Radiotherapy and Oncology 125 (2017) 392–397

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Decision support system

Clinical decision support of radiotherapy treatment planning: A data-driven machine learning strategy for patient-specific dosimetric decision making

Gilmer Valdes [a,*], Charles B. Simone II [b], Josephine Chen [a], Alexander Lin [c], Sue S. Yom [a,d], Adam J. Pattison [e], Colin M. Carpenter [e], Timothy D. Solberg [a]

[a] Department of Radiation Oncology, University of California, San Francisco; [b] University of Maryland Medical Center, Baltimore; [c] Department of Radiation Oncology, University of Pennsylvania, Philadelphia; [d] Department of Otolaryngology-Head and Neck Surgery, San Francisco; and [e] Siris Medical, Redwood City, United States

ARTICLE INFO

Article history:
Received 31 May 2017
Received in revised form 10 October 2017
Accepted 10 October 2017
Available online 20 November 2017

Keywords:
Decision support system
Knowledge-based planning
Dosimetric tradeoffs

ABSTRACT

Background and purpose: Clinical decision support systems are a growing class of tools with the potential to impact healthcare. This study investigates the construction of a decision support system through which clinicians can efficiently identify which previously approved historical treatment plans are achievable for a new patient to aid in selection of therapy.

Material and methods: Treatment data were collected for early-stage lung and postoperative oropharyngeal cancers treated using photon (lung and head and neck) and proton (head and neck) radiotherapy. Machine-learning classifiers were constructed using patient-specific feature-sets and a library of historical plans. Model accuracy was analyzed using learning curves, and historical treatment plan matching was investigated.

Results: Learning curves demonstrate that for these datasets, approximately 45, 60, and 30 patients are needed for a sufficiently accurate classification model for radiotherapy for early-stage lung, postoperative oropharyngeal photon, and postoperative oropharyngeal proton, respectively. The resulting classification model provides a database of previously approved treatment plans that are achievable for a new patient. An exemplary case, highlighting tradeoffs between the heart and chest wall dose while holding target dose constant in two historical plans is provided.

Conclusions: We report on the first artificial-intelligence based clinical decision support system that connects patients to past discrete treatment plans in radiation oncology and demonstrate for the first time how this tool can enable clinicians to use past decisions to help inform current assessments. Clinicians can be informed of dose tradeoffs between critical structures early in the treatment process, enabling more time spent on finding the optimal course of treatment for individual patients.

The ideal radiotherapy treatment plan should be personalized, delivering a potentially curative tumor dose while minimizing toxicity based on the individual patient's specific anatomy and underlying medical condition. Traditionally, treatment planning decisions are guided by high-quality scientific studies that map quantities of radiation dose, e.g. Dose to 20% of the Lung Volume (V20) or prescribed dose, to the likelihood of tumor control and normal tissue toxicity. While the challenge of dosimetrically-based planning is a solvable computational problem, the underlying clinical challenge lies in understanding the best treatment plan that can be achieved for a specific patient, related to differences in patient anatomy, tradeoffs in the weighting of planning constraints, and conscious and unconscious biases on the part of the prescribing physicians [1]. Moreover, the process of creating a treatment plan requires close communications between practitioners with different areas of expertise in clinical medicine (physician), radiation delivery (physicist), and treatment planning software (dosimetrist).

Clinical decision support systems leverage the history of past decisions by a clinical team and quickly provide reference data informed by past successes at a given clinic or shared between clinics. Combined with contemporary machine learning (also known as artificial intelligence) algorithms and large data stores, these expert systems have begun to impact clinical practice, with examples such as the triage of patients in the Emergency Department [2] or highlighting of calcifications in breast mammography

* Corresponding author at: Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA, United States.
E-mail address: gilmer.valdes@ucsf.edu (G. Valdes).

https://doi.org/10.1016/j.radonc.2017.10.014
0167-8140/

[3]. A key element of these systems is the ability to augment clinicians' knowledge by processing previous decision records to identify those prior decisions and accompanying parameters that are relevant to the current patient. Together with new algorithm development, these systems promise to change the way decisions are made in medicine [4–6]. In radiation oncology, machine learning has been used in applications ranging from quality assurance to patient toxicity but clinical decision support systems (CDS) that empower physicians have not reached widespread use [7–13].

