# Machine learning and statistical prediction of patient quality-of-life after prostate radiation therapy

Zhijian Yang [a,b,1], Daniel Olszewski [c,d,2], Chujun He [e], Giulia Pintea [f,g,3], Jun Lian [h], Tom Chou [i], Ronald C. Chen [j], Blerta Shtylla [k,l,4,*]

[a] New York University, New York, NY, 10012, USA
[b] Applied Mathematics and Computational Science Program, University of Pennsylvania, Philadelphia, PA, 19104, USA
[c] Carroll College, Helena, MT, 59625, USA
[d] Computer, Information Science and Engineering Department, University of Florida, Gainesville, FL, 32611, USA
[e] Smith College, Northampton, MA, 01063, USA
[f] Simmons University, Boston, MA, USA
[g] Department of Psychology, Tufts University, Boston, MA, 02111, USA
[h] Department of Radiation Oncology, The University of North Carolina, Chapel Hill, NC, 27599, USA
[i] Depts. of Computational Medicine and Mathematics, UCLA, Los Angeles, CA, 90095-1766, USA
[j] Department of Radiation Oncology, University of Kansas Medical Center, Kansas City, KS, 66160, USA
[k] Department of Mathematics, Pomona College, Claremont, CA, 91711, USA
[l] Early Clinical Development, Pfizer Worldwide Research, Development, and Medical, Pfizer Inc, San Diego, CA, 92121, USA

## ARTICLE INFO

## ABSTRACT

Thanks to advancements in diagnosis and treatment, prostate cancer patients have high long-term survival rates. Currently, an important goal is to preserve quality of life during and after treatment. The relationship between the radiation a patient receives and the subsequent side effects he experiences is complex and difficult to model or predict. Here, we use machine learning algorithms and statistical models to explore the connection between radiation treatment and post-treatment gastro-urinary function. Since only a limited number of patient datasets are currently available, we used image flipping and curvature-based interpolation methods to generate more data to leverage transfer learning. Using interpolated and augmented data, we trained a convolutional autoencoder network to obtain near-optimal starting points for the weights. A convolutional neural network then analyzed the relationship between patient-reported quality-of-life and radiation doses to the bladder and rectum. We also used analysis of variance and logistic regression to explore organ sensitivity to radiation and to develop dosage thresholds for each organ region. Our findings show no statistically significant association between the bladder and quality-of-life scores. However, we found a statistically significant association between the radiation applied to posterior and anterior rectal regions and changes in quality of life. Finally, we estimated radiation therapy dose thresholds for each organ. Our analysis connects machine learning methods with organ sensitivity, thus providing a framework for informing cancer patient care using patient reported quality-of-life metrics.

## 1. Introduction

Approximately 175 thousand new cases of prostate cancer were reported in 2019 in the United States [18]. Depending the patient's age and cancer stage, first-line treatments include prostatectomy, radiation treatment, and androgen ablation. Each of these treatments carries different side effects. For some patients, a prostatectomy is followed by radiation treatment (RT) to minimize the possibility of recurrence. Radiation planning for each patient begins with a CT scan, which is followed by the demarcation of the prostate or prostate bed (radiation

\* Corresponding author. Department of Mathematics, Pomona College, Claremont, CA, 91711, USA.
*E-mail addresses:* shtyllab@pomona.edu, shtyllab@pomona.edu (B. Shtylla).
[1] Current address: Applied Mathematics and Computational Science Program, University of Pennsylvania, Philadelphia, PA, 19104.
[2] Current address: Computer, Information Science and Engineering Department, University of Florida, Gainesville, FL, 32611.
[3] Current address: Department of Psychology, Tufts University, Boston, MA, 02111.
[4] Current address: Early Clinical Development, Pfizer Worldwide Research, Development, and Medical, Pfizer Inc., San Diego, CA, 92121.

target) and the surrounding organs (bladder and rectum) by a physician. Each RT plan is customized to a patient to spatially optimize the dosing to the target organ without overexposing and damaging surrounding organs and structures. RT plans are developed using Dose Volume Histograms (DVHs). DVHs discard all organ-specific spatial information, and they are usually based on a single planning CT scan that does not account for anatomical variations over the course of several weeks of therapy [11]. Various metrics have been developed in order to translate the information from a DVH into a computed probability of uncomplicated tumor control using normal tissue complication probability models (NTCP) [11]. These efforts are necessary as the relationship between exposure to radiation of the surrounding organs/structures and the severity and probability of toxicity (urinary and bowel) is still not fully understood. From a physiological perspective, it is not clear how radiation dosage affects tissues and organ function.

In order to develop a radiation plan that minimizes patient side effects, one needs to quantify these side effects post-radiation. Traditionally, physician-assessed scoring systems have been widely used to measure patient side effects following cancer treatment [3]. However, more recent evidence indicates that clinicians often do not fully ascertain the frequency and severity of patients' treatment-related symptoms [3,14]. Patient-reported quality-of-life (QoL) surveys are becoming an important tool in measuring outcomes after cancer treatment [17,20]. For example, a study by K. Diao et al. [5] explored urinary and bowel symptom development during treatment using patient-reported QoL scores (with 1 indicating no symptoms, to 5 indicating high frequency of symptoms). An IRB waiver was received at the University of North Carolina for this retrospective study using anonymized data. The results showed that average scores progressively increased from baseline throughout treatment, but all symptoms resolved to baseline levels by follow-up. In the context of RT, NTCP models can be correlated to patient-reported QoL data, as was done in Ref. [12]. However, their analysis focused only on urinary symptoms during post-prostatectomy radiotherapy, and NTCP models rely on already reduced DVH information. Given the rich organ-specific information obtained during treatment planning, it could be desirable to more directly connect 3D patient CT scans and dosing with QoL scores.

Machine learning methods have become state of the art in many applications with impressive results; deep convolutional networks have many times outperformed traditional methods for diagnosis [15] or visual recognition tools [9]. In biomedical imaging, deep convolutional neural network (CNN) algorithms have been applied to a wide array of problems. Of particular interest in our context are medical image segmentation machine learning algorithms, such as the U-net [16] architecture that focuses on semantic segmentation of biomedical images. The U-net architecture has appealing features that we will employ in our approach as it uses a large number of max-pooling operations to allow for the identification of global, non-local features and up-convolution to return images to their original size. In the work of Nguyen et al. [13], a modified U-net architecture was shown to accurately predict voxel-level dose distributions for intensity-modulated radiation therapy for prostate cancer patients. These prior studies indicate tremendous promise of guidance of radiation treatment planning with artificial intelligence-based algorithms. Nevertheless, these algorithms can be of significant use in treatment planning if they can also incorporate QoL predictions that can provide immediate guidance for the dosimetrist during clinical plan optimization. To our knowledge, there are no prior approaches that integrate QoL in machine learning algorithms in the context of RT for prostate cancer.

In the department of Radiation Oncology at the University of North Carolina, QoL scores were collected using a validated questionnaire [4, 19] administered during weekly treatment visits as part of the routine clinical work-flow for prostate cancer patients. While QoL data have been studied *after* prostate cancer radiation treatment, data collected *during* the course of treatment can convey important information about symptom development, which can be a fertile ground for the use of

quantitative modeling to guide optimal RT dosing. In this paper, we analyze data from a 14-question quality-of-life prostate cancer patient survey that was collected over the span of five years (2010–2015). The data we examined tracked patient urinary and bowel side effects before and during the course treatment (about seven weeks). Along with each patient's QoL, we also examined the associated anatomical CT scans and radiation dosing patterns for approximately 50 patients.

In this paper, we propose a CNN algorithm to explore the connection between the spatial distribution of the RT dose and the QoL outcomes reported from patients in our dataset. A significant problem with CNN algorithms in our context is the need for a sufficiently large dataset; to resolve this issue, we augmented our patient datasets using interpolation algorithms that generated synthetic patients by combining existing patient data. In addition, we used transfer learning in order to improve the performance of our CNN algorithm and implemented steps to avoid overfitting problem. A key goal for our study was to generate insight into the most radiation-sensitive tissue regions.

