Received: 28 June 2024 | Revised: 9 December 2024 | Accepted: 2 February 2025

DOI: 10.1002/mp.17692

RESEARCH ARTICLE

MEDICAL PHYSICS

Dose prediction via deep learning to enhance treatment planning of lung radiotherapy including simultaneous integrated boost techniques

Wenhua Cao[1] | **Mary Gronberg**[2] | **Stephen Bilton**[1] | **Hana Baroudi**[1] | **Skylar Gay**[1] | **Christopher Peeler**[1] | **Zhongxing Liao**[3] | **Thomas J. Whitaker**[1] | **Karen Hoffman**[3] | **Laurence E. Court**[1]

[1]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[2]Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, Texas, USA

[3]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Correspondence
Wenhua Cao, Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA.
Email: wcao1@mdanderson.org

Funding information
Varian Medical Systems; the NIH/NCI, Grant/Award Number: P30CA016672; Cancer Prevention Research Institute of Texas, Grant/Award Number: RP200395; the National Center for Advancing Translational Sciences of the NIH, Grant/Award Numbers: TL1TR003169, UL1TR003167

Abstract

Background: Recent studies have shown deep learning techniques are able to predict three-dimensional (3D) dose distributions of radiotherapy treatment plans. However, their use in dose prediction for treatments with varied prescription doses including simultaneous integrated boost (SIB), that is, using multiple prescription doses within the same plan, and benefit in improving plan quality should be validated.

Purpose: To investigate the feasibility and potential benefit of using deep learning to predict dose distribution of volumetric modulated arc therapy (VMAT) including SIB techniques and improve treatment planning for patients with lung cancer.

Methods: The dose prediction model was trained with 93 retrospective clinical VMAT plans for patients with lung cancer from our institutional patient database. The prescription doses of these plans ranged from 35 to 72 Gy, with various fractionation schemes. We used a 3D U-Net architecture to predict 3D dose distributions with 75 plans for training and 18 plans for testing. Model input consisted of computed tomography (CT) images, target and normal tissue contours and prescription doses. We first evaluated model accuracy by comparing the predicted and clinical plan doses for the test set, and then performed replanning according to predicted dose distributions. Furthermore, we evaluated the model prospectively in an additional set of 10 patients from our institution by two approaches where dose prediction was either blinded or provided to treatment planners. We then assessed whether dose prediction could identify suboptimal plan quality and how it affects plan quality if adopted in clinical planning workflow.

Results: The dose prediction model achieved good agreement between the predicted and clinical plan dose distributions, with a mean dose difference of -0.49 ± 0.54 Gy across the test set. The replanning study guided by dose prediction showed that a small subset of the original plans could benefit from improvements regarding sparing of the spinal cord and esophagus. The analysis of the prospective dataset, with initial and final clinical plans generated in the absence of dose prediction, showed that the predicted doses were able to identify possible improvements of target coverage and normal tissue sparing in the initial plans similar to those made by the final plans for majority of the patients, but in varied magnitudes. Moreover, the plans generated with dose prediction guidance were able to consistently improve normal tissue sparing compared to the plans generated without dose prediction guidance.

Conclusions: We demonstrated that our deep learning model can consistently predict high quality VMAT lung plans for a variety of prescription doses. The dose prediction tool was also effective in identifying suboptimal plan quality, suggesting its potential benefit in automated treatment planning and evaluation.

KEYWORDS
deep learning, dose prediction, lung cancer, plan quality improvement, volumetric modulated arc therapy

