Domain knowledge driven 3D dose prediction using moment-based loss function

Gourav Jhanwar¹,³, Navdeep Dahiya²,³, Parmida Ghahremani¹, Masoud Zarepisheh¹,⁴, and Saad Nadeem¹,⁴

¹ Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY, United States of America
² Department of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America
³ Co-first authors.
⁴ Co-senior authors.
* Author to whom any correspondence should be addressed.

E-mail: zarepis@mskcc.org

Keywords: deep learning dose prediction, automated radiotherapy treatment planning, external photon treatment planning

Abstract

Objective. To propose a novel moment-based loss function for predicting 3D dose distribution for the challenging conventional lung intensity modulated radiation therapy plans. The moment-based loss function is convex and differentiable and can easily incorporate clinical dose volume histogram (DVH) domain knowledge in any deep learning (DL) framework without computational overhead. Approach. We used a large dataset of 360 (240 for training, 50 for validation and 70 for testing) conventional lung patients with 2 Gy × 30 fractions to train the DL model using clinically treated plans at our institution. We trained a UNet like convolutional neural network architecture using computed tomography, planning target volume and organ-at-risk contours as input to infer corresponding voxel-wise 3D dose distribution. We evaluated three different loss functions: (1) the popular mean absolute error (MAE) loss, (2) the recently developed MAE + DVH loss, and (3) the proposed MAE + moments loss. The quality of the predictions was compared using different DVH metrics as well as dose-score and DVH-score, recently introduced by the AAPM knowledge-based planning grand challenge. Main results. Model with (MAE + moment) loss function outperformed the model with MAE loss by significantly improving the DVH-score (11%, p < 0.01) while having similar computational cost. It also outperformed the model trained with (MAE + DVH) and (3) the proposed MAE + moments loss. The quality of the predictions was significantly improving the computational cost (48%) and the DVH-score (8%, p < 0.01). Significance. DVH metrics are widely accepted evaluation criteria in the clinic. However, incorporating them into the 3D dose prediction model is challenging due to their non-convexity and non-differentiability. Moments provide a mathematically rigorous and computationally efficient way to incorporate DVH information in any DL architecture. The code, pretrained models, docker container, and Google Colab project along with a sample dataset are available on our DoseRTX GitHub (https://github.com/nadeemlab/DoseRTX)

1. Introduction

Despite recent advances in optimization and treatment planning, intensity modulated radiation therapy (IMRT) treatment planning remains a time-consuming and resource-demanding task with the plan quality heavily dependent on the planner’s experience and expertise (Das et al 2009, Nelms et al 2012, Berry et al 2016). This problem is even more pronounced for challenging clinical cases such as conventional lung with complex geometry and intense conflict between the objectives of irradiating planning target volume (PTV) and sparing organ at risk structures (OARs). Conventional lung plans are generally considered challenging in the clinic due to large PTV sizes, variation in the location of PTV, and having many sensitive nearby structures, making the planning process time-consuming and difficult. Balancing the trade-off between conflicting objectives can lead to sub-optimal plans (Moore et al 2015), sacrificing the plan quality.
In the last decade, several techniques have been developed to automate or facilitate the radiotherapy treatment planning process. Multi-criteria optimization (Craft and Bortfeld 2008) facilitates the planning by generating a set of Pareto optimal plans upfront and allowing the user to navigate among them offline. Hierarchical constrained optimization enforces the critical clinical constraints using hard constraints and improves the other desirable criteria as much as possible by sequentially optimizing these (Breedveld et al 2009, Zarepisheh et al 2019). Knowledge-based planning (KBP) is a data-driven approach to automate the planning process by leveraging a database of pre-existing patients and learning a map between the patient anatomical features and some dose distribution characteristics. The earlier KBP methods used machine learning methods such as linear regression, principal component analysis, random forests, and neural networks to predict DVH as a main metric to characterize the dose distribution (Appenzoller et al 2012, Zarepisheh et al 2014, Fogliata et al 2015, Tol et al 2015, Chin Snyder et al 2016, Valdes et al 2017, Liu et al 2020). However, DVH lacks any spatial information and only predicts dose for the delineated structures.

