*, 93-104.

Butte, A. J., & Kohane, I. S. (1999). Unsupervised knowledge discovery in medical databases using relevance networks. *Proc AMIA Symp*, 711-715.

Caudell, J. J., De Los Santos, J. F., Keene, K. S., Fiveash, J. B., Wang, W., Carlisle, J. D., & Popple, R. (2007). A dosimetric comparison of electronic compensation, conventional intensity modulated radiotherapy, and tomotherapy in patients with early-stage carcinoma of the left breast. *Int J Radiat Oncol Biol Phys, 68*(5), 1505-1511. doi:10.1016/j.ijrobp.2007.04.026

Censor, Y., Bortfeld, T., Martin, B., & Trofimov, A. (2006). A unified approach for inversion problems in intensity-modulated radiation therapy. *Phys Med Biol, 51*(10), 2353-2365. doi:10.1088/0031-9155/51/10/001

Chankong, V., & Haimes, Y. (1983). Multiobjective Decision Making: Theory and Methodology (Amsterdam, North Holland).

Chu, W. W., Cárdenas, A. F., & Taira, R. K. (1995). KMeD: A knowledge-based multimedia medical distributed database system. *Information Systems, 20*(2), 75-96. doi:http://dx.doi.org/10.1016/0306-4379(95)98556-S

Chui, C. S., Hong, L., Hunt, M., & McCormick, B. (2002). A simplified intensity modulated radiation therapy technique for the breast. *Med Phys, 29*(4), 522-529. doi:10.1118/1.1460875

Craft, D., Halabi, T., Shih, H. A., & Bortfeld, T. (2007). An Approach for Practical Multiobjective IMRT Treatment Planning. *International Journal of Radiation Oncology*Biology*Physics, 69*(5), 1600-1607. doi:http://dx.doi.org/10.1016/j.ijrobp.2007.08.019

David, C., Tarek, H., & Thomas, B. (2005). Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys Med Biol, 50*(24), 5857.

Delaney, A. R., Tol, J. P., Dahele, M., Cuijpers, J., Slotman, B. J., & Verbakel, W. F. (2016). Effect of Dosimetric Outliers on the Performance of a Commercial Knowledge-Based Planning Solution. *Int J Radiat Oncol Biol Phys, 94*(3), 469-477. doi:10.1016/j.ijrobp.2015.11.011

Emmens, D. J., & James, H. V. (2010). Irregular surface compensation for radiotherapy of the breast: correlating depth of the compensation surface with breast size and resultant dose distribution. *Br J Radiol, 83*(986), 159-165. doi:10.1259/bjr/65264916

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. doi:10.1038/nature21056

Etheredge, L. M. (2007). A rapid-learning health system. *Health Aff (Millwood), 26*(2), w107-118. doi:10.1377/hlthaff.26.2.w107

Evans, P. M., Donovan, E. M., Fenton, N., Hansen, V. N., Moore, I., Partridge, M., . . . Yarnold, J. R. (1998). Practical implementation of compensators in breast radiotherapy. *Radiat Oncol, 49*, 255-265.

Fan, H., Zaiane, O. R., Foss, A., & Wu, J. (2006). A nonparametric outlier detection for efficiently discovering top-N outliers from engineering data. *In Proc. PAKDD*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, 17*(3), 37-54.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *Knowledge discovery and data mining: towards a unifying framework*. Paper presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon.

Fayyad, U., & Stolorz, P. (1997). Data mining and KDD: Promise and challenges. *Future Generation Computer Systems, 13*(2), 99-115. doi:http://dx.doi.org/10.1016/S0167-739X(97)00015-0

Fayyad, U. M. (1996). Data mining and knowledge discovery: making sense out of data. *IEEE Expert, 11*(5), 20-25. doi:10.1109/64.539013

Fayyad, U. M. (1997). *Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases*. Paper presented at the Proceedings of the Ninth International Conference on Scientific and Statistical Database Management.