As a comparable alternative to the CNN approach, we also used analysis of variance and logistic regression to explore organ sensitivity to radiation and develop dose thresholds for each organ region. We identified regions of the rectum that were highly correlated with changes in individual patient symptoms. Finally, we estimated radiation therapy dose thresholds that could trigger collateral symptoms. Combining results from machine learning and direct analyses of organ sensitivity provides a powerful framework to inform patient care in the quality of life context.

This paper is organized in three sections: In the Methods section, we formulate convolutional neural network algorithms and statistical models. In the Results section, we demonstrate that the CNN algorithm we developed can identify correlations between bowel-related symptoms and radiation. We support these findings through statistical analyses that explore organ sensitivity to radiation dosage. Finally, in the Discussion and Conclusions section, we compare our results with those of previous studies.

## 2. Methods

### 2.1. Quality-of-life data

Our data are extracted from patients' answers to a 14-question quality-of-life survey. Seven questions address urinary symptoms and address bowel symptoms. The specific survey questions and possible responses are shown in Fig. A.12 in Appendix A. Patients took the survey before undergoing radiation therapy, once a week during their seven-week treatment, and after completing therapy. Answers by patient $j$ to question $i$ at time point $t = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $a_{ij}(t)$, were scored on a discrete scale ranging from 1 to 4 or 5 (with 1 indicating the least severity in symptoms and 5 indicating very high severity in symptoms). We used this quality-of-life survey and de-identified CT scans and treatment plans for a total of 52 patients (57 patients were provided; however, five patients had incomplete data and were discarded in our analysis).

The answers $a_{ij}(t = 0)$ to the first survey taken before treatment were used as the baseline of symptoms before radiation therapy. Subsequent answers $a_{ij}(t \geq 1)$ provide information on patient $j$'s symptoms associated with question $i$. In order to reduce the dimensionality of the dataset collected over several weeks, we developed a single score for urinary-related symptoms and bowel-related symptoms.

For each patient, we divided the questions and associated answers into those concerning urinary or bowel function and then identified the worst (maximum) score $a_{ij}^* \equiv \max_t a_{ij}(t)$ a patient reported throughout the multi-week treatment for each question. A single *total difference score* for each patient $j$, $\Delta_j$, was computed as the difference $a_{ij}^* - a_{ij}(0)$, summed over the urinary- or bowel-subset of answers $i$:

$$\Delta_j^{\mathrm{u}} = \sum_{i=1}^{7} \left[ a_{ij}^* - a_{ij}(0) \right], \quad \Delta_j^{\mathrm{b}} = \sum_{i=8}^{14} \left[ a_{ij}^* - a_{ij}(0) \right]. \tag{1}$$

here, we have ordered questions and corresponding answers associated with urinary function as $i = \{1, 2, 3, 4, 5, 6, 7\}$ and the bowel-associated questions as $i = \{8, 9, 10, 11, 12, 13, 14\}$. The total difference scores, $\Delta_j^{\mathrm{u}}$ and $\Delta_j^{\mathrm{b}}$, represent the total change in the quality of life associated with urinary and bowel function, respectively.

Next, we used a cut-off (or threshold) value of 6 to convert the patient's quality-of-life responses into a binary classifier defined by the discrete Heaviside function

$$y_j \equiv H(\Delta_j - 6) = \begin{cases} 0 & \Delta_j < 6, \\ 1 & \Delta_j \geq 6. \end{cases} \tag{2}$$

Score thresholding is visualized in Fig. 1, where both the total changes in the bowel and urinary questionnaire scores are plotted for each patient, overlaid with colored boxes that mark the score cut-off value. A mid-point value of 6 was initially chosen for the threshold; however, we tested other threshold values in order to evaluate the sensitivity of the binary classifier, as detailed in the Results section.

## 2.2. Image augmentation

Previous biomedical studies focusing on image processing have used datasets as small as 85 data points [8] and as large as 100,000 data points [15]. Since the amount of data points we have is even smaller than 85, we decided to enrich our dataset prior to training our algorithms. We used the Fischer-Modersitzki (FM) curvature-based image registration technique [6] in order to generate new *in silico* patients as interpolations of existing patient data. The FM approach was developed in the context of image registration. Although this algorithm has not yet been implemented in open-source toolboxes, we chose it for its mathematical simplicity and efficient implementation.

For completeness, we outline the core ideas of image registration algorithms here. Let $d$-dimensional images be represented by compactly supported mappings $T, R : \Omega \rightarrow \mathbb{R}$ where $\Omega = (0, 1)^d$. Specifically, the quantity $T(\mathbf{x})$ is the intensity or image grey value at the spatial position



**Fig. 1.** Total difference of qualiy-of-life survey scores for patients 1–52. Patients who were classified as having a significant change in symptoms, defined by $\Delta_j^{\mathrm{b}} \geq 6$ or $\Delta_j^{\mathrm{u}} \geq 6$, are shown in red, while those with little symptom change, $\Delta_j^{\mathrm{b}}, \Delta_j^{\mathrm{u}} \leq 5$ are shown in blue. Total change in (a) bowel symptom scores, and (b) urinary symptom scores for each of the analyzed patients.

$\mathbf{x} \in \Omega$. Given a reference image $R(\mathbf{x})$ and a deformable template image $T(\mathbf{x})$, a registration algorithm outputs a deformation, or displacement field, $\mathbf{u} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that when applied to the template image, $T(\mathbf{x} - \mathbf{u}(\mathbf{x}))$ a resulting modified template more closely matches the reference, $R(\mathbf{x})$. The problem is then how to find a desired deformation $\mathbf{u} = (u_1, ..., u_d)$. This becomes an optimization problem of minimizing the difference between the deformed template $T_{\mathbf{u}} := T(\mathbf{x} - \mathbf{u}(\mathbf{x}))$ and the reference $R(\mathbf{x})$.

For any optimization technique, variations in registration methods arise and a metric for measuring the goodness of a deformation must be defined. Let $D$ be the distance measure between the reference $R$ and deformed template $T$, and let $S$ be a measure of the smoothness of the deformation $\mathbf{u}$. The FM approach consists of finding $\mathbf{u}$ by minimizing the joint functional $J[\mathbf{u}]$,

$$J[\mathbf{u}] := D[R, T; \mathbf{u}] + \alpha S[\mathbf{u}]. \tag{3}$$

The regularization parameter $\alpha$ is used to control the strength of the smoothness of the displacement versus the similarity of the images. The difference-squared measure $D$ is given by

$$D[R, T; \mathbf{u}] := \frac{1}{2} \| R - T_{\mathbf{u}} \|^2 = \frac{1}{2} \int_{\Omega} (R(\mathbf{x}) - T(\mathbf{x} - \mathbf{u}(\mathbf{x})))^2 d\mathbf{x}, \tag{4}$$

and the curvature-based smoothness

$$S[\mathbf{u}] := \frac{1}{2} \sum_{j=1}^{d} \int_{\Omega} (\Delta u_j)^2 d\mathbf{x}, \tag{5}$$

with Neumann boundary conditions defined by

$$\nabla u_j(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega, \quad j = 1, ..., d. \tag{6}$$

We find a minimizer $\mathbf{u}$ by first ensuring that the Gâteaux derivative of the objective function vanishes. The resulting *Euler-Lagrange* equations are

$$f(\mathbf{x}, \mathbf{u}(\mathbf{x})) + \alpha A[\mathbf{u}](\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega, \tag{7}$$

with

$$f(\mathbf{x}, \mathbf{u}(\mathbf{x})) = (R - T_{\mathbf{u}}) \cdot \nabla T_{\mathbf{u}}(\mathbf{x})$$
$$= (R(\mathbf{x}) - T(\mathbf{x} - \mathbf{u}(\mathbf{x}))) \cdot \nabla T(\mathbf{x} - \mathbf{u}(\mathbf{x})), \tag{8}$$