More recently, deep learning (DL) methods have been successfully used in radiation oncology for automated image contouring/segmentation (Han 2017, Ibragimov and Xing 2017, Men et al 2017) as well as 3D voxel-level dose prediction (Wang et al 2020). A typical DL dose prediction method uses a convolutional neural network (CNN) model which receives a 2D or 3D input in the form of planning CT with OAR/PTV masks and produces a voxel-level dose distribution as its output. The predicted dose is compared to the real dose using some form of loss function such as mean absolute error (MAE) or mean square error (MSE). The loss function in fact quantifies the goodness of the prediction by comparing that to the delivered dose voxel-by-voxel. While MAE and MSE are powerful and easy-to-use loss functions, they fail to integrate any domain specific knowledge about the quality of dose distribution including maximum/mean dose at each structure. The direct representation of DVH results in a discontinuous, non-differentiable, and non-convex function which makes it difficult to integrate it into any DL model. Nguyen et al (2020) proposed a continuous and differentiable, yet non-convex, DVH-based loss function (not to be confused with predicting DVH). In this paper, we propose a differentiable and convex surrogate loss function for DVH using multiple moments of dose distribution. It has been previously shown that moments can approximate a DVH to an arbitrary accuracy (Zinchenko et al 2008), and also moments have been successfully used to replicate a desired DVH (Zarepisheh et al 2013). The convexity and differentiability of the moment-based loss function makes it computationally appealing and also less prone to the local optimality. Furthermore, using different moments for different structures allows the DL model to drive the prediction according to the clinical priorities.

2. Materials and method

2.1. Loss functions

We use three types of loss functions in our study. First, we use MAE that measures the error between paired observations of real and predicted 3D dose. MAE is defined as

\[
L_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^{N} |d_{\text{pred}}^i - d_{\text{ref}}^i|, 
\]

where \( d_{\text{pred}}^i \) is the predicted dose and \( d_{\text{ref}}^i \) is the reference dose for the \( i \)th voxel.

2.1.1. Sigmoid-based DVH loss

Nguyen et al (2020) proposed approximating the heaviside step function by the readily differentiable sigmoid function to address discontinuity and non-differentiability issues of DVH function. Borrowing the notation from Nguyen et al (2020), for a given volumetric dose distribution \( D \) and a segmentation mask \( B_s \) for the \( s \)th structure, the volume-at-dose with respect to the dose \( d_s \), denoted by \( v_{s, d}(D, B_s) \), is defined as the volume fraction of a given region-of-interest (OARs or PTV) which receives a dose of at least \( d_s \) or higher which can be approximated as:

\[
v_{s, d}(D, B_s) = \frac{\sum_i \sigma(\frac{D(i) - d_s}{\beta}) B_s(i)}{\sum_i B_s(i)},
\]

where \( \sigma \) is the sigmoid function, \( \sigma(x) = \frac{1}{1 + e^{-x}} \), \( \beta \) is histogram bin width, and \( i \) loops over the voxel indices of the dose distribution. Based on this, the DVH for the structures \( s \) is defined as
The DVH loss can be calculated using MSE between the real and predicted dose DVH and is defined as follows:

\[ L_{\text{DVH}}(D_r, B_r, D_p, B_p) = \frac{1}{n_t} \sum_{i} ||\text{DVH}(D_r, B_r) - \text{DVH}(D_p, B_p)||^2_2. \]  

Total loss for training the UNET using sigmoid based non-convex DVH loss is then given by

\[ L_{\text{Total}} = L_{\text{MAE}} + w_{\text{DVH}} \cdot L_{\text{DVH}}, \]  

where \( w_{\text{DVH}} \) is the weight for DVH loss (in our experiments we used \( w_{\text{DVH}} = 10 \) based on validation results).