1 | INTRODUCTION

Intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) are common technologies of today's radiotherapy for treating lung cancer. These modalities, empowered by highly modulated treatment fields, can deliver higher doses to the tumor and lower doses to healthy surrounding tissues compared to conventional three-dimensional (3D) conformal radiotherapy. Prescription doses employed in lung cancer treatments can vary from 30 to 70 Gy or higher, which are determined by physicians to best suit individual patients' clinical scenarios, for example, retreatment, involvement of lymph nodes, dose escalation, and so forth. Moreover, a planning technique called simultaneous integrated boosting (SIB) has been increasingly adopted for treating lung cancer patients.[1] Toward the aim of improving tumor control while limiting normal tissue complications, an SIB plan is optimized to deliver a higher dose to the gross tumor volume (GTV) and a lower dose to the surrounding clinical target volume (CTV), in other words, prescribing different dose levels to specific target volumes. An IMRT or VMAT plan is typically created by inverse treatment planning, that is, the plan is optimized based on a set of dosimetric objectives and constraints provided by the treatment planner. However, radiotherapy plan quality is greatly dependent on the experience and skill of the planner, as well as the resources allocated in planning.[2–4] As more and more complex plans such as VMAT and SIB are utilized, the development of efficient methods to ensure consistent plan quality becomes critical.

Studies have extensively shown the promising use of knowledge-based planning (KBP) techniques to improve efficiency and consistency of treatment planning.[5–8] The concept of KBP is essentially to guide the treatment planning process by model-based goals or objectives, minimizing manual intervention by the planner. Initial studies of KBP mainly explored models that associate patient geometric features with dose distribution parameters such as the dose-volume histogram (DVH).[9–11] More recently, substantial progress has been made in using deep learning models to predict 3D dose distributions.[12–22] 3D dose prediction, by virtue of providing the spatial distribution of dose in the entire treatment volume, is superior to DVH prediction for evaluating plan quality, because it can better help KBP or the planner know where the tradeoffs of dose distributions among structures exist and how the plan could potentially be improved.[23–25] For example, the study by Gronberg et al showed the benefit of 3D dose prediction in identifying tradeoffs in planning to reduce dose to certain non-anatomical regions for head and neck cancer patients. However, dose prediction using deep learning for lung radiotherapy has not been well studied. Barragán-Montero et al. introduced a dose prediction model incorporating beam arrangement in lung IMRT treatments as a model input.[26] Shao et al. developed an asymmetrical network to predict dose comparing with other typically used UNet networks.[27] Jhanwar et al. investigated the benefit of using a moment-based loss function in their dose prediction model.[28] All these works focusing on lung radiotherapy considered only IMRT plans with a single prescription, whereas the present study focused on VMAT with multiple-prescription treatments.

In this work, we aimed to investigate the feasibility and potential benefit of using deep learning to predict dose distributions and improve treatment planning for patients with lung cancer treated with VMAT. Our training and test datasets include a variety of prescription doses, including multiple prescriptions per plan. The key contribution of our work was to test the feasibility of dose prediction methods for SIB treatment planning, which has not been well addressed in the literature. In addition, we reported a comprehensive prospective study evaluating the potential effectiveness of applying the prediction model in clinical workflow. Specifically, we tested the use of dose prediction either blinded or provided to treatment planners in the planning process; thus, we assessed whether it could identify suboptimal plan quality and be effective in clinical application.

2 | METHODS AND MATERIALS

2.1 | Patient data

We collected a cohort of 93 lung cancer patients that had been treated with VMAT at our institution between January 2020 through December 2022. Prescription doses used for treating these patients ranged from 35 to

TABLE 1 Treatment plan information of the training and test sets.

	Training set	Test set
Number of patients	75	18
Prescription doses used (Gy)	35, 37.5, 42, 45, 50, 52.5, 54, 56, 60, 63, 66, 70, 72	45, 50, 54, 56, 60, 66, 70
Number of SIB plans	56	13

Abbreviation: SIB, simultaneous integrated boost.

72 Gy, including 14 different levels, with various conventional fractionation schemes. We did not limit the patient selection by requiring certain prescription doses. However, we did exclude patients who received stereotactic body radiation therapy (SBRT) treatments because they are fundamentally different than conventional treatment strategies. Therefore, this dataset represents nearly all types of lung VMAT plans involved in our clinic except SBRT. All plans were optimized according to our clinical planning protocol and approved by attending physicians for treatment. The plans were typically delivered by 2–4 arcs. Among the VMAT plans, 69 of 93 used SIB, where a lower dose was prescribed to the CTV and the planning target volume (PTV), and a boosted dose was prescribed to the GTV. We randomly split the dataset into a training set of 75 plans and a test set of 18 plans. More details of the two subsets are described in Table 1.