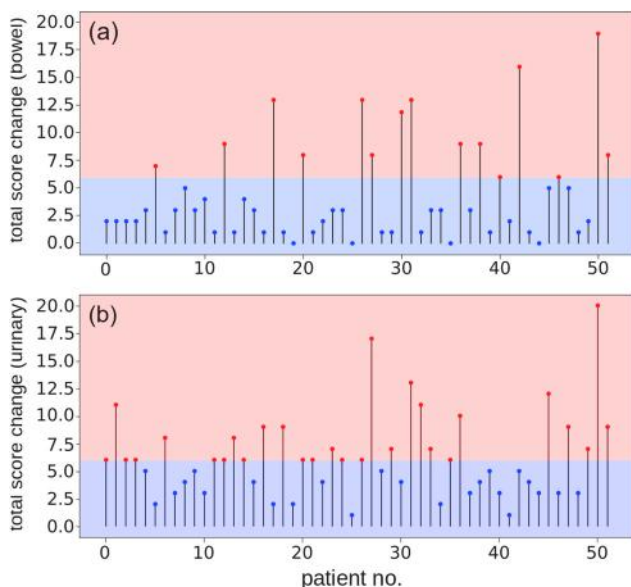$$A[\mathbf{u}](\mathbf{x}) = \Delta^2 \mathbf{u}. \tag{9}$$

The above semi-linear partial differential equations (PDE) are known as the *Navier-Lame* biharmonic and diffusion equations. The Euler-Lagrange equations can be solved using the following fixed-point iteration

$$\alpha A[\mathbf{u}^{k+1}](\mathbf{x}, t) = -f(\mathbf{x}, \mathbf{u}^k(\mathbf{x}, t)), \quad k \geq 0, \tag{10}$$

$$\mathbf{u}^0 = 0. \tag{11}$$

Since the computational domain $\Omega$ is of a simple geometry in this case, a finite difference approximation for the derivatives can be used. This yields a linear system of equations that are solved in each iteration step to obtain the deformation $\mathbf{u}$; more details of the discretization scheme we implemented to solve for $\mathbf{u}$ are given in Ref. [6].

Returning to our problem, the goal is to take each CT slice of patient A and interpolate it with each CT slice of patient B creating a new stack of CT image slices for a new "patient" C. We achieved this interpolation by applying the FM registration approach and selecting one stack of images to serve as the reference and another to serve as the template. We implemented FM in MATLAB, using the discretization approach outlined above and also the implementation outlined in Ref. [10]. A new interpolated patient C was obtained by applying the deformation to the template CT image stacks. We performed this procedure on the 52

patients' cropped CT image stacks for the bladder and the rectum respectively, thus producing 1326 new images for each organ. An example of the interpolated CT images we obtained using this method is shown in Fig. 2.

### 2.3. Radiation plans

Radiation plans were overlaid on a baseline CT scan to give the spatial distribution of applied radiation doses over the course of treatment. When the plan is mapped onto a CT image, high radiation dosage regions are denoted by high pixel intensity (white), while low dosages are represented by darker pixels. A cross-section of a representative radiation plan is shown in Fig. 3. We applied the same FM interpolation used before onto the radiation plan images for each *in silico* patient. Specifically, we interpolate between radiation plans of patient A and patient B to create a new stack of radiation plans for a new "patient" C.

### 2.4. Convolutional neural network model

Next, we constructed separate 3D, three-layer (three-level) convolutional neural networks (CNN) for each organ. Each CNN had two channels as part of its input layer: one to process information related to the patient CT scans and another to process information related to the patient RT plans. The architecture of the model is illustrated in Fig. 4, and more details can be found in Table D2.

The convolutional layers use filters of size $3 \times 3 \times 3$ and strides of 1. Each of the three convolutional layers is followed by max pooling, which reduced the feature size by a factor of two on all three dimensions. We chose to use max pooling for our pooling method so that the filters captured the strongest (and thus highest) pixel value for each stride (we used strides of 1 for each filter and a pooling size of [2,2,2]). In addition, each convolutional layer is followed by an exponential linear unit (ELU) activation function defined as

$$f(x) = \begin{cases} e^x - 1 & x < 0, \\ x & x \geq 0. \end{cases} \tag{12}$$

We used batch normalization (BN) after the convolution and ELU operations, which have been shown to update weights equally throughout the CNN, resulting in faster convergence [13]. In addition, we used drop-out (40% rate) to reduce overfitting since our dataset was small. Because our model is a classifier, we use a cross-entropy loss function that the network minimizes through back-propagation:

$$\text{Loss}(k) = -\frac{1}{N} \sum_{j=1}^{N} y_j \log p_k(y_j) + (1 - y_j) \log(1 - p_k(y_j)), \tag{13}$$

where $k$ is the step number, $y_j = \{0, 1\}$ classifier label of the $j^{\text{th}}$ patient, and $p_k(y_j)$ represents the predicted probability for the corresponding label at the $k^{\text{th}}$ iteration.

Our input images are three-dimensional and of different sizes, $48 \times 48 \times 32$ for the rectum and $34 \times 42 \times 40$ for the bladder. The images used in our study came from patients undergoing radiation after
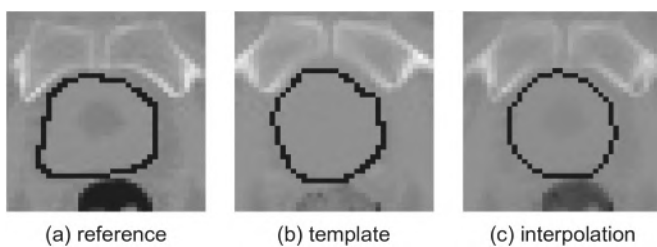


**Fig. 2.** Interpolation of CT images of the bladder using the FM algorithm. The interpolated image (c) is the deformed version of the template image (b) against the reference image (a).



**Fig. 3.** A representative radiation plan. The highest doses correspond to the greatest pixel intensity (in white). Black pixels correspond to no radiation.
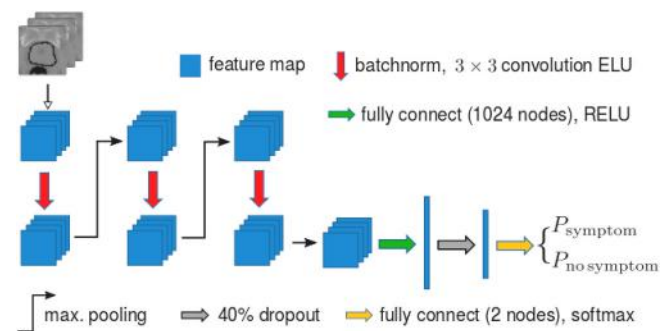


**Fig. 4.** Architecture of the CNN classification model. We used three layers that had convolution, activation, and pooling. The last convolution layer was connected by fully-connected layers and a drop-out layer which dropped out 40% information to avoid over-fitting. The model outputs the probability that a patient will experience urinary or bowel symptoms and probability that a patient will not experience symptoms.

prostatectomy, so no prostates are included in our images or analyses. Both CT scan stacks and RT plans were cropped around the organ of interest (either the bladder or the rectum) with the respective doctor-annotated organ contours for each CT scan. The minimum size was chosen so that the organ of each patients could fit into the cropped images. We used cropped CT and RT images because the information outside of the organs of interest was not deemed useful for the purposes of this study. Cropping also minimized the computational power needed to run our algorithm. The final layer of the model consisted of two nodes: one providing the predicted probability that a patient would manifest a change in (urinary or bowel) symptoms throughout treatment and another giving the predicted probability that a patient would not manifest a change in (urinary or bowel) symptoms throughout treatment. In addition, the CNN produced a confusion matrix (for either the urinary or bowel symptoms) outlining how many patients it accurately predicted from the testing set.

To assess the overall performance of our model, the CNN trained on 39 patients with a batch size of 20 and learning rate of 0.001 for approximately 12 hours. It was then tested against the remaining 13 patients. The CNN was trained on the original patient dataset, not the augmented one through FN interpolation, as only the 52 patient data had QoL information that could be used for model prediction. Patients were shuffled and randomly assigned to the training and testing sets to avoid bias. The CNN also employed a 5-fold cross-validation procedure on the training set, similar to the approach in Jiang et al. [7]. Each of the 5 folds splits the training set into 31 training patients and 8 validation

patients, respectively. Every fold initialized a classifier (for a total of 5 classifiers), from which we could select the model that performed best, based on its accuracy and number of true-positives, and evaluated it on the testing set. The cross-validation model and corresponding loss functions used can be visualized in Fig. 5.