### 2.1.2. Moment loss

Moment loss is based on the idea that a DVH can be well-approximated using a few moments (Zinchenko et al. 2008, Zarepisheh et al. 2013):

\[ \text{DVH} \sim (M_1, M_2, M_3, ..., M_p), \]

where \( M_p \) represents the moment of order \( p \) defined as:

\[ M_p = \left( \frac{1}{|V_s|} \sum_{j \in V_s} d^p_j \right)^{\frac{1}{p}}, \]  

where \( V_s \) is a set of voxels belonging to the structure \( s \), and \( d \) is the dose. \( M_1 \) is simply the mean dose of a structure whereas \( M_{\infty} \) represents the max dose, and for \( p > 1 \), \( M_p \) represents a value between mean and max doses.

In our experiments, we used a combination of three moments \( P = \{1, 2, 10\} \) for the critical OARs and PTV, where \( M_1 \) is exactly the mean dose, \( M_2 \) is the dose somewhere between the mean and max dose and \( M_{10} \) approximates the max dose.

The moment loss is calculated using MSE between the actual and predicted moment for the structure:

\[ L_{\text{Moment}} = \sum_{p \in P} ||M_p - \bar{M}_p||^2_2, \]  

where \( M_p \) and \( \bar{M}_p \) are the \( p \)th moment of the actual dose and the predicted dose of a given structure, respectively.

Total loss for training the neural network using moment based convex loss function is then defined as

\[ L_{\text{Total}} = L_{\text{MAE}} + w_{\text{Moment}} \cdot L_{\text{Moment}}, \]  

where \( w_{\text{Moment}} \) is the weight for the moment loss (in our experiments we used \( w_{\text{Moment}} = 0.01 \) based on validation results).

### 2.2. Patient dataset

We used 360 randomly selected lung cancer patients treated with conventional IMRT with 60 Gy in 30 fractions at Memorial Sloan Kettering Cancer Center between the years 2017 and 2020. All these patients received treatment and therefore included the treated plans which were manually generated by experienced planners using 5–7 coplanar beams and 6 MV energy. Table 1 refers to the clinical criteria used at our institution. All these plans were generated using Eclipse™ V13.7-V15.5 (Varian Medical Systems, Palo Alto, CA, USA).

### 2.3. Inputs and preprocessing

Structure contours and the 3D dose distribution were extracted from the Eclipse V15.5 (Varian Medical Systems, Palo Alto, CA, USA). Each patient has a planning CT and manually delineated contours of PTV and OARs which may differ from patient to patient depending on the location and size of the tumor. However, all patients have esophagus, spinal cord, heart, left lung, right lung and PTV delineated. Hence, we use these five

| Structure     | Max (Gy) | Mean (Gy) | Dose-volume |
|---------------|----------|-----------|-------------|
| PTV           | 72       |           |             |
| Lungs not GTV | 66       | 21        | V (20 Gy) < = 37% |
| Heart         | 66       | 20        | V (30 Gy) < = 50% |
| Stomach       | 54       | 30        |             |
| Esophagus     | 66       | 34        |             |
| Liver         | 66       |           | V (30 Gy) < = 50% |
| Cord          | 50       |           |             |
| Brachial plexus| 65       |           |             |
OARs and PTV as inputs in addition to the planning CT. Figure 1 shows the overall workflow to train a CNN to generate voxel-wise dose distribution. The CT images may have different spatial resolutions but have the same in-plane matrix dimensions of $512 \times 512$. The PTV and OAR segmentation dimensions match those of the corresponding planning CTs. The intensity values of the input CT images are first clipped to have range of $[-1024, 3071]$ and then re-scaled to range $[0, 1]$ for input to the DL network. The OAR segmentations are converted to a one-hot encoding scheme with value of 1 inside each anatomy and 0 outside. The PTV segmentation is then added as an extra channel to the one-hot encoded OAR segmentation.

The dose data have different resolutions than the corresponding CT images. Each pair of the doses is first re-sampled to match the corresponding CT image. For easier training and comparison between different patients, the mean dose inside PTV of all patients is re-scaled to 60 Gy. This serves as a normalization for comparison between patients and can be easily shifted to a different prescription dose by a simple re-scaling inside the PTV region.

Finally, in order to account for the GPU RAM budget, we crop a $300 \times 300 \times 128$ region from all the input matrices (CT/OAR/PTV/Dose) and re-sample it to a consistent $128 \times 128 \times 128$ dimensions. We used the OAR/PTV segmentation masks to guide the cropping to avoid removing any critical regions of interest.