2.2 | Deep learning approach

We used a 3D dense dilated U-Net architecture to predict dose distributions based on the input of computed tomography (CT) images, and contours of target and organ at risk (OAR) structures. This architecture has been validated with high accuracy in our previous works on VMAT dose prediction for head and neck[16,23] and gynecologic[24] cancers. It uses a densely connected sequence of dilated convolutions as the bottleneck level and each convolution is connected to all previous ones within the level (Figure 1). The implementation is fully 3D, with the CT input in 3D volume instead of 2D slices. In model training, we normalized the voxel intensities from (−1000, 1000) to (0, 1). The input of OARs as image masks included body, spinal cord, left and right lungs, heart, esophagus, liver, and left and right brachial plexuses. The input of the target was an image array of the PTVs and GTVs with the values of the corresponding prescription doses.

We trained the model on a server computer using an Nvidia V100 graphical processing unit with 16 GB memory. The Adam optimizer was used with a base learning rate of 0.001. The maximum epochs allowed was set as 1000. Patches of size (64, 64, 64) were randomly generated and used for model training.

The loss function in this work was selected from several variants based on performance. Those variants include mean squared error (MSE) based on the predicted and real doses, MSE plus a minimization term for the target, as well as DVH based losses using different weights factors for different structures. They were outperformed by the following loss function incorporating target prescribed doses.

As shown in Equation (1), the selected loss function L consists of MSE between the predicted and clinical dose distributions and the weighted sum of MSE between the predicted and prescribed doses in all target structures. Equal weights (denoted by λ_m for target m) for different target structures were found effective and therefore used in model training.

$$L = \mathrm{MSE}(D_{\mathrm{pred}}, D_{\mathrm{real}}) + \sum_m \lambda_m \mathrm{MSE}(D_{\mathrm{pred}}(\mathrm{Target}_{m,}), D_{\mathrm{prescribed}}(\mathrm{Target}_{m,})) \quad (1)$$

The training set was randomly grouped into 3 folds of 25 plans each. Three models were trained in which each was trained with data from 2 of the 3 folds and validated with data from the remaining fold. The final dose prediction for the test set was the average of the dose predictions from the 3 models.

2.3 | Test set evaluation and retrospective replanning study

For the 18 patients in the test set, we compared the dose predictions with the corresponding clinical plans by evaluating their DVH curves and 3D dose distributions and analyzing mean and maximum doses to structures including all targets, body, spinal cords, lungs, heart and liver, V5 and V20 of lungs, V30 and V40 of heart, V5 of liver, as well as additional target dose metrics including D2%, D98%, and a homogeneity index (HI) defined as $D5\%/D95\%$. A paired two samples one-sided t-test was performed with an equivalence bound of 1 Gy to determine the equivalence of the predicted and clinical mean doses.

Based on the predicted dose distributions, replanning of test patients was performed to assess if the predictions were achievable. A clinically commissioned RayStation treatment planning system (RaySearch Laboratories AB, Stockholm, Sweden), in which the original plans were planned, was used for this study. The replanning was performed when a clinical plan was identified inferior to the predicted dose in any of the dose volume metrics that had been used as clinical planning goals. As long as one criterion was found to be possibly inferior, even if all other criteria were superior, compared to dose prediction, the plan would still be flagged for re-optimization. That is, any potential trade-off between clinical goals was not considered at this stage.