*2.5. Autoencoder*

Due to the small training set and complexity of imaging data, directly training the CNN model resulted in overfitting and poor generalization to the validation set. Therefore, we decided to use transfer learning as a way to improve classification performance. To this end, we employed a convolutional autoencoder network that used a portion of the CNN architecture. Specifically, instead of connecting to a fully-connected layer after all the convolution layers, as we did with the CNN, the autoencoder was used to pre-train the network on unlabeled augmented image data (i.e., data including "synthetic" patients obtained from FM interpolation) by reconstructing the images. The convolutional autoencoder network architecture is illustrated in Fig. 6 and is similar to the U-net architecture used in related segmentation problems [2,8]. More details on autoencoder implementation can be found in Table D3. After the autoencoder was trained, we truncated the network at the start of the deconvolution layers without changing parameters and concatenated it to a fully connected layer that served as the output for our new CNN model.
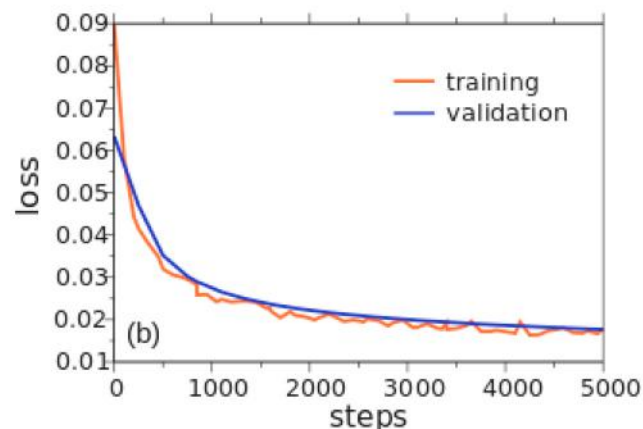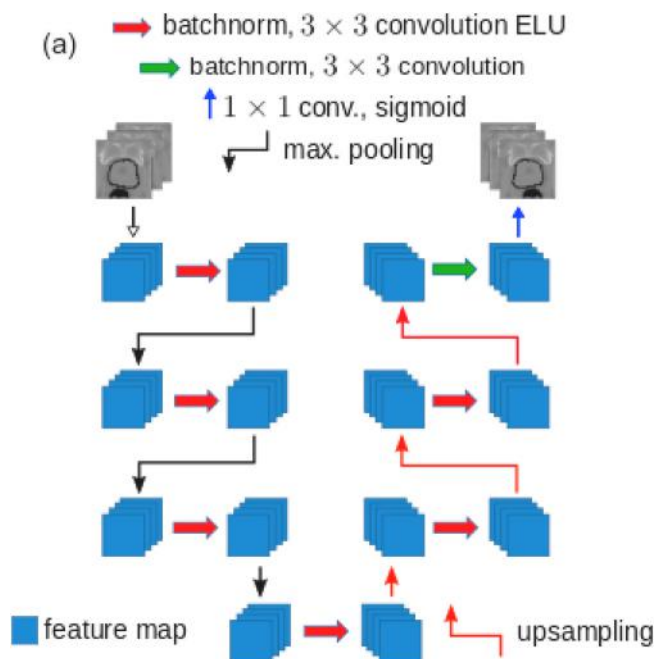




**Fig. 6.** Convolutional Autoencoder. (a) Schematic of the architecture of the convolutional autoencoder network. Three layers have convolution, activation, and pooling. The network deconvolves with activation and pooling for three more layers. The network aims to reconstruct the input image and learn the key features of the CT scans and RT plans. (b) Autoencoder loss while training on the bladder data. The blue curve shows the loss for the validation set; the red curve shows the loss for the training set. Both decrease with the number of iterations of the autoencoder.

Training of the autoencoder allowed us to implement a transfer learning approach where we first trained the autoencoder network to reconstruct augmented patient images. We then assigned the near-optimal weights obtained from this autoencoder training as initial weights for the CNN network with the binary classifier. This pre-training turned out to be necessary in our study due to the small size of the original patient dataset.





**Fig. 5.** (a) Model cross-validation. We initialized 5 different models. For each model, in order to fully sample the data and evaluate model performance, the validation set was selected from a different training subset as illustrated in the diagram. (b) Training (red) and validation (blue) loss for the best model chosen through cross-validation. As shown in red, the training loss decreased significantly, but the validation loss remained around 0.2.

*2.6. Statistical analysis for dose thresholding*

In order to further explore the relationship between QoL scores and RT dosage, we investigated whether there were any correlations between RT dosage in certain organ regions and whether the patients experienced a change in symptoms related to specific organs. Based on correlations, we then computed ranges of RT dosage that could trigger
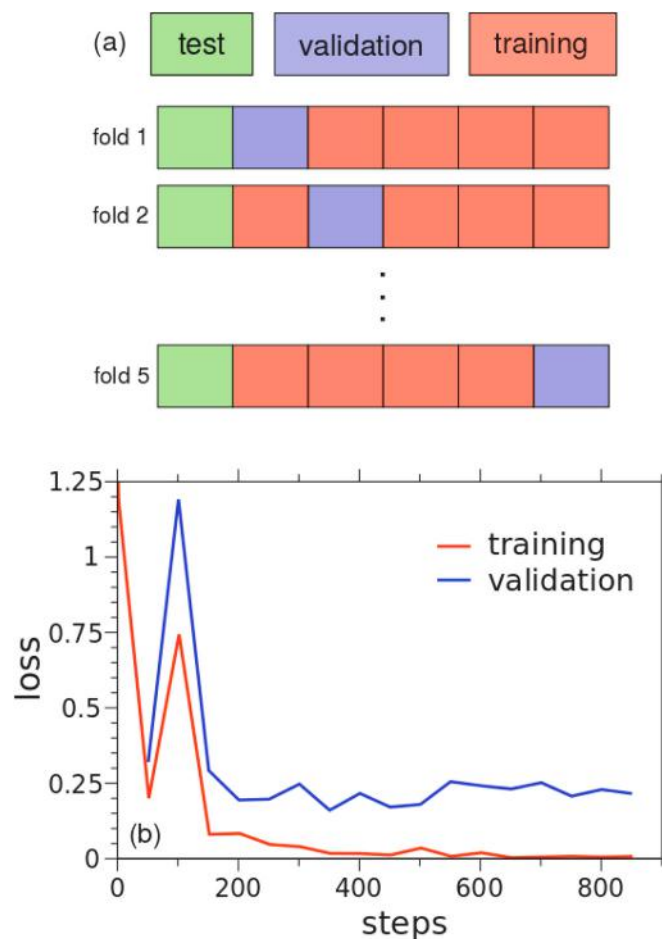
these changes in QoL.

We conducted our analysis on larger (thicker contour) regions surrounding doctor-annotated organ contours for two reasons. First, doctor-annotated organ contours were established before treatment; however, organ shape and position may shift between daily treatments. Moreover, we expanded the region beyond annotated organ boundaries because inflammation can arise from radiation delivered outside the organ. This organ-exterior radiation can impact the organ and corresponding reported symptoms. The region outside the contour was constructed in such way so that the farthest boundary points in the surrounding region fell within a pre-selected distance (pixels) from the doctor-annotated contour (note that our regions included the doctor-annotated contour). To avoid arbitrary selection of the surrounding region, we systematically increased the size of the surrounding organ regions and conducted statistical analysis for each case. This allowed us to obtain more information about the RT dosage located surrounding the organ positions, as originally annotated. Had we analyzed RT dose only on the original organ doctor-annotated contours, we may not have had the most useful information because of the uncertainty in the organs' positions during treatment. Some representative regions with the corresponding RT dosage are shown in Fig. 7.

Next, we defined $d(R_{jk}^a)$ as the average RT dose measured in cGy for each patient $j$ and for each selected region $R_{jk}^a$ with maximum surrounding region thickness of $k$ pixels. We use $a = 1$ to represent regions around the bladder and $a = 2$ to represent regions around the rectum. We examined regions with maximum thickness between $k = 1$ to $k = 8$ pixels.

*Organ sensitivity.* On each region, we investigated whether differences in average RT dosage across different patients had an impact on their QoL scores throughout treatment. To accomplish this, we converted the QoL survey scores into a binary classifier per question as follows

$$y_{ij} \equiv H\left(\left(a_{ij}^* - a_{ij}(0)\right) - 2\right) = \begin{cases} 0 & a_{ij}^* - a_{ij}(0) < 2, \\ 1 & a_{ij}^* - a_{ij}(0) \geq 2. \end{cases} \tag{14}$$

Using this approach, we obtained a binary value $y_{ij}$ for each survey question $i$ and patient $j$ so that we could identify which symptoms were significantly affected by the corresponding RT dosage.