2.4. CNN architecture

Unet is a fully connected network which has been widely used in the medical image segmentation. We train a UNet like CNN architecture (Ronneberger et al 2015, Isola et al 2017) to output the voxel-wise 3D dose prediction corresponding to an input comprising of 3D CT/contours which are concatenated along the channel dimension. The network follows a common encoder–decoder style architecture which is composed of a series of layers which progressively downsample the input (encoder) using max pooling operation, until a bottleneck layer, where the process is reversed (decoder). Additionally, UNet-like skip connections are added between corresponding layers of encoder and decoder. This is done to share low-level information between the encoder and decoder counterparts.

The network (figure 2) uses Convolution-BatchNorm-ReLU-Dropout as a block to perform series of convolution. Dropout is used with a dropout rate of 50%. Maxpool is used to downsample the image by 2 in each
spatial level of encoder. All the convolutions in the encoder are $3 \times 3 \times 3$ 3D spatial filters with a stride of 1 in all 3 directions. In the decoder we use trilinear upsampling followed by regular $2 \times 2 \times 2$ stride 1 convolution. The last layer in the decoder maps its input to a one channel output $(128^3, 1)$.

2.5. Evaluation criteria
To evaluate the quality of the predicted doses, we adopt the metrics used in recent AAPM ‘open-access KBP grand challenge’ (OpenKBP) (Babier et al 2021). This competition was designed to advance fair and consistent comparisons of dose prediction methods for KBP in radiation therapy research. The competition organizers used two separate scores to evaluate dose prediction models: dose score, which evaluates the overall 3D dose distribution and a DVH score, which evaluates a set of DVH metrics. The dose score was simply the MAE between real dose and predicted dose. The DVH score which was chosen as a radiation therapy specific clinical measure of prediction quality involved a set of DVH criteria for each OAR and target PTV. Mean $\bar{D}$ dose received by OAR was used as the DVH criteria for OAR while PTV had three criteria: $D_{99}$, $D_{95}$ and $D_1$ which are the doses received by 99% (1st percentile), 95% (5th percentile), and 1% (99th percentile) of voxels in the target PTV. DVH error, the absolute difference between the DVH criteria for real and predicted dose, was used to evaluate the DVHs. Average of all DVH errors was taken to encapsulate the different DVH criteria into a single score defined as DVH score, measuring the DVH quality of the predicted dose distributions. We also report clinical evaluation criteria such as homogeneity index (HI) (Hodapp 2012) given by $(\frac{D_2 - D_{90}}{D_{90}})$ and paddick conformity index (PCI) (Paddick 2000) which is given by, $PCI = \frac{TV_{PIV}}{TV_{PIV} + \text{PTV}}$, where $TV = \text{target volume}, \text{PIV} = \text{prescription isodose volume and } TV_{\text{PTV}} = \text{volume of the target covered by the prescription isodose}.$

2.6. DL settings
In our experiments, we used Stochastic Gradient Descent (SGD), with a batch size of 1, and Adam optimizer (Kingma and Ba 2015) with an initial learning rate of 0.0002, and momentum parameters $\beta_1 = 0.5, \beta_2 = 0.999$. We trained the network for total of 200 epochs. We used a constant learning rate of 0.0002 for the first 100 epochs and then let the learning rate linearly decay to 0 for the final 100 epochs. When using the MAE and DVH
combined loss, we scaled the DVH component of the loss by a factor of 10. Also, when using the MAE and Moment combined loss, we used the weight of 0.01 for moment loss based upon our validation results.

We divided our training set of 290 images into train/validation set of 240 and 50 images respectively and determined the best learning rate and scaling factor for \((\text{MAE} + \text{DVH})\) loss and \((\text{MAE} + \text{moment})\) loss. Afterwards, we trained all our models using all 290 training datasets and tested on the holdout 70 datasets used for reporting results.

We created the implementations of the CNN model, loss functions and other related training/testing scripts in pytorch and we conducted all our experiments on an NVIDIA A40 GPU with 48 GB VRAM.