Fiandra, C., Filippi, A. R., Catuzzo, P., Botticella, A., Ciammella, P., Franco, P., . . . Ricardi, U. (2012). Different IMRT solutions vs. 3D-Conformal Radiotherapy in early stage Hodgkin's lymphoma: dosimetric comparison and clinical considerations. *Radiat Oncol, 7*(1), 1-9. doi:10.1186/1748-717x-7-186

Flejmer, A. M., Josefsson, D., Nilsson, M., Stenmarker, M., & Dasu, A. (2014). Clinical Implications of the ISC Technique for Breast Cancer Radiotherapy and Comparison with Clinical Recommendations. *ANTICANCER RESEARCH, 34*, 3563-3568.

Fogliata, A., Belosi, F., Clivio, A., Navarria, P., Nicolini, G., Scorsetti, M., . . . Cozzi, L. (2014). On the pre-clinical validation of a commercial model-based optimisation engine: application to volumetric modulated arc therapy for patients with lung or prostate cancer. *Radiother Oncol, 113*(3), 385-391. doi:10.1016/j.radonc.2014.11.009

Fogliata, A., Wang, P. M., Belosi, F., Clivio, A., Nicolini, G., Vanetti, E., & Cozzi, L. (2014). Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. *Radiat Oncol, 9*. doi:10.1186/s13014-014-0236-0

Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: an overview. *AI Mag., 13*(3), 57-70.

Geng, X., & Smith-Miles, K. (2009). Incremental Learning. In S. Z. Li & A. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 731-735). Boston, MA: Springer US.

Han, J., Cai, Y., & Cercone, N. (1992). *Knowledge Discovery in Databases: An Attribute-Oriented Approach*. Paper presented at the Proceedings of the 18th International Conference on Very Large Data Bases.

Han, J., & Fu, Y. (1994). *Dynamic Generation and Refinement of Concept Hierarchies for Knowledge Discovery in Databases.* Paper presented at the KDD Workshop.

Han, X., Hoogeman, M. S., Levendag, P. C., Hibbard, L. S., Teguh, D. N., Voet, P., . . . Wolf, T. K. (2008). Atlas-based auto-segmentation of head and neck CT images. *Med Image Comput Comput Assist Interv, 11*(Pt 2), 434-441.

Hawkins, D. (1980). Identification of Outliers. *Chapman and Hall, London*.

Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review, 22*(2), 85-126.

Jain, A., Duin, R., & Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 22*(1), 4-37.

Kaufman, L., & Rousseeuw, P. (1987). *Clustering by Means of Medoids*: Fac., Univ.

Khoo, C. S. G., Chan, S., & Niu, Y. (2000). *Extracting causal knowledge from a medical database using graphical patterns*. Paper presented at the Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong.

Khoo, V. S., Bedford, J. L., Webb, S., & Dearnaley, D. P. (2003). Class solutions for conformal external beam prostate radiotherapy. *Int J Radiat Oncol Biol Phys, 55*, 1109-1120.

Knorr, E. M., & Ng, R. T. (1997). A unified approach for mining outliers. *In Proc. CASCON*.

Knorr, E. M., & Ng, R. T. (1998). Algorithms for Mining Distance-Based Outliers in Large Datasets. *In Proceedings of the 24th International Conference on Very Large Data Bases*, 392-403.

Kovalerchuk, B., Vityaev, E., & Ruiz, J. F. (2000). Consistent knowledge discovery in medical diagnosis. *IEEE Eng Med Biol Mag, 19*(4), 26-37.

Krayenbuehl, J., Norton, I., Studer, G., & Guckenberger, M. (2015). Evaluation of an automated knowledge based treatment planning system for head and neck. *Radiat Oncol, 10*, 226. doi:10.1186/s13014-015-0533-2

Langerak, T. R., Berendsen, F. F., Van der Heide, U. A., Kotte, A. N., & Pluim, J. P. (2013). Multiatlas-based segmentation with preregistration atlas selection. *Med Phys, 40*(9), 091701. doi:10.1118/1.4816654

Lessard, E., Kwa, S. L., Pickett, B., Roach, M., 3rd, & Pouliot, J. (2006). Class solution for inversely planned permanent prostate implants to mimic an experienced dosimetrist. *Med Phys, 33*(8), 2773-2782. doi:10.1118/1.2210565

Li, T., Thongphiew, D., Zhu, X., Lee, W. R., Vujaskovic, Z., Yin, F. F., & Wu, Q. J. (2011). Adaptive prostate IGRT combining online re-optimization and re-positioning: a feasibility study. *Phys Med Biol, 56*(5), 1243-1258. doi:S0031-9155(11)73337-6 [pii] 10.1088/0031-9155/56/5/002