For each selected region with maximum distance $k$, we first compared the mean and range of average doses between the groups of patients who had significant change in all symptoms and those who did not (i.e., $y_j = 1$ vs $y_j = 0$ from Eq. (2)).

This analysis allowed us to understand how the group difference was affected by the thickness of the organ region, $k$. Then, the binary data for each survey question were used in a one-way ANOVA, which compared the distribution of RT dosage $d(R_{jk}^a)$ between the two patient groups ($y_{ij} = 1$ vs $y_{ij} = 0$) for a given symptom question $i$. Recall that using our convention urinary questions, $i = 1 : 7$ correspond to regions $R_{jk}^1$ and bowel-associated questions, $i = 8 : 14$ correspond to regions $R_{jk}^2$. This analysis allowed us to identify which specific symptoms were significantly affected by the corresponding RT dosage for each given region

and what size of surrounding region, $k$, captures any correlations, if any, between symptoms and RT dosage.

*Dosage Thresholding.* Using the most effective size $k$ of the region surrounding an organ (inferred from previous analysis), we next investigated the range of RT dosages that could trigger significant changes in symptoms. To perform this analysis, we first separated organ regions into two parts (or subregions).

To obtain each organ subregion, we combined all CT images for each patient into a cube and separated it into either top and bottom, or front and back regions. Since the bottom of the bladder is closer to the prostate, we speculated this region would be exposed to more radiation and thus be associated with a higher incidence of collateral urinary symptoms. Therefore, we split the bladder into top and bottom regions. Since the front of the rectum is closer to the prostate, we anticipated it would be exposed to more radiation. Thus, we split the rectum into front and back regions. Organ regions are illustrated in Fig. 7(c).

For each organ region and its two corresponding subregions, we identified the lowest RT dose (zero false positive point) for patients with changes in symptoms ($y_j = 1$) and the highest RT dose (zero false negative point) for patients that did not have a change in symptoms ($y_j = 0$) and from which we inferred how high RT dosage had to be in different parts in order to trigger collateral symptoms.

### 2.7. Summary of study workflow

Fig. 8 shows a flowchart that summarizes all the steps we took in this study. Specifically, there were two main branches for our approach. One branch included image analysis and prediction of QoL scores using image augmentation and transfer learning for construction of a predictive CNN algorithm. A second branch used statistical analysis to capture specific organ regions that seemed to correlate most with QoL changes; this statistical analysis could be used to predict organ sensitivity to radiation. Both analysis branches complement each other and enrich our analysis of QoL scores.

## 3. Results and discussion

### 3.1. CNN results

We evaluated the performance of the CNN network using a measure of accuracy defined to be the number of patients with correctly predicted outcomes over the total number of patients. We estimated the accuracy of our results for the bladder and the rectum symptoms by running our algorithm 10 times. We did not find consistent patterns for the bladder, as we obtained a median accuracy of 38% with a range of 23%–53%. This indicated that the CNN model, using the available data, did not find
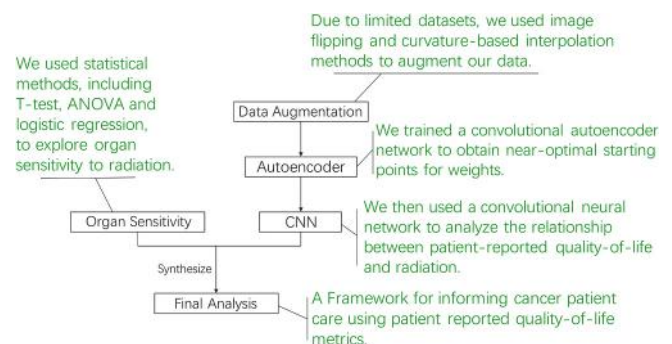
**Fig. 7.** Organ contouring and organ regions. (a) A CT slice of the bladder wall of patient 32 showing his organ contour. (b) Same as (a) but showing the thicker contour ($k = +8$ pixels) for RT-symptom analyses. (c) Lateral view diagram showing how the bladder (yellow) and rectum (red) were split for spatial RT analyses. Image from http://libcat.org/anatomy-of-prostate-cancer.

**Fig. 8.** A flowchart summarizing our analysis of patient images and quality-of-life surveys. Two analysis branches are highlighted. One branch encompasses image analysis through an autoencoder and a CNN for patient QoL predictions. A second statistical analysis branch examines correlations of organ regions with changes in QoL scores as a way to predict organ sensitivity. Both approaches are used complementary for our QoL analysis.

any patterns to classify the patients for bladder symptoms. For the patients with a change in bladder symptoms (denoted by bladder$_{symp}$ in Fig. 9), there was a lot of variability in model predictions ranging anywhere from 0% to 50% with a median of 27%.

In contrast, promising results were obtained for classification of rectal symptoms. Our model with cross-validation accurately predicted a median of 74% changes in rectal symptoms with a range of 62%–84.5%, as illustrated in Fig. 9. For patients with a change in symptoms, (denoted by rectum$_{symp}$ in Fig. 9), the model was accurately predicting the change in symptoms with a median of 56% with a range from 25% to 100%.
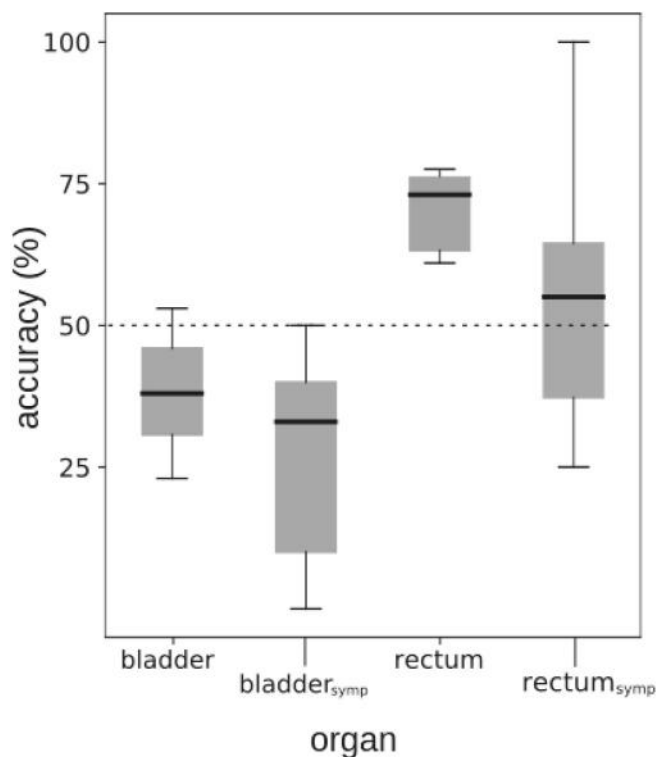
In Table 1, we show the confusion matrix for the rectum model with the validation set that resulted in an 84.6% accuracy. Of the 10 patients without a change in symptoms, 9 of them were accurately predicted. Of the 3 patients with a change in symptoms, 2 were accurately predicted. The result is promising because the model is not always predicting one of the classes; instead, it is picking up some patterns from the patients' data so that it can classify the patients in either category.

We discuss next the steps that we took to ensure that our results were not influenced by our choice of threshold value used to delineate changes in quality of life.

### 3.2. Outcome thresholding

In addition to classifying patients based on their quality-of-life score with a cut-off value of 6, we also used thresholds of 5 and 7 to train the classification model. We found accuracy rates similar to those found using a threshold of 6.