This paper demonstrates a clinical decision support system utilizing machine-learning for patient-specific treatment planning in radiation oncology with the purpose of assisting the radiotherapy planning team in making better treatment plan decisions by leveraging past data. The key differentiation of our system from other "knowledge based solutions" (KBS) to treatment planning is the focus on helping physicians navigate expectations about dosimetric tradeoffs before the treatment planning process. While other existing approaches focus on determining dose–volume histogram (DVH) expectations for individual organs at risk [14–16], our approach allows prospective, expectant navigation of the inherent tradeoffs between those expectations. The CDS described here is aimed at helping physicians and the clinical team to determine the best course of treatment before expending resources on a lengthy treatment planning process, and to better define expectations among the radiation oncology team during the course of plan development. While treatment planning is still required, the knowledge ascertained from a CDS has the potential to guide therapy and decrease the time needed to reach an acceptable plan.

Methods

Data collection

Data from 2009 to 2015 from the University of Pennsylvania Health System was used in this Institutional Review Board-approved retrospective study. The dataset was comprised of 104 consecutive early-stage lung cancer patients treated with stereotactic body radiation therapy (SBRT). Of these data, 81 received a prescription of 5000 cGy in 4 or 5 fractions to the planning target volume (PTV) (peripheral lesions), with the remaining receiving a prescription of 6000 cGy in 8 fractions (central lesions). An additional dataset was comprised of 40 patients with advanced-stage squamous cell carcinoma of the oropharynx who received postoperative proton radiotherapy. Of these data, 38 patients were prescribed between 6000 and 6600 cGy to the PTV (proton Radiobiological Effect Dose, RBE = 1.1). For each oropharynx patient, there also existed a volumetric modulated arc therapy (VMAT) clinical backup photon plan. Patients were identified through a database query (ARIA, Varian Medical Systems, CA). These data were exported, anonymized, and accumulated for processing by commercial software designed for the purposes of this study (QuickMatch, Siris Medical, CA).

Patient treatment plan classification

The goal of the CDS system is to match the current patient to previously treated patients with similar characteristics, such that previously achieved treatment plans and tradeoffs can be explored. This is represented schematically in Fig. 1. Current planning approaches either do not algorithmically use past data (Fig. 1A), or use past data to understand trends from DVH subpopulations (Fig. 1B), primarily as a quality assurance tool. In contrast, plan classification identifies discrete historical plans that can include dose tradeoffs between the target and various organs-at-risk. The requirements for plan classification are an accurate classification algorithm combined with a knowledge database of previous patient treatments. With a sufficiently large database, various achievable results will be proposed by the algorithm such that the clinical team has multiple reference points to use for optimally choosing the appropriate dose trade-offs for a given patient (Fig. 1C and E).

Consider a database of plans, $\boldsymbol{P}$, from which we seek matches to a specific plan p, where matching is defined according to a dissimilarity index of the dosimetric indices between plans. Addressing this problem as a classification problem similar to nearest neighbor classification, we would like to find the plans in $\boldsymbol{P}$ that are closer to p in terms of the dissimilarity index. A probability threshold, τ, is set for the dissimilarity index, and prior treatments that are within the threshold produce treatment plan matches. Formally, the subset $P_p \subseteq P$ of plans in $\boldsymbol{P}$ that matches plan p can be defined as:

$$P_p \subseteq P : \text{dissimilarity}(P_i, p | F_p = f_p, D_p = d_p, F_i = f_i, D_i = d_i) \leqslant \tau \tag{1}$$

where f are the features, d are the dosimetric indices, and the indices "i" denote different plans in the database. For the current patient, F_p and D_p are the features and dosimetric indices. A threshold, τ, is set for classification, and prior treatments that are within the threshold produce treatment plan matches. This probability threshold defines achievability and incorporates known sources of dosimetric variability in planning, including the repeatability of the treatment plan produced by the treatment planning system, v_{TPS}, and the variation in treatment planning preference between clinicians, v_C. The variability in the treatment planning system is found empirically by repeatedly running the treatment planning system for a given set of dosimetric objectives and priorities on exemplary treatment plans. The variability in clinician preference can be learned by calculating the variability in prediction for prediction models built on different subgroups of the dataset, stratified by, for example, treatment planner or physician. This threshold is defined as:

$$T = v_{TPS} + v_C \tag{2}$$

More specifically, the dissimilarity between the new patient plan p and a historical patient plan P_i is determined by calculating the difference between the j dosimetric indices of patient p and patient P_i. A historical patient is a match if,

$$\text{distance}\{d_{p_j}, d_{i_j}\} \leqslant T, \forall j \tag{3}$$

For an historical patient, d_i are the dosimetric indices from the historical plan. For the new patient,

$$d_p = \mathcal{F}(f_p) \tag{4}$$

where f_p are the generated features for the new patient. As noted above, a boosting framework is used to predict the dosimetric indices:

$$\mathcal{F}(f_p) = \sum_j T_j(f_p) \tag{5}$$

where $T(f)$ is a weak learner. Because of the modern machine learning approach that was used (boosting) and our extensive feature-set, the summation in Eq. (5) is over thousands of decision trees that take input from thousands of features. Therefore, a more detailed description of the function represented in Eq. (5) is not practical in this manuscript, and powerful computation is needed to obtain its value. Boosting is a well-known technique in modern machine learning and has been proven to be one of the most accurate, but powerful computing is required to generate the result [19].

Creating the features that account for data variability is a critical aspect of an accurate classifier; this process is often viewed as the most important aspect of a machine learning algorithm [19]. For an accurate feature set, we perform analysis on DICOM images,