Li, T., Zhu, X., Thongphiew, D., Lee, W. R., Vujaskovic, Z., Wu, Q., . . . Wu, Q. J. (2010). On-line adaptive radiation therapy: feasibility and clinical study. *J Oncol, 2010*, 407236. doi:10.1155/2010/407236

Li, Y., & Maguire, L. (2011). Selecting critical patterns based on local geometrical and statistical information. *IEEE Trans Pattern Anal Mach Intell, 33*(6), 1189-1201. doi:10.1109/TPAMI.2010.188

Likhacheva, A., Palmer, M., Du, W., Brown, P. D., & Mahajan, A. (2012). Intensity modulated radiation therapy class solutions in Philips Pinnacle treatment planning for central nervous system malignancies: Standardized, efficient, and effective. *Pract Radiat Oncol, 2*(4), e145-153. doi:10.1016/j.prro.2012.03.010

Lu, C. T., Chen, D., & Kou, Y. (2003, 19-22 Nov. 2003). *Algorithms for spatial outlier detection.* Paper presented at the Third IEEE International Conference on Data Mining.

Mackinnon, M. J., & Glick, N. (1999). Applications: Data Mining and Knowledge Discovery in Databases – An Overview. *Australian & New Zealand Journal of Statistics, 41*(3), 255-275. doi:10.1111/1467-842X.00081

Matheus, C. J., Chan, P. K., & Piatetsky-Shapiro, G. (1993). Systems for Knowledge Discovery in Databases. *IEEE Trans. on Knowl. and Data Eng., 5*(6), 903-913. doi:10.1109/69.250073

Mott, J. H., Livsey, J. E., & Logue, J. P. (2004). Development of a simultaneous boost IMRT class solution for a hypofractionated prostate cancer protocol. *Br J Radiol, 77*(917), 377-386. doi:doi:10.1259/bjr/66104316

Motulsky, H. J., & Brown, R. E. (2006). Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics, 7*, 123. doi:10.1186/1471-2105-7-123

Murray, L. J., Cosgrove, V., Lilley, J., Sykes, J., Thompson, C. M., Franks, K., . . . Henry, A. M. (2014). Developing a class solution for Prostate Stereotactic Ablative Body Radiotherapy (SABR) using Volumetric Modulated Arc Therapy (VMAT). *Radiother Oncol, 110*(2), 298-302. doi:10.1016/j.radonc.2013.10.036

Niemierko, A. (1999). A generalized concept of equivalent uniform dose (EUD) *Medical Physics, 26*, 1100.

Ohsaki, M., Abe, H., Tsumoto, S., Yokoi, H., & Yamaguchi, T. (2007). Evaluation of rule interestingness measures in medical knowledge discovery in databases. *Artificial Intelligence in Medicine, 41*(3), 177-196. doi:http://dx.doi.org/10.1016/j.artmed.2007.07.005

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation, 9*(6), 1-12.

Paddick, I. (2000). A simple scoring ratio to index the conformity of radiosurgical treatment plans. Technical note. *J Neurosurg, 93 Suppl 3*, 219-222. doi:10.3171/jns.2000.93.supplement 3.0219

Piatetsky-Shapiro, G. (2000). Knowledge discovery in databases: 10 years after. *SIGKDD Explor. Newsl., 1*(2), 59-61. doi:10.1145/846183.846197

Piatetsky-Shapiro, G., Matheus, C., Smyth, P., & Uthurusamy, R. (1994). KDD-93: progress and challenges in knowledge discovery in databases. *AI Mag., 15*(3), 77-82.

Prather, J. C., Lobach, D. F., Goodwin, L. K., Hales, J. W., Hage, M. L., & Hammond, W. E. (1997). Medical data mining: knowledge discovery in a clinical data warehouse. *Proc AMIA Annu Fall Symp*, 101-105.

Purdie, T. G., Dinniwell, R. E., Fyles, A., & Sharpe, M. B. (2014). Automation and intensity modulated radiation therapy for individualized high-quality tangent breast treatment plans. *Int J Radiat Oncol Biol Phys, 90*(3), 688-695. doi:10.1016/j.ijrobp.2014.06.056

Purdie, T. G., Dinniwell, R. E., Letourneau, D., Hill, C., & Sharpe, M. B. (2011). Automated planning of tangential breast intensity-modulated radiotherapy using heuristic optimization. *Int J Radiat Oncol Biol Phys, 81*(2), 575-583. doi:10.1016/j.ijrobp.2010.11.016

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of clustere analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65.