For the data reclassified with a threshold of 5, our model with cross-validation accurately predicted an average of 69% changes in rectal symptoms. For the data reclassified with a threshold of 7, we found an
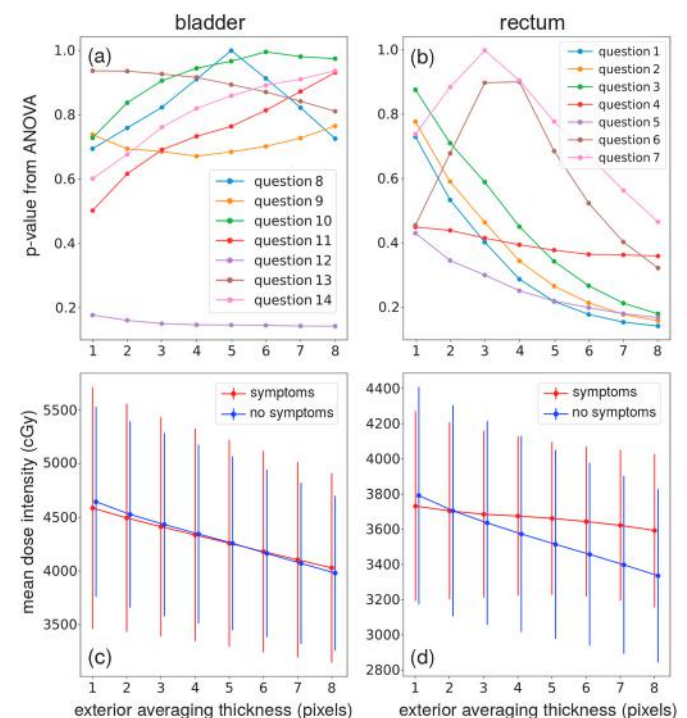
**Table 1**

Confusion matrix for rectum model. Table shows the accuracy for one completely validated model. We show the actual classification of the patient, and what the model predicted.

| Confusion Matrix (Rectum Symptoms) | | |
|---|---|---|
| Actual | Predicted | |
| | No Change | Change |
| No Change | 9 | 1 |
| Change | 1 | 2 |

average accuracy rate of 69%. As there are no considerable differences in results when we change the threshold, reinforcing that our original choice of a cut-off-value, the half of the greatest sum of changes, is a reasonable way to classify patients based on their quality-of-life scores.

### 3.3. Statistical analysis

In Fig. 10(a)-(b) we show the p-values obtained from one-way ANOVA comparing the distribution of average RT dosage in surrounding organ shells (for varying thickness) between patients with or without significant changes in symptoms (for each of the QoL questions). We observed no significant correlations for the bladder. A more interesting signal arises between bowel symptoms and average shell doses where we see that $p-$ values decrease with increasing shell size (or as we include more surrounding regions around the rectum). Fig. 10(c)–(d) show the average doses for patients with and without symptoms, as a function of averaging volume. Patients with symptoms appear to have larger rectum-exterior radiation doses, indicating that bowel function may be sensitive to radiation delivered to tissues outside the rectum. Since the applied RT dosage is defined relative to the anatomy before treatment and the organs may shift between treatment, the result depicted in



**Fig. 9.** Accuracy of our trained classification model. The overall accuracy for the bladder and rectum and the model accuracy within patients who experienced symptoms (denoted with "symp"). For the bladder and bladder symptom accuracies, there were no significant differences, as the median accuracy were below the 50% line. For the rectum, both the overall accuracy for predicting rectum and median rectum symptoms exceed the 50% line.



**Fig. 10.** Correlations between mean doses and symptoms for each organ. (a–b) p-values for each question as a function of the volume taken outside the bladder and rectum "shells" for inclusion in the average radiation dose received, respectively. (c–d) The mean dose as a function of averaging volume divided into patients with and without symptoms. Note that larger averaging volumes outside the rectum are associated with higher average doses and higher average doses are seen to correlate with bowel symptoms.

Fig. 10(d) may also be a consequence of a smaller (decreasing) gradient of dosage for symptomatic patients.

Based on our results thus far we decided to fix the size of the surrounding organ region for analysis to $k = 8$ pixels since this value corresponds to the region thickness with the lowest p values shown in Fig. 10. We next proceed to compute dose thresholds for this fixed surrounding region thickness for each organ. In Fig. 11(a), we assumed zero false positive and zero false negatives to find a "safe" range of mean doses $3081 - 4175$ cGy for the expanded rectum region, implying that a dose greater than ~4200 cGy could induce the development of collateral symptoms. If we further assume the radiation dosages to the different parts of the rectum are uncorrelated, we can also independently find dosage thresholds for the front and back of the expanded rectum regions. We observed that the front of the expanded rectum region could tolerate a higher range of mean dosage (4123–5586 cGy) than the back of the expanded rectum region (901–3009 cGy). These threshold values are listed in Fig. 11(a) and similar results for the top, bottom, and total bladder regions are shown in Fig. C.14 in Appendix B.

While this conclusion is consistent with the published correlations found between the rectal symptoms and rectum dosages in the interval 2500–4200 cGy [1], we cannot rule out that it could result from a possible collinear effect in which patients received high doses to both front and back of the rectum. Since the symptoms cannot be associated with excess radiation to the front or back of the rectum, the region-dependent dosage thresholds are likely a function of the total radiation delivered to all regions. The dosage-symptom instances are shown in a scatter plot in Fig. 11(b). Here, symptoms are typically associated with larger overall dosages. The current patient data are not sufficient to resolve independent region-specific thresholds.

For the bladder, the dose thresholds for the top and bottom of the bladder showed significant overlap: the top of the bladder could tolerate a mean dose of 0–5147 cGy while the bottom could tolerate a mean dose of 3622–6141 cGy (see Appendix B). The corresponding scatter plot shows no discernible correlation between symptoms and sampled doses.

Finally, we computed the correlations between changes in symptoms for the rectum questions and similarly for urinary symptoms, shown in Table D.4. We observed correlations in between questions for rectum and only two questions in bladder. We also computed correlations between sum of changes in rectal and urinary symptoms and observed little correlation there. We also did not obtain correlations between symptom changes and organ fractions that received more than common dose
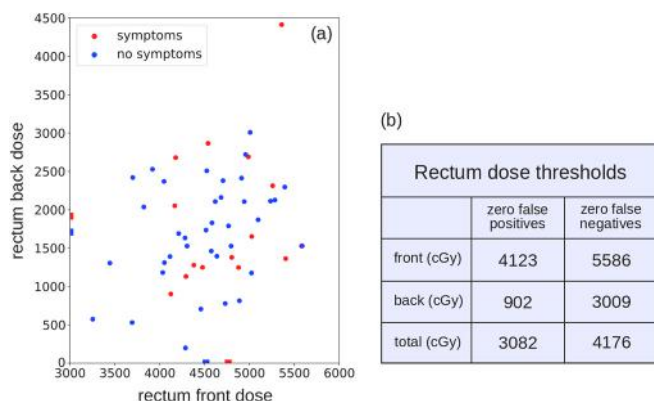
constraints of 40 and 65 Gy, as illustrated by scatterplots of the changes in QoL scores as a function of the fraction of the organ receiving more that the threshold levels in Figure B13. Taken together, these data seem to indicate that bowel and urinary symptoms can be analyzed independently and that there is possibly more consistent reporting of bowel symptoms, or, that there is overlap in the information being covered between bowel symptom questions.

## 4. Summary and conclusions

With the lowering of the prostate cancer mortality rate, an emphasis has been placed on increasing the quality of life for patients undergoing radiation treatment. Utilizing machine learning algorithms and statistical methods, we provide an in-depth analysis on the spatial dosage provided to each patient. By analyzing a patient's anatomical CT image and the radiation therapy dosing, we were able to connect understanding of how radiation influences side-effects. We were able to do this by using a convolutional neural network that analyzed the CT image and associated radiation dosage. Our second method used ANOVA analysis on summarized spatial information. Using a brute-force technique, we were able to identify that splitting the bladder into a top region and bottom and the rectum into a front and back region was the best approach. Our outcomes from ANOVA agreed with our convolutional neural network and also provided dosage thresholds for each region. These results for the dosage thresholds for the rectum and bladder align with the results we obtained from our CNN prediction model, but should be interpreted with care. The thresholds across different regions should not be thought of as independent parameters because the doses applied in the patient samples are correlated and the binary, whole patient symptom indicators are not attributed to any region. Moreover, the number of patients and the range of radiation doses they received are not large enough clearly resolve sharper thresholds. This explains the wide range for the bladder dosage thresholds and the overlap we observed between the top and bottom of the bladder.

On the other hand, our CNN prediction model found that the radiation dose (and the CT scan features) does in fact play a large role in explaining the differences in symptom development across patients. Physiologically, it is possible that radiation dose causing an inflammatory reaction to the area just outside the rectum may actually impact the rectum and patient symptoms. This is an interesting and novel finding, and would need further studies to confirm.