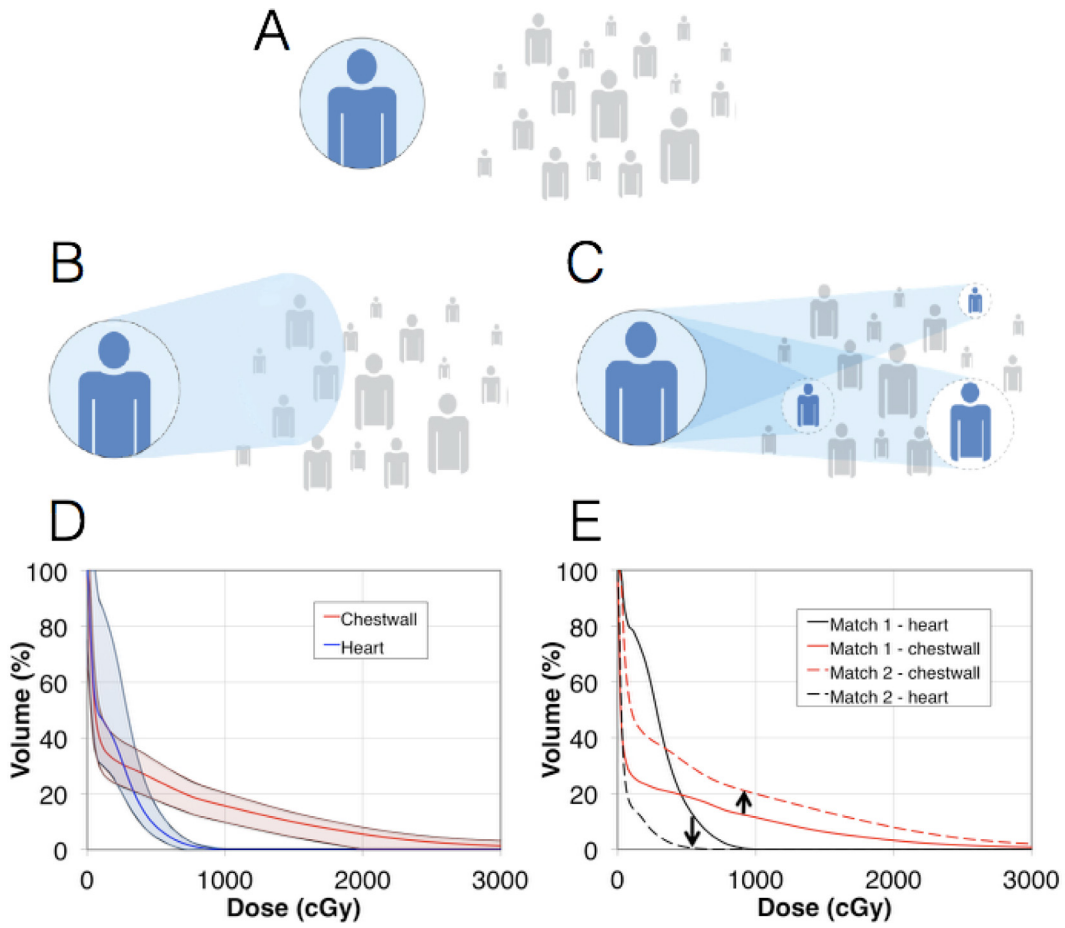

Fig. 1. (A) Standard treatment planning where past patients are not recalled during the planning process. (B) Knowledge-based planning typified by several alternative approaches [17,18]. (C) Treatment plan outcome decision support enabled using treatment plan classification. (D) Dose–volume histogram illustrating the range of values that are acceptable for a certain plan provided by other knowledge-based planning approaches. (E) Dose–volume histogram illustrating distinct tradeoffs for two separate plans that can be delivered using a classification technique.

Table 1
Feature-set categories used to predict dose for a radiation treatment plan.

Feature category	Example features
Anatomical information	Distance, volume, geometric relationship, and importance of structures and surrounding structures
Medical record	ICD-9/10 code, gender, ethnicity
Treatment intent	Fractionation schedule, treatment margin, number of beams/arcs, and the clinicians who are part of the team creating the radiation treatment plan
Radiation transport	Penumbra, aperture, incident angle, beam energy, radiation type (proton vs photon), depth of structure, and existence of bolus

medical record data (e.g. ICD-9/10 diagnosis codes), radiation-transport parameters (e.g. beam energy), and physician treatment intent (e.g. prescription). Several important feature-sets are shown below in Table 1. Additionally, the dosimetric indices for the gross tumor volume (GTV), clinical target volume (CTV), PTV, and organs at risk (OARs) are analyzed for each plan in the knowledge database and aggregated to form the patient response or treatment plan matrix. In this work, the predictive models are built using a boosting framework which internally incorporates feature selection to avoid overfitting [20] and that has been previously used in Radiation Oncology data [9].

Model assessment

An adequately large and heterogeneous database is needed to ensure accuracy in plan classification and to ensure identification of previously achievable treatment plans. To prevent over-fitting during model creation, statistical models were constructed utilizing five-fold cross-validation techniques for each dataset [19].

Learning curves were constructed to analyze the effect of dataset size on dosimetric model accuracy and misclassification error; this technique is a standard method in the machine learning literature for analyzing model error with respect to dataset size [19]. For all dosimetric indices, a decreasing exponential was fit to the data and extrapolated to analyze model error beyond the study dataset size. In general, as the dataset size increases, the model is able to capture the heterogeneity in the data, and thus generate more accurate predictions. For classification, a larger dataset enables more potential past patient treatments for classification.

In order to validate the classification of a treatment plan (to find a match), validation datasets consisting of 10 SBRT lung patients and 7 head and neck (HNC) patients were used. These patient plans were independent of the original data (104 lung plans and 40 head and neck plans) that were used to construct the statistical models,

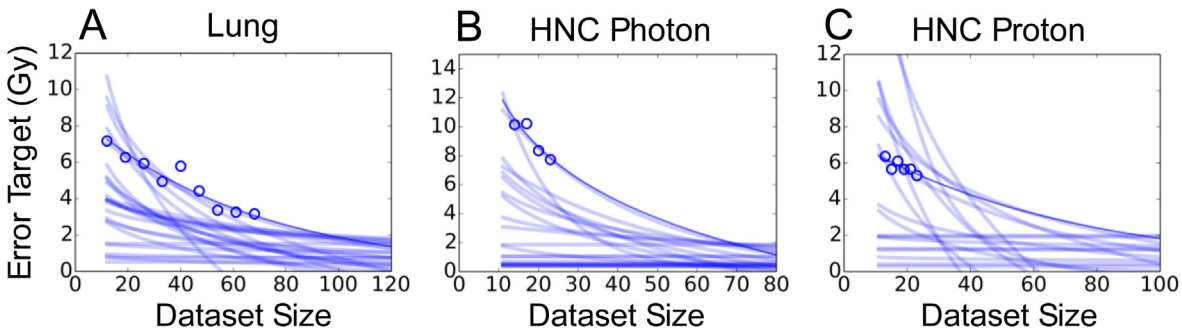

Fig. 2. Learning curves relating the number of data points (patients) needed to achieve an expected error target. Each line represents a different dosimetric index for each patient cohort (shown on the *Y*-axis).

and thus represent the clinical scenario of new patients being assessed by the system. Doses were predicted for each of the validation patients based on the calculated features and the results were compared with the known validation patient doses.