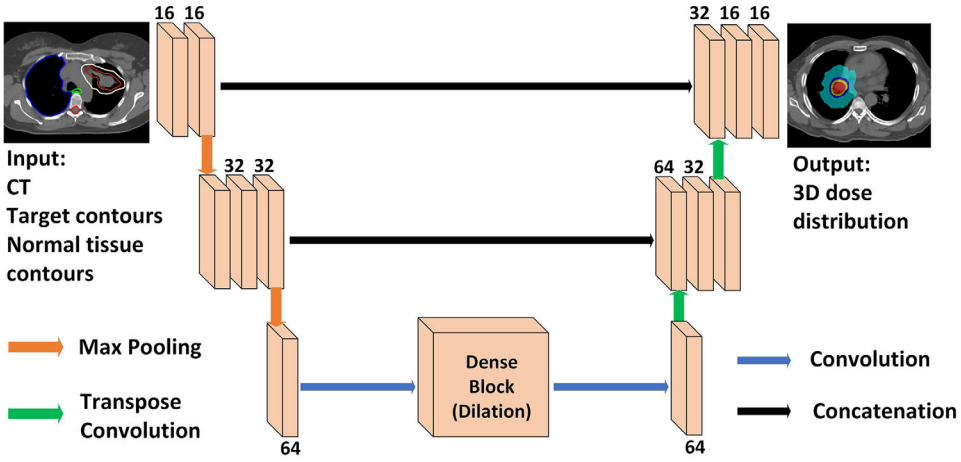

FIGURE 1 Schematic of a 3D dense dilated U-Net architecture. The dense block consists of densely connected dilated convolutions in two repeated sequences of convolutions with dilation rates of 1, 2, 5, and 9. The inputs of the architecture include the CT, target, and normal tissue contours, and the outputs are 3D dose distributions. 3D, three-dimensional; CT, computed tomography.

We took a manual iterative approach for replanning, as done in the clinical workflow. A "warm start" with the clinical plan was used to begin the optimization process. New objectives were added to the optimization, derived from the differences between predicted and clinical doses, indicating possible planning improvements. For example, if dose prediction indicated that a maximum dose of 25 Gy is potentially achievable for the spinal cord for a given patient and this was lower than that of the clinical plan, then a maximum dose objective for the spinal cord with a reference dose of 25 Gy or lower was added. When adjusting priorities of criteria, the general process was to adjust one priority at a time and increase priorities gradually until other criteria were affected. Priorities were set to be consistent with the ones used in the original planning. Typically, re-optimization iterations began with the structure or criterion for which the largest difference between the predicted dose and the current plan had been identified. This was repeated until the optimization of all flagged structures or criteria were evaluated. Ultimately, if there was a trade-off, that is, improvement on one dose criterion was at the cost of worsening one or more other criteria, the re-optimized plan would not be accepted. The final re-optimized plan was finalized after no further improvement could be achieved without compromising the clinical goals that had been achieved in the original clinical plan. The main steps of plan re-optimization are showed by the flowchart in Figure 2.

2.4 | Prospective planning study

2.4.1 | Planning without guidance of dose prediction

To assess the utility of dose prediction for identifying suboptimal plans and the potential of integrating dose prediction into existing clinical workflows, we collected initial optimized plans (i.e., the first try) and final approved clinical plans for 10 patients who received VMAT for lung cancer at our clinic between February through April 2024 and compared those plans with predicted dose distributions generated by our model. The initial plans were optimized using both target and normal tissue criteria according to each treatment planner's preference and experience to reach reasonably good quality (determined by the planner as well), but they were likely to be suboptimal as no extensive fine tuning or trial-and-error was done and physician feedback was not yet incorporated. On the other hand, the final approved plans were fully optimized, having undergone extensive trial and error as well as checks by physicians and physicists. During this process, optimization parameters were further adjusted, including the addition of objectives based on tuning structures (non-anatomical planning structures) until the identified dose-volume measures could no longer be improved. By comparing those plans with the predicted dose distributions, we evaluated whether the dosimetric improvements made within the current clinical workflow could be predicted by dose prediction. Both initial and final plans were generated by six experienced dosimetrists blinded to the predicted dose distributions.

2.4.2 | Planning with guidance of dose prediction

To further investigate the feasibility and impact of utilizing the dose prediction tool in a clinical setting, we conducted an additional study assimilating a clinical implementation with the same 10 prospective lung patients studied in Section 2.4.1. Two senior graduate students in medical physics were invited as treatment planner for this study. They both have worked on treatment planning