3. Results

Figure 3(a) compares the model trained using MAE loss versus \((\text{MAE} + \text{moment})\) loss with respect to different metrics. The y-axis represents the relative improvement obtained using \((\text{MAE} + \text{moment})\) loss, comparing the prediction and ground truth (manual plan). While both models performed similarly with respect to the dose-score and training time (~7 h), the \((\text{MAE} + \text{moment})\) loss improved the DVH score, homogeneity error and conformity error by 11%, 32%, and 21% respectively. For statistical tests, Wilcoxon signed-rank test was performed for different metrics and \(p = 0.05\) was considered as statistical significance. The difference in DVH-score and conformity error was found statistically significant \(p < 0.01\).

Similarly, figure 3(b) compares the model trained using \((\text{MAE} + \text{DVH})\) loss versus \((\text{MAE} + \text{moment})\) loss and shows the relative improvement obtained by \((\text{MAE} + \text{moment})\) loss. \((\text{MAE} + \text{moment})\) significantly improves the training time (almost half), while modestly improves the DVH score, homogeneity and conformity errors (about 7%–8%). The significant improvement in training time for \((\text{MAE} + \text{moment})\) loss owes to its convexity and simplicity. Statistically significant difference \(p < 0.01\) was observed for DVH score.

Figure 4 shows the average absolute error between actual and predicted dose in terms of percentage of prescription for different clinically relevant criteria. Critical OARs like cord and esophagus showed substantial improvement in max and mean absolute dose error respectively using \((\text{MAE} + \text{moment})\) loss compared to other two in the category. PTV D95 and mean dose showed marginal improvements in the dose prediction quality compared to MAE loss. There was small/no-improvement in the MAE for other healthy organs (i.e. left lung, right lung, heart, etc.).
right lung, heart). Figure 5 compares the DVH of an actual dose (ground-truth here) with two predictions obtained from two different loss functions for a patient. As can been seen, in general, the prediction generated with (MAE + moment) loss resembles the actual ground-truth dose more than the other model, for this particular patient, especially for PTV.

Figure 6 shows the comparison of the absolute error for the model trained with default moments \((p = 1, 2, 10)\) for all the structures (red bar) and the model that used different moments for cord \((p = 5, 10)\) and heart \((p = 1, 2)\) and default moments \((p = 1, 2, 10)\) for all other structures (blue bar). As can be seen in the figure, using higher-order moments for cord improves the maximum dose prediction, while using lower-order moments for heart improves the mean dose prediction. Hence, choosing the right set of moments for different structures, one can incorporate different clinically relevant criteria into the model e.g. if max dose for the structure is important, one can use higher-order moments \((p = 5, 10)\) for the structure in loss function and on the other hand, if mean dose is important one can use lower-order moment \((p = 1, 2)\).

Figure 5. DVH plots for different structures using (i) actual dose, predicted dose using (ii) MAE + DVH loss, and (iii) MAE + moment loss, for one patient.

4. Discussion

In this study, we have employed moments as a surrogate loss function to integrate DVH into DL (DL) 3D dose prediction. Moments provide a mathematically rigorous and computationally efficient way to incorporate DVH information in any DL architecture without any computational overhead. This allows us to incorporate the domain specific knowledge and clinical priorities into the DL model. Using MAE + moment loss means the DL model tries to match the actual dose (ground-truth) not only at a micro-level (voxel-by-voxel using MAE loss) but also at a macro-level (structure-by-structure using representative moments).

Moments are essentially simple polynomial functions which can be calculated efficiently. Given their convexity, they do not suffer from the local optimality issue, making them more reliable choices with more robust behaviour against the stochastic nature of the optimization techniques, commonly used in DL models. The computational efficiency of the moments allows training of large DL models. They also offer better fine-tuning of the hyper parameters.
The moments in conjunction with MAE help to incorporate DVH information into the DL model, however, the MAE loss still plays the central role in the prediction. In particular, the moments lack any spatial information about the dose distribution which is provided by the MAE loss. The MAE loss has also been successfully used across many applications and its performance is well-understood. Further research is needed to investigate the performance of the moment loss on more data especially with different disease sites.