In conclusion, we developed a deep learning framework and complementary statistical methods to identify the connection between spatial dosage and symptoms caused by prostate radiation therapy. A strength of machine learning is that it can produce accurate predictions if presented with sufficiently large datasets; however, the underlying mechanisms or specific features are difficult to discern in these approaches. In our application, it has the potential not only to accurately predict patient side effects, but also to learn what regions of the organs might be responsible for specific side effects. As there is significant interest in integrating machine learning approaches with more traditional modeling approaches, we also found that classical statistical approaches was also useful in our problem. We expect that our CNN results will be much more accurate upon subsequent training on larger patient datasets and can be extended to predicting specific question scores to further refine treatment planning.



**Fig. 11.** Scatter plots, thresholds and corresponding RT dosages for each rectum region. (a) Scatter plot for patients and their corresponding front and back RT doses for rectum. Patients with symptoms and without symptoms are shown in red and blue, respectively. The distribution of sampled RT doses are just broad enough to observe that higher doses lead to symptoms. The mean front and back radiation doses of patients with and without symptoms are indicated by the thick red and blue bars on the *x*- and *y*-axes, respectively. (b) Table of computed thresholds assuming independence. The current data are insufficient to resolve anything other than a total dosage thresholding effect.

### Acknowledgements

## Appendix A. Quality-of-Life Survey

**Rectal Symptoms**
*In the past week have you…*

| | | | | | |
|---|---|---|---|---|---|
| 1. Had diarrhea or loose watery stools? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 2. Had a sense of urgency that you move your bowels? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 3. Had any tenderness or pain when you move your bowels? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 4. Had bleeding with your bowel movements? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 5. Had abdominal cramping or pain? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 6. Passed mucus from your rectum? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 7. Had the urge to move your bowels but had nothing to pass? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |

**Urinary Symptoms**
*In the past week…*

| | | | | | |
|---|---|---|---|---|---|
| 1. How easy has your urine flow been? | Very easy<br><br>1 | Fairly easy<br><br>2 | Slow, but I don't have to strain or bear down<br>3 | Very slow, and I do have to strain or bear down<br>4 | Very slow, and I have to strain or bear down hard<br>5 |
| 2. How often did you urinate at night? | Seldom or never<br>1 | Once a night<br><br>2 | 2 to 3 times a night<br>3 | More than 3 times a night<br>4 | //////// |
| 3. How often did you urinate? | 4 or fewer times a day<br>1 | 5 to 8 times a day<br>2 | 9 to 12 times a day<br>3 | More than 12 times a day<br>4 | //////// |
| 4. How often have you felt pain or burning during urination? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 5. How often did you have the feeling that it is urgent that you pass your urine? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |
| 6. How much control did you have over your urine? | Had complete control (no leaking)<br>1 | Leaked urine, but only at certain times<br>2 | Leaked urine most of the time<br>3 | Little or no control<br><br>4 | //////// |
| 7. How often did you leak urine? | Not at all<br><br>1 | Occasionally (once or twice)<br>2 | Fairly frequently (several times)<br>3 | Frequently (at least once a day)<br>4 | Very frequently (several times a day)<br>5 |

**Fig. A.12.** Quality-of-Life surveys given to patients before and weekly after RT. In our analysis, we assigned the bowel-related questions 1–7, and relabeled the urinary-related questions 8–14.

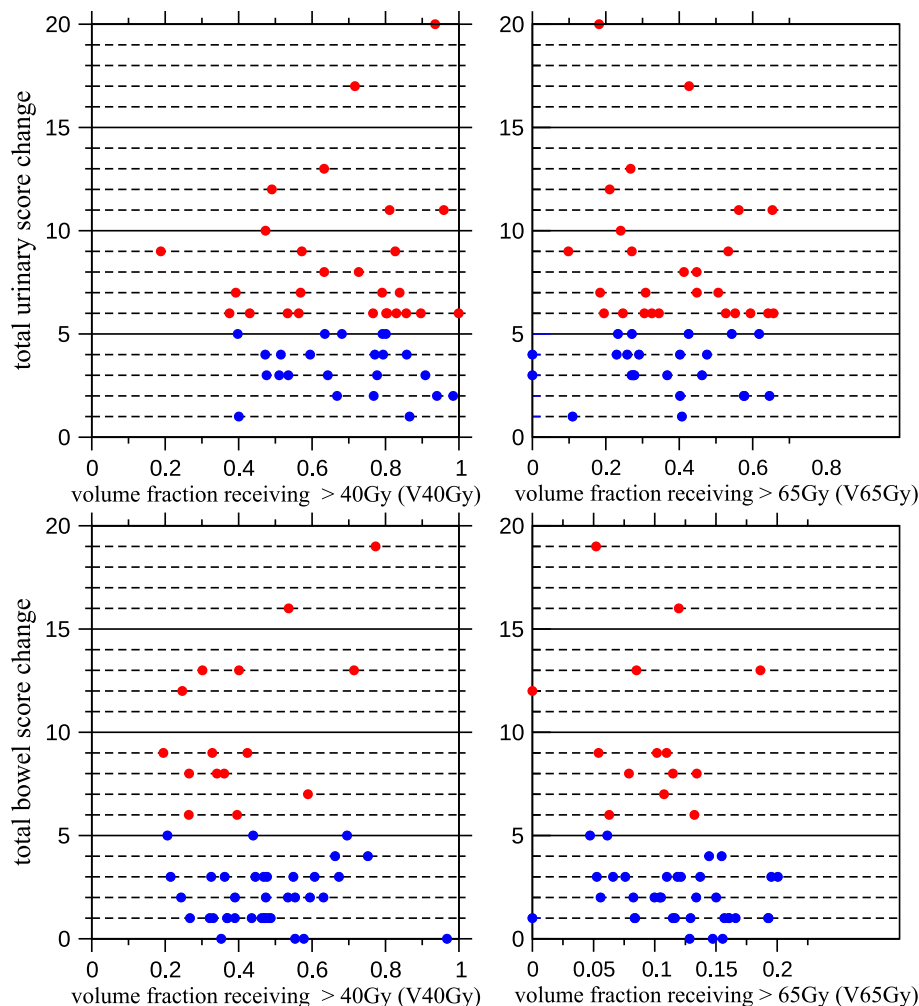**Appendix B. No correlation with V40Gy and V65Gy radiation thresholds**



**Fig. B.13.** Scatter plots of the changes in QoL scores as a function of the fraction of the organ receiving more that the threshold levels 40Gy and 65Gy. There is no apparent correlation with how much of the organ is exposed beyond these threshold levels.

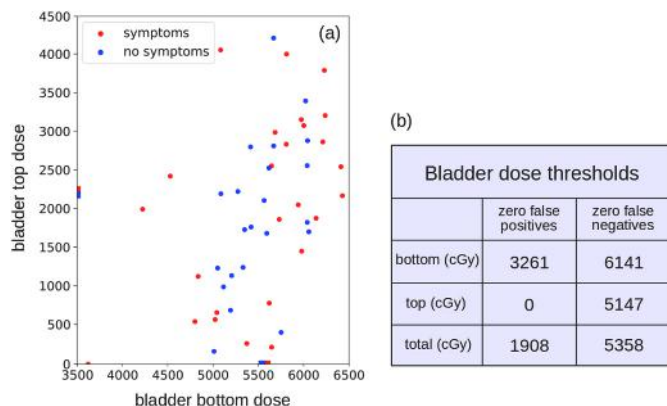**Appendix C. Bladder region thresholds**



**Fig. C.14.** Scatter plots and thresholds of RT doses for each bladder region. (a) Scatter plot of RT doses for the bottom and top of the bladder. The mean front and back radiation doses of patients with and without symptoms are indicated by the thick red and blue bars on the *x*- and *y*-axes, respectively. We see less separation of doses between patients with symptoms and without symptoms for this organ, compared to rectum. (b) Table of independently computed thresholds. These thresholds were estimated from logistic analysis with significant dosage overlap. For the bladder, there is a much smaller range in bottom bladder RT dose and no clear thresholds, i.e., the patients with and without symptoms have significant overlap in their RT dosages.