Results

Dosimetric model accuracy and treatment plan misclassification error were investigated using learning curve analyses [19]. Fig. 2 shows the learning curves for (A) lung SBRT, (B) HNC photon-based IMRT, and (C) HNC proton therapy, for all dosimetric indices. The discrete model error for the dosimetric index that required the largest dataset (amount of learning) is overlaid as circles on the appropriate extrapolated curve. For lung patients, this index is the maximum dose to the ipsilateral brachial plexus; for HNC photon it is the mean dose to the contralateral submandibular gland; for HNC proton, it is the mean dose to the ipsilateral parotid. It can be seen that in order for dose prediction error to meet a 2 Gy threshold, chosen as approximately 10% of the maximum dose constraints for the serial structures, a library of 97 patients is needed for lung, 68 patients for HNC photon, and 92 patients for HNC proton.

The validation datasets described above (10 lung patients, 7 HNC patients) were used to build plan classification learning curves to determine the effect of database size on misclassification error; these validation patients were not used in constructing the statistical model. Fig. 3A shows the misclassification error for SBRT lung plans, in which a dataset size of 45 was needed to correctly classify all patients. For HNC photon plans, a dataset size of approximately 60 patients was needed for accurate classification. For HNC proton plans, a dataset size of 30 patients was needed.

The classification algorithm was used to investigate the number of potential treatment plan matches for each patient in the valida-

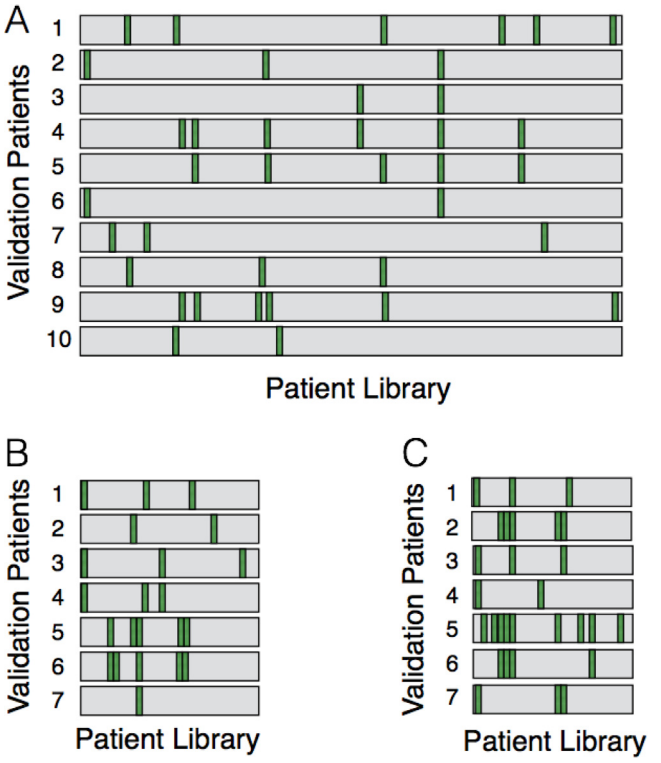

Fig. 4. Treatment classification of the validation patients to the library for (A) lung SBRT, (B) HNC photon-based IMRT, and (C) HNC proton therapy plans. Green lines indicate that the patient in the library (column) was matched to a validation patient. The gray area indicates that the patient in the library was not matched to the validation patient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