Schreibmann, E., & Xing, L. (2004). Feasibility study of beam orientation class-solutions for prostate IMRT. *Med Phys, 31*(10), 2863-2870. doi:10.1118/1.1797571

Slutsky, J. R. (2007). Moving closer to a rapid-learning health care system. *Health Aff (Millwood), 26*(2), w122-124. doi:10.1377/hlthaff.26.2.w122

Soibelman, L., & Kim, H. (2002). Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases. *Journal of Computing in Civil Engineering, 16*(1), 39-48. doi:doi:10.1061/(ASCE)0887-3801(2002)16:1(39)

Stapleford, L. J., Lawson, J. D., Perkins, C., Edelman, S., Davis, L., McDonald, M. W., . . . Fox, T. (2010). Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *Int J Radiat Oncol Biol Phys, 77*(3), 959-966. doi:10.1016/j.ijrobp.2009.09.023

Strassmann, G., Abdellaoui, S., Richter, D., Bekkaoui, F., Haderlein, M., Fokas, E., . . . Engenhart-Cabillic, R. (2010). Atlas-based semiautomatic target volume definition (CTV) for head-and-neck tumors. *Int J Radiat Oncol Biol Phys, 78*(4), 1270-1276. doi:10.1016/j.ijrobp.2010.01.029

Tan, K. C., Yu, Q., Heng, C. M., & Lee, T. H. (2003). Evolutionary computing for knowledge discovery in medical diagnosis. *Artificial Intelligence in Medicine, 27*(2), 129-154. doi:http://dx.doi.org/10.1016/S0933-3657(03)00002-2

Thongphiew, D., Wu, Q. J., Lee, W. R., Chankong, V., Yoo, S., McMahon, R., & Yin, F. F. (2009). Comparison of online IGRT techniques for prostate IMRT treatment: adaptive vs repositioning correction. *Med Phys, 36*(5), 1651-1662.

Tol, J. P., Delaney, A. R., Dahele, M., Slotman, B. J., & Verbakel, W. F. (2015). Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol Biol Phys, 91*(3), 612-620. doi:10.1016/j.ijrobp.2014.11.014

Tsumoto, S. (2000a). Automated discovery of positive and negative knowledge in clinical databases. *IEEE Eng Med Biol Mag, 19*(4), 56-62.

Tsumoto, S. (2000b). Knowledge discovery in clinical databases and evaluation of discovered knowledge in outpatient clinic. *Information Sciences, 124*, 125-137.

Weksberg, D. C., Palmer, M. B., Vu, K. N., Rebueno, N. C., Sharp, H. J., Luo, D., . . . Chang, E. L. (2012). Generalizable class solutions for treatment planning of spinal stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys, 84*(3), 847-853. doi:10.1016/j.ijrobp.2011.12.060

Wille, R. (2002). Why can concept lattices support knowledge discovery in databases? *Journal of Experimental & Theoretical Artificial Intelligence, 14*(2-3), 81-92. doi:10.1080/09528130210164161

Wu, B., Ricchetti, F., Sanguineti, G., Kazhdan, M., Simari, P., Jacques, R., . . . McNutt, T. (2011). Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning. *Int J Radiat Oncol Biol Phys, 79*(4), 1241-1247. doi:10.1016/j.ijrobp.2010.05.026

Wu, Q., Mohan, R., Morris, M., Lauve, A., & Schmidt-Ullrich, R. (2003). Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas. I: dosimetric results. *Int J Radiat Oncol Biol Phys, 56*(2), 573-585.