## Appendix D. Autoencoder and CNN details

**Table D.2**
Parameters of the CNN prediction model.

| Layer | Filter | Number of | Stride | Pooling | Dropout |
|---|---|---|---|---|---|
| | Size | filters | Size | Size | Rate |
| Conv1+Elu | $3 \times 3 \times 3$ | 16 | 1 | 1 | – |
| MaxPool1+BN | $2 \times 2 \times 2$ | – | 2 | 0 | – |
| Conv2+Elu | $3 \times 3 \times 3$ | 32 | 1 | 1 | – |
| MaxPool2+BN | $2 \times 2 \times 2$ | – | 2 | 0 | – |
| Conv3+Elu | $3 \times 3 \times 3$ | 64 | 1 | 1 | – |
| MaxPool3+BN | $2 \times 2 \times 2$ | – | 2 | 0 | – |
| FC + Relu | – | 1024 | – | – | – |
| Dropout | – | – | – | – | 0.4 |
| Softmax | – | – | – | – | – |

*Details of CNN prediction model*

**Table D.3**
Parameters of the autoencoder pre-train model.

| Layer | Filter (Upsampling) | Number of | Stride | Pooling |
|---|---|---|---|---|
| | Size | filters | Size | Size |
| Conv1+Elu | $3 \times 3 \times 3$ | 16 | 1 | 1 |
| MaxPool1+BN | $2 \times 2 \times 2$ | – | 2 | 0 |
| Conv2+Elu | $3 \times 3 \times 3$ | 32 | 1 | 1 |
| MaxPool2+BN | $2 \times 2 \times 2$ | – | 2 | 0 |
| Conv3+Elu | $3 \times 3 \times 3$ | 64 | 1 | 1 |
| MaxPool3+BN | $2 \times 2 \times 2$ | – | 2 | 0 |
| Upsample1 | $2 \times 2 \times 2$ | – | – | – |
| Conv4+Elu | $3 \times 3 \times 3$ | 64 | 1 | 1 |
| Upsample2 | $2 \times 2 \times 2$ | – | – | – |
| Conv5+Elu | $3 \times 3 \times 3$ | 32 | 1 | 1 |
| Upsample3 | $2 \times 2 \times 2$ | – | – | – |
| Conv6+Elu | $3 \times 3 \times 3$ | 16 | 1 | 1 |
| Conv7+Sigmoid | $3 \times 3 \times 3$ | 2 | 1 | 1 |

*Details of autoencoder pre-train model*

**Table D.4**
Spearman correlation between rectal and urinary symptoms and correlation among different types of symptoms within them. For comparison among different types of symtoms, only correlations $> 0.4$ are shown in the table.

| Spearman correlation among symptoms | | | |
|---|---|---|---|
| Change in Rectal Symptoms | | | |
| Q1 vs Q2 | 0.744 | Q2 vs Q4 | 0.458 |
| Q1 vs Q3 | 0.404 | Q2 vs Q5 | 0.457 |
| Q1 vs Q4 | 0.517 | Q3 vs Q7 | 0.441 |
| Q1 vs Q5 | 0.532 | Q6 vs Q7 | 0.431 |
| Q2 vs Q3 | 0.426 | | |
| Change in urinary symptoms | | | |
| Q8 vs Q11 | 0.459 | | |
| Sum of changes in rectal and urinary symptoms | | | |
| Rectal vs Urinary | | 0.338 | |

## References

[1] Massoud Al-Abany, R. Asgeir, Helgason, Anna-Karin Agren Cronqvist, Bengt Lind, Panayiotis Mavroidis, Peter Wersll, Helena Lind, Eva Qvanta, and Gunnar Steineck. Toward a definition of a threshold for harmless doses to the anal-sphincter region and the rectum, Int. J. Radiat. Oncol. Biol. Phys. 61 (4) (mar 2005) 1035–1044.

[2] Anjali Balagopal, Samaneh Kazemifar, Dan Nguyen, Mu-Han Lin, Raquibul Hannan, Amir Owrangi, Steve Jiang, Fully automated organ segmentation in male pelvic CT images, Phys. Med. Biol. 63 (24) (2018) 245015.

[3] Ethan Basch, The missing voice of patients in drug-safety reporting, N. Engl. J. Med. 362 (10) (2010) 865–869.

[4] J.A. Clark, J.A. Talcott, Symptom indexes to assess outcomes of treatment for early prostate cancer, Med. Care 39 (2001) 1118–1130.

[5] Kevin Diao, Emily A. Lobos, Eda Yirmibesoglu, Ram Basak, Laura H. Hendrix, Brittney Barbosa, Seth M. Miller, Kevin A. Pearlstein, Gregg H. Goldin, Andrew Z. Wang, Ronald C. Chen, Patient-reported quality of life during definitive and postprostatectomy image-guided radiation therapy for prostate cancer, Practical Radiation Oncology 7 (2) (2017) e117–e124, 2019/05/20.

[6] Bernd Fischer, Modersitzki Jan, A unified approach to fast image registration and a new curvature based registration technique, Lin. Algebra Appl. 380 (2004) 107–124.

[7] Z. Jiang, J.-F. Witz, P. Lecomte-Grosbras, J. Dequidt, C. Duriez, M. Cosson, S. Cotin, M. Brieu, B-spline based multi-organ detection in magnetic resonance imaging, Strain 51 (3) (2015) 235–247.

[8] Samaneh Kazemifar, Anjali Balagopal, Dan Nguyen, Sarah McGuire, Raquibul Hannan, Steve Jiang, Amir Owrangi, Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning, Biomedical Physics & Engineering Express 4 (5) (jul 2018), 055003.

[9] Alex Krizhevsky, Ilya Sutskever, E. Geoffrey, Hinton. Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[10] Stuart Alexander MacGillivray, Curvature-based Image Registration: Review and Extensions, Master's thesis, University of Waterloo, 2009.

[11] Lawrence B. Marks, Ellen D. Yorke, Andrew Jackson, Randall K. TenHaken, Louis S. Constine, Avraham Eisbruch, Søren M. Bentzen, Jiho Nam, Joseph O. Deasy, Use of normal tissue complication probability models in the clinic, Int. J. Radiat. Oncol. Biol. Phys. 76 (3) (2010) S10–S19, 2019/05/20.

[12] Panayiotis Mavroidis, Kevin A. Pearlstein, John Dooley, Jasmine Sun, Saripalli Srinivas, Shiva K. Das, Andrew Z. Wang, Ronald C. Chen, Fitting NTCP models to bladder doses and acute urinary symptoms during post-prostatectomy radiotherapy, Radiat. Oncol. 13 (1) (2018) 17.

[13] Dan Nguyen, Troy Long, Xun Jia, Weiguo Lu, Xuejun Gu, Zohaib Iqbal, Steve Jiang, A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning, Sci. Rep. 9 (2019) 1076.

[14] Serguei V. Pakhomov, Steven J. Jacobsen, Christopher G. Chute, Veronique L. Roger, Agreement between patient-reported symptoms and their documentation in the medical record, Am. J. Manag. Care 14 (8) (2008) 530–539.

[15] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Langlotz Curtis, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng, Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning, ArXiv (2017) 1–7.

[16] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: convolutional networks for biomedical image segmentation, in: Nassir Navab, Joachim Hornegger, William M. Wells, Alejandro F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.

[17] Jeff A. Sloan, Lawrence Berk, Roscoe Joseph, Michael J. Fisch, Edward G. Shaw, Gwen Wyatt, Gary R. Morrow, Amylou C. Dueck, Integrating patient-reported outcomes into cancer symptom management clinical trials supported by the national cancer institutesponsored clinical trials networks, J. Clin. Oncol. 25 (32) (2007) 5070–5077.

[18] American Cancer Society, American Cancer Society Facts & Figures 2019, 2019 [Online; posted January 8 2019].

[19] James A. Talcott, Jack A. Clark, Judith Manola, Sonya P. Mitchell, Bringing prostate cancer quality of life research back to the bedside: translating numbers into a format that patients can understand, J. Urol. 176 (4) (2006) 1558–1564.

[20] Lynne I. Wagner, Lari Wenzel, Edward Shaw, David Cella, Patient-reported outcomes in phase II cancer clinical trials: lessons learned and future directions, J. Clin. Oncol. 25 (32) (2007) 5058–5062.