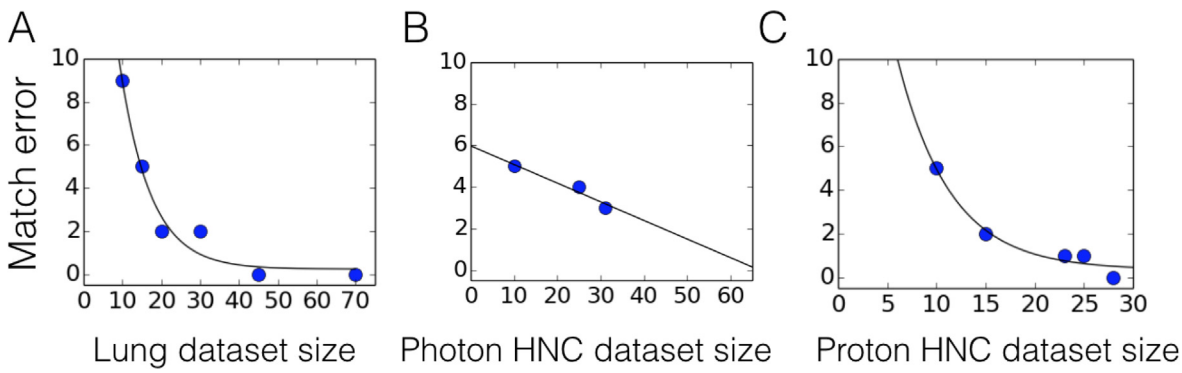

Fig. 3. Learning curves relating misclassification error for different dataset (patient) sizes for (A) lung SBRT, (B) HNC photon-based IMRT, and (C) HNC proton therapy.

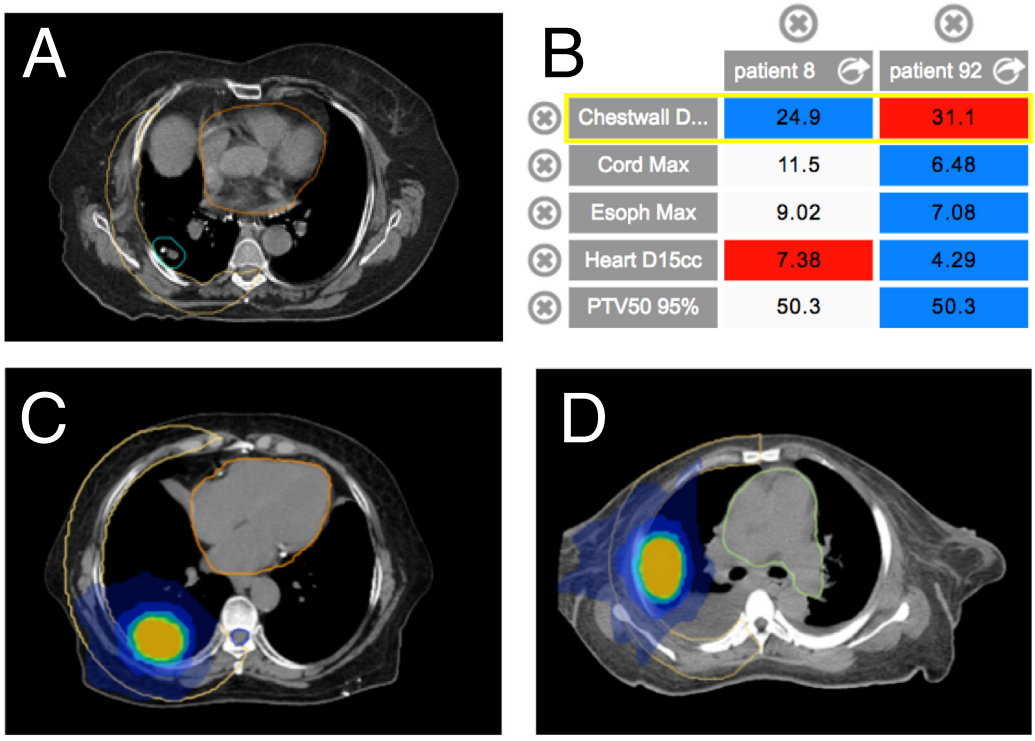

Fig. 5. (A) Validation lung patient evaluated for treatment matches; (B) Possible dosimetric tradeoffs achievable between heart and chest wall for the specific validation patient; (C) One such match (number 8) in which the treatment reduced chest wall dose; (D) Another match (number 92) where the dose to the heart is reduced.

tion sets. The same patient libraries previously used to create the classification models were used as potential matches for the validation set. Fig. 4 shows the patient matches (columns) available for each of the validation patients (rows) for lung SBRT, HNC photon-based IMRT, and HNC proton therapy, respectively. Green lines indicate which library patients matched to the validation patients; gray regions indicate that the corresponding library patients were not a match.