Wu, Q. J., Thongphiew, D., Wang, Z., Mathayomchan, B., Chankong, V., Yoo, S., . . . Yin, F. F. (2008). On-line re-optimization of prostate IMRT plans for adaptive radiation therapy. *Phys Med Biol, 53*(3), 673-691. doi:10.1088/0031-9155/53/3/011

Xia, P., Lee, N., Liu, Y. M., Poon, I., Weinberg, V., Shin, E., . . . Verhey, L. J. (2004). A study of planning dose constraints for treatment of nasopharyngeal carcinoma using a commercial inverse treatment planning system. *Int J Radiat Oncol Biol Phys, 59*(3), 886-896. doi:10.1016/j.ijrobp.2004.02.040

Xiao, Y., Wang, H., & Xu, W. (2015). Parameter selection of Gaussian kernel for one-class SVM. *IEEE Trans Cybern, 45*(5), 927-939. doi:10.1109/TCYB.2014.2340433

Yang, X., Wu, N., Cheng, G., Zhou, Z., Yu, D. S., Beitler, J. J., . . . Liu, T. (2014). Automated Segmentation of the Parotid Gland Based on Atlas Registration and Machine Learning: A Longitudinal MRI Study in Head-and-Neck Radiation

Therapy. *Int J Radiat Oncol Biol Phys, 90*(5), 1225-1233. doi:10.1016/j.ijrobp.2014.08.350

Yang, Y., Li, T., Yuan, L., Ge, Y., Yin, F., Lee, W. R., & Wu, Q. J. (2015). Quantitative comparison of automatic and manual IMRT optimization for prostate cancer: the benefits of DVH prediction. *JOURNAL OF APPLIED CLINICAL MEDICAL PHYSICS, 16*, 241-250.

Yoon, J. P., & Kerschberg, L. (1993). A Framework for Knowledge Discovery and Evolution in Databases. *IEEE Trans. on Knowl. and Data Eng., 5*(6), 973-979. doi:10.1109/69.250080

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*, 32-35.

Yuan, L., Ge, Y., Lee, W., Yin, F. F., Kirkpatrick, J., & Wu, Q. (2012a). Quantitative analysis of the factors which affect the inter-patient organ-at risk dose sparing variation in IMRT plans. *Med Phys, 39*. doi:10.1118/1.4757927

Yuan, L., Ge, Y., Lee, W. R., Yin, F. F., Kirkpatrick, J. P., & Wu, Q. J. (2012b). Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. *Med Phys, 39*(11), 6868-6878. doi:10.1118/1.4757927

Yuan, L., Wu, Q. J., Yin, F., Li, Y., Sheng, Y., Kelsey, C. R., & Ge, Y. (2015). Standardized beam bouquets for lung IMRT planning. *Phys Med Biol, 60*(5), 1831-1843. doi:10.1088/0031-9155/60/5/1831

Yuan, L., Wu, Q. J., Yin, F. F., Jiang, Y., Yoo, D., & Ge, Y. (2014). Incorporating single-side sparing in models for predicting parotid dose sparing in head and neck IMRT. *Med Phys, 41*(2), 021728. doi:10.1118/1.4862075

Zhu, X., Ge, Y., Li, T., Thongphiew, D., Yin, F. F., & Wu, Q. J. (2011). A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys, 38*(2), 719-726.

Biography

Yang Sheng was born on June 29th, 1990 at Zhenjiang in Jiangsu Province, a city in eastern part of People' Republic of China, 200 miles west of Shanghai. He joined Nanjing University in 2008 to pursue his Bachelor's degree. He graduated in June 2012 and earned his Bachelors of Science degree in physics with several national scholarships. Soon after graduation from the undergraduate institution, he came to Duke University to pursue his M.S. degree in medical physics. In May 2014, he graduated with the degree of Master of Science in medical physics with the Outstanding Academic Award from Duke University Medical Physics Graduate Program. He has then been pursuing his Ph.D. degree at Duke University Medical Physics Graduate Program since August 2014. During five years study at Duke, he published three peer-reviewed articles including two first-author scientific papers: "Atlas-Guided Prostate Intensity Modulated Radiation Therapy (IMRT) Planning" on Physics in Medicine and Biology (PMB), "Exploring the Margin Recipe for Online Adaptive Radiation Therapy for Intermediate-Risk Prostate Cancer: An Intra-fractional Seminal Vesicles Motion Analysis" on International Journal of Radiation Oncology * Biology * Physics (IJROBP). He also had 9 conference abstracts including 4 oral presentations at the annual meeting of American Association of Physicist in Medicine (AAPM) and American Society of Radiation Oncology (ASTRO). He also earned travel award from Southeast AAPM (SEAAPM).