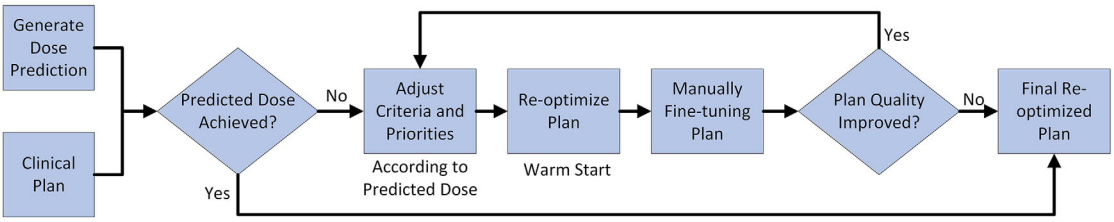

FIGURE 2 Flowchart illustrating the iterative process for plan re-optimization based on dose prediction.

research and received training for the clinical treatment planning system. Although their experiences were not as extended as clinical dosimetrists, they optimized the plans with the help of generic protocol dose criteria and predicted dose distributions so that all dosimetric goals prescribed per patient were met. Two plans were generated for each patient, that is, a standard plan, where its optimization was blinded to dose prediction, and a deep learning-guided (DL-guided) plan, where its optimization was guided by dose prediction. Specifically, the standard plans were generated first, then dose predictions were given, and the DL-guided plans were generated second. The optimization of DL-guided plan followed the similar approach as done in the plan re-optimization study (Figure 2), except that clinical plan was not used here.

3 | RESULTS

3.1 | Evaluation of predicted and clinical dose distributions on the test set

Dose prediction accuracy was evaluated on 18 test patients in this study. Among those 18 patients, 13 used SIB treatments, with prescribed PTV doses ranging from 45 to 60 Gy, boosting GTV doses ranging from 52.5 to 66 Gy. In the comparison of the predicted and clinical voxel doses within the treatment volume for all 18 patients, the mean dose difference was -0.49 ± 0.54 Gy. The paired one-sided *t*-test showed that the predicted and clinical dose distributions were equivalent ($p < 0.0001$). For target dose metrics, the percentage dose differences in D98% and D2% of PTVs were $-1.59\% \pm 1.93\%$ and $0.76\% \pm 1.51\%$, respectively. The mean HI for the predicted dose distributions was 1.10 ± 0.06, while the mean HI for the clinical plans was 1.08 ± 0.06.

Figure 3 shows the differences in mean and maximum doses and other dose volume indices for PTVs and OARs between the clinical plans and predicted doses. The median values of mean dose differences were slightly above zero for all structures, which indicates the predicted dose was lower than the clinical dose, except for the liver. However, for the majority of the structures, these differences were within ranges of 2 Gy for the mean dose and 4 Gy for the maximum dose. The largest differences in dose were found for the spinal cord, with a difference in mean dose of more than 5 Gy for two patients and a difference in max dose of more than 10 Gy for five patients. One possible explanation for the larger prediction error for the spinal cord is its smaller size compared to other structures. Additionally, there may be greater variation in spinal cord sparing in the clinical plans, as the maximum dose to the spinal cord might not be consistently prioritized during treatment planning when it falls well below the clinical limit of 45 Gy. This variation will also be discussed further in the section on replanning. The differences between the predicted and clinical plan dose in V5 were significantly larger than the differences in V20 for lungs. This reflects dose prediction in terms of low dose regions can vary greatly and may associate with large uncertainties. The differences in heart V30 and V40 showed good correlation with differences in heart mean dose (including outliers) in similar magnitude.

Figure 4 illustrates the dose distributions from the clinical plan and the prediction on an axial slice and DVHs for one of the test patients. The dose distributions show that the predicted dose was slightly more conformal to the PTV than the clinical plan, while sparing of the esophagus and spinal cord shows no difference. The DVHs also demonstrate the two dose distributions are not different in a manner that is clinically significant.

3.2 | Retrospective replanning study

A study was conducted on the test set plans to evaluate whether the predicted doses could be achieved through replanning the clinical plans. The reoptimized, predicted and original clinical plan doses for the 18 test patients are intercompared. Figure 5 shows both the prediction and reoptimized plan have good agreement with the clinical plan for PTV mean dose and D98%, except in one case with lower D98% in prediction than the clinical plan. This indicates that all the original plans had reasonable doses to the targets. Hence the plan re-optimization primarily focused on normal tissue sparing. The differences in mean or maximum doses to the OARs (esophagus, spinal cord, lung, and heart) between