A case of a patient with several applicable treatment dose tradeoffs is shown in Fig. 5. Fig. 5A presents the CT of a patient with stage I non-small cell lung cancer that is matched to the library. The dosimetric indices of two achievable matches are shown in Fig. 5B; values that are above the treatment constraint limit are marked in red. The treatment plans for these matched patients are shown below, specifically patient 8 in Fig. 5C and patient 92 in Fig. 5D. Fig. 5B shows two different potential outcomes in which a variety of outcomes may be explored. In the case of patient 8, the decision was made historically to reduce the chest wall D30cc below 25 Gy. This resulted in increased dose to the heart. Conversely, an alternative treatment plan outcome is shown for patient 92, where the heart D15cc was reduced compared to that of patient 8 (7.38 Gy decreased to 4.29 Gy), but with an increased dose to the chest wall (24.9 Gy increased to 31.1 Gy) and the same PTV coverage (50.3 Gy). It is important to note that although the geometric position of the tumor on the validation set is not exactly the same as that of the patient in panel D, the algorithm effectively identifies the tradeoff between dose to chest wall and dose to the heart for the case under interrogation. This tradeoff is visually translated for the clinician by showing previous cases where the compromises were made. A diagram such as that presented in Fig. 5 can also assist a physician in informing a patient about the inherent tradeoffs of a given clinical decision. This could shift the current clinical decision making process to one in which both physicians and patients may actively visualize the tradeoffs and implications.

Discussion

The power of decision support is the ability to connect a current patient to past decisions [2], which can be recalled from within the same institution or across different institutions. The impact is that decision support tools can enable investigation of prior applicable decisions in an efficient manner by aiding expert opinion and reducing the labor required to reach decision points. With more expertise spent on the clinical decisions themselves, this approach has the potential to improve efficiency, reduce costs, and importantly improve patient outcomes.

Indeed, decision support systems across oncology have proven to be capable of having impact on workflow and quality of care. For example, predictive systems have been used to estimate the risk of breast lesion malignancy [3], predict the outcomes of patients with lung cancer treated with chemoradiation therapy [9,13], predict the survival of patients with skeletal metastases [21] or predict which radiotherapy technique is more appropriate [12,22,23]. Brodin et al. developed a decision support tool to investigate dose–response relationships in Hodgkin's lymphoma, while Cheng et al. have developed decision support systems to evaluate the choice of proton vs photon [24,25]; this tool additionally aided in the optimization of the treatment plan. Others have proposed predictive models of radiotherapy treatment response for genitourinary and thoracic sites [9,13]. Uniquely, the work described herein is the first demonstration of a tool to enable patient-specific dosimetric decision support in a generalized manner across multiple anatomical sites, multiple treatment modalities, and multiple fractionation schemes. In that regard, it is important to highlight the differences between knowledge based solutions (KBS) that to date have been used in radiation oncology and the unique approach presented here. KBS focuses on finding thresholds for error bars and is an excellent tool to evaluate the quality of the planning process [14–16]. In this paper, we demonstrate the novel feasibility of providing matches, which can aid physicians and

patients in decision making prior to the treatment planning process [26].

While the ability to investigate dose trade-offs offers similar functionality to the 'pareto' approach to treatment planning [27], a significant difference is that historical treatment decision support can recall previous treatment decisions, acting as reference to past-approved decisions, and may be applied across institutions.

There are drawbacks to the proposed technique related to database size and quality. While the database only includes clinically-approved plans, there is an inherent variability in the quality of these cases; furthermore, an insufficient database size can restrict the clinician to a match that may be suboptimal. In addition, a reduced number of matches reduces the number of dose trade-offs that can be explored, though these matches can still provide the clinician with information that could benefit decision making. A larger database, especially across multiple clinics, would provide a more expansive group of matches which would capture better quality plans and enable exploration of dosimetric trade-offs in greater resolution.