The 3D dose prediction can facilitate and accelerate the treatment planning process by providing a reference plan which can be fed into a treatment planning optimization framework to be converted into a deliverable Pareto optimal plan. The dose-mimicking approach has been commonly used in the literature Fan et al (2019), seeking the closest deliverable plan to the reference plan using quadratic function as a measure of distance. Babier et al (2020) proposed an inverse optimization framework which estimates the objective weights from the reference plan and then generates the deliverable plan by solving the corresponding optimization problem. Any improvements in the prediction, including the ones we have obtained using our proposed moment loss, needs to be ultimately evaluated using the entire pipeline of predicting a plan and converting that into a deliverable plan.

5. Conclusion

This work shows that the moments are powerful tools with sound mathematical properties to integrate DVH as an important domain knowledge into 3D dose prediction without any computational overhead. The idea has been validated on a large dataset of 360 conventional lung patients. The conventional lung patients are usually considered challenging cases in the clinic due to their large PTV sizes and sensitive nearby critical structures (e.g. esophagus, heart).

Acknowledgments

This project was partially supported by MSK Cancer Center Support Grant/Core Grant (P30 CA008748).
References

Appenzoller J M, Michalski J M, Thorstad W L, Mutic S and Moore K L 2012 Predicting dose-volume histograms for organs-at-risk in IMRT planning Med. Phys. 39 7446–61

Babier A, Mahmood R, McNiven A L, Diamant A and Chan T C 2020 The importance of evaluating the complete automated knowledge-based planning pipeline Phys. Med. 72 73–9

Babier A, Zhang B, Mahmood R, Moore K L, Purdie T G, McNiven A L and Chan T C 2021 OpenKBP: the open-access knowledge-based planning grand challenge and dataset Med. Phys. 48 5549–61

Berry S L, Bozcokowski A, Ma R, Mechelakos J and Hunt M 2016 Interobserver variability in radiation therapy plan output: results of a single-institution study Pract. Radiat. Oncol. 6 442–9

Breedveld S, Storchi P R and Heijmen B J 2009 The equivalence of multi-criteria methods for radiotherapy plan optimization Phys. Med. Biol. 54 7199–209

Chin Snyder K, Kim J, Reding A, Fraser C, Gordon J, Ajlouni M, Movsas B and Chetty I 2016 Development and evaluation of a clinical model for lung cancer patients using stereotactic body radiotherapy (SBRT) within a knowledge-based algorithm for treatment planning J. Appl. Clin. Med. Phys. 17 263–75

Craft D and Bortfeld T 2008 How many plans are needed in an IMRT multi-objective plan database? Phys. Med. Biol. 53 2785–96

Das I J, Moskvin V and Johnstone P A 2009 Analysis of treatment planning time among systems and planners for intensity-modulated radiation therapy J. Am. Coll. Radiol. 6 514–7

Fan J, Wang J, Chen Z, Hu C, Zhang Z and Hu W 2019 Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique Med. Phys. 46 570–81

Fogliata A et al 2015 Performance of a knowledge-based model for optimization of volumetric modulated arc therapy plans for single and bilateral breast irradiation PLoS One 10 10.1371/journal.pone.0145137

Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method Med. Phys. 44 1408–19

Hodapp N 2012 The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft... [et al] 188 97–9

Ibragimov B and Xing L 2017 Segmentation of organs-at-risk in head and neck CT images using convolutional neural networks Med. Phys. 44 547–57

Isola P, Zhu J-Y, Zhou T and Efros A A 2017 Image-to-Image Translation with Conditional Adversarial Networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1125–1134 (https://openaccess.thecvf.com/content_cvpr_2017/html/Isola IMAGE-TO-IMAGE_TRANSLATION_with_CVPR_2017_paper.html)

Kingma D P and Ba J 2014 Adam: a method for stochastic optimization 10.48550 arkiv:/87299-321

Liu Z, Chen X, Men K, Yi J and Dai J 2020 A deep learning model to predict dose-volume histograms of organs at risk in radiotherapy treatment plans Med. Phys. 47 5467–81

Men K, Dai J and Li Y 2017 Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks Med. Phys. 44 6377–89