One limitation of the current study is that some of the learning curves needed to be extrapolated to reach the desired reported error. While this is standard practice in the machine learning literature [19], there is uncertainly in the exact numbers of patient data needed to reach the stated error for such curves.

This study demonstrates the utility of the data-driven machine learning strategy in multiple anatomical sites, fractionation schemes, and treatment modalities. It is, however, limited to the datasets studied herein. Future investigations will assess the impact in additional anatomical sites, fractionation schemes, and treatment modalities, as well as the application to treatment decisions such as which technique or delivery modality is optimal. Further work will also investigate the effect of dataset size to matching; it is anticipated that the impact of this technique will strengthen with larger dataset sizes and across multiple institutions, as the number of matches and thus the degree of clinical insight will increase. Additionally, future work will investigate the impact on enabling remote consultation and remote planning.

Conflict of interest statements

Authors Adam J Pattison PhD and Colin M Carpenter PhD work for Siris Medical and have potential conflict of interest on this publication.

All other authors do not have conflict of interest.

References

[1] Simone 2nd CB et al. Comparison of intensity-modulated radiotherapy, adaptive radiotherapy, proton radiotherapy, and adaptive proton radiotherapy for treatment of locally advanced head and neck cancer. Radiother Oncol 2011;101:376–82.

[2] Garg AX et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. Jama 2005;293:1223–38.

[3] Burnside ES et al. Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings 1. Radiology 2009;251:663–72.

[4] Valdes G et al. MediBoost: a patient stratification tool for interpretable decision making in the era of precision medicine. Sci Rep 2016;6:37854.

[5] Caruana, R., et al. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 201ACM.

[6] Li Q, Chan MF. Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study. Ann N Y Acad Sci 2017;1387:84–94.

[7] Valdes G et al. A mathematical framework for virtual IMRT QA using machine learning. Med Phys 2016;43:4323–34.

[8] Valdes G et al. Use of TrueBeam developer mode for imaging QA. J Appl Clin Med Phys 2015;16:5363.

[9] Valdes G et al. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. Phys Med Biol 2016;61:6105.

[10] El Naqa I et al. Datamining approaches for modeling tumor control probability. Acta Oncol 2010;49:1363–73.

[11] El Naqa I et al. Predicting radiotherapy outcomes using statistical learning techniquesPart of this work was first presented at the Seventh International Conference on Machine Learning and Applications, San Diego, CA, USA, 11–13 December 2008. Phys Med Biol 2009;54:S9.

[12] Lambin P et al. Rapid Learning health care in oncology'–an approach towards decision support systems enabling customised radiotherapy. Radiother Oncol 2013;109:159–64.

[13] Oberije C et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. Radiother Oncol 2014;112:37–43.

[14] Zhu X et al. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys 2011;38:719–26.

[15] Moore KL et al. Experience-based quality control of clinical intensity-modulated radiotherapy planning. Int J Radiat Oncol Biol Phys 2011;81:545–51.

[16] Good D et al. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: an example application to prostate cancer planning. Int J Radiat Oncol Biol Phys 2013;87:176–81.

[17] Tol JP et al. Evaluation of a knowledge-based planning solution for head and neck cancer. Int J Radiat Oncol Biol Phys 2015;91:612–20.

[18] Wu B et al. Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: a head-and-neck case study. Med Phys 2013;40.

[19] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Springer series in statistics. Berlin: Springer; 2001.

[20] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001:1189–232.

[21] Forsberg JA et al. Estimating survival in patients with operable skeletal metastases: an application of a bayesian belief network. PLoS One 2011;6:e19956.

[22] Blanchard P et al. Toward a model-based patient selection strategy for proton therapy: external validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort. Radiother Oncol 2016;121:381–6.

[23] Langendijk JA et al. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.

[24] Brodin NP et al. Interactive decision-support tool for risk-based radiation therapy plan comparison for Hodgkin lymphoma. Int J Radiat Oncol Biol Phys 2014;88:433–45.

[25] Cheng Q et al. Development and evaluation of an online three-level proton vs photon decision support prototype for head and neck cancer–comparison of dose, toxicity and cost-effectiveness. Radiother Oncol 2016;118:281–5.

[26] Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? Phys Med Biol 2008;53:2785.

[27] Thieke C et al. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. Radiother Oncol 2007;85:292–8.