Moore K L et al 2015 Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126 Int. J. Radiat. Oncol., Biol. Phys. 92 226–35

Nelms B E, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, Wheeler J and Sobczak M L 2012 Variation in external beam treatment planning: the expedited constrained hierarchical optimization (ECHO) system Med. Phys. 39 2944–54

Nemoto H, Hashimoto S, Sato S, Diez T, Tojo T, Higuchi Y, Renshaw AA, Dogan A, Gatenby RA, Chappuis S, Leibel SA, Arnesen T, Okuno Y, Oka T, Kato N, Arai K, Minami K, Matsui T, Kato K, Yatabe Y, Nishiyama T, Sugawara T, Vargas-Guzman C, Moats A, Horgan OJ, Kuhlman J, Gevaert K, Desai J, van der Graaf Y, Douglas-Denton R, Jordan SC, Ewald F, Kopp J, Rogers J, de Ridder D, Bortfeld T, Yura A, Mercier S, Dabbeke M, Chmait R, Yoon HW, Stiglic G, Togashi K, Sen K, Seto H, Weichselbaum RR, Hasegawa M, Kaye FJ, Leszczynski JA, Verheyen J, Midena E, Davis-Anderson C, Ouyang H, Togashi K, Eberhard MW, Deasy JO, Dessy E, Meinzer HP, McLellan SL, Krag D, Zabar RA, Pajitnov A, Saad H, Kralik D, Fenton M, Talari S, Kvarda B, Studer M, Fowlkes JR, LeBauer JA, Hui Y, Jin C, Bajpai S, Desai D, Guckenberger M, Paliwal B, Badlani G, Chaddad K, Zaidi M, Heemskerk MM, Zwart JJ, Chin AY, Schild SE, Bajwa J, Kilpatrick JL, Jiang S, Hnatowich D, Duan X, Cai X, Narayan S, Liu X, Ito T, Zhang X, Tandan R, Bortfeld T, Artzler M, Xi W, Hamilton P, Prince MA, Togashi K, Duda R, Iyengar A, Ramanathan R, Krajewska D, McDonald RJ, Schiller G, Weichselbaum RR, Minami Y, Byrnes K and Cunat J 2019 Variation in external beam treatment planning and adaptive radiotherapy replanning Med. Phys. 41 061711

Paddick I 2000 A simple scoring ratio to index the conformity of radiosurgical treatment plans J. Neurosurg. 93 219–22

Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation Medical Image Computing and Computer-Assisted Intervention – MICCAI (https://doi.org/10.1007/978-3-319-24574-4_28)

Tol J P, Delaney A R, Dahele M, Slotman B J and Verbakel W F 2015 Evaluation of a knowledge-based planning solution for head and neck cancer Int. J. Radiat. Oncol., Biol. Phys. 91 612–20

Valdes G, Simone C B, Chen J, Lin A, Yom S S, Pattison A J, Carpenter C M and Solberg T D 2017 Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making Radiat. Oncol. 125 392–7

Wang M, Zhang Q, Lam S, Cai J and Yang R 2020 A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning Front. Oncol. 10 580919–580919

Zarepisheh M, Hong L, Zhou Y, Oh H, Mechelakos J G, Hunt M A, Mageras G S and Desay J O 2019 Automated intensity modulated treatment planning: the expedited constrained hierarchical optimization (ECHO) system Med. Phys. 46 2944–54

Zarepisheh M, Long T, Li N, Tian Z, Romeijn H E, Jia X and Jiang S B 2014 A DVH-guided IMRT optimization algorithm for automatic treatment planning and adaptive radiotherapy replanning Med. Phys. 41 061711

Zarepisheh M, Shakouirifar M, Trigila G, Ghomi P, Couzens S, Abebe A, Nooreña L, Shang W, Jiang S B and Zinchenko Y 2013 A moment-based approach for DVH-guided radiotherapy treatment plan optimization Phys. Med. Biol. 58 1869–87

Zinchenko Y, Craig T, Keller H, Terlaky T and Sharpe M 2008 Controlling the dose distribution with gEUD-type constraints within the convex radiotherapy optimization framework Phys. Med. Biol. 53 5231–50