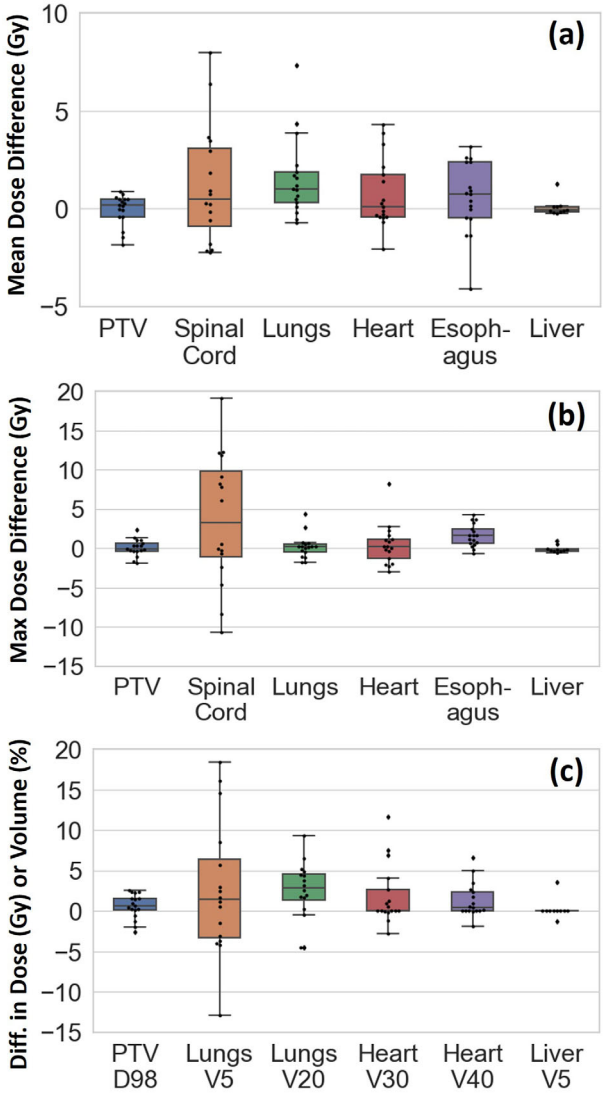

FIGURE 3 Differences in mean dose (a), maximum dose (b), and selected dose volume indices (c) for PTV and OARs between the clinical (achieved) and predicted dose distributions (achieved minus predicted) for the 18 patients in the test set are summarized in box plots. Individual data points are shown by black dots. OAR; organ at risk; PTV, planning target volume.

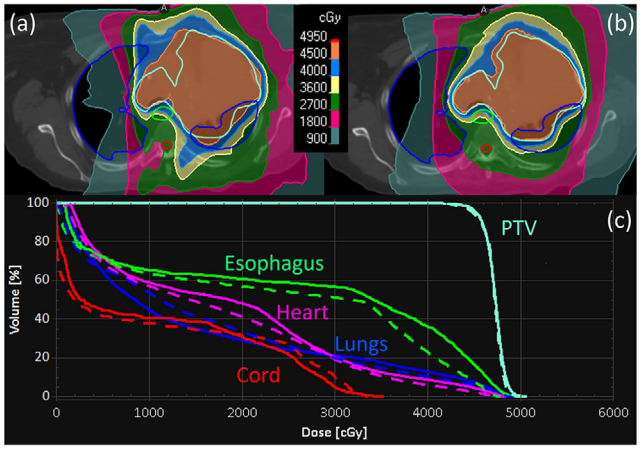

FIGURE 4 Comparison of clinical and predicted dose distributions for an example case in the test set. Shown are dose distributions in an axial view between the clinical (achieved) plan (a) and the prediction (b), where the PTV, esophagus, lungs, and spinal cord are outlined in cyan, green, blue, and red. In the DVHs (c), the clinical plan is shown by solid lines while the predicted dose is shown by dashed lines. DVH, dose-volume histogram; PTV, planning target volume.

the clinical or reoptimized plans and the predicted dose for each patient are shown in Figure 6. In this comparison, dose prediction found 32% (35 out of 108) of the OARs among all patients where the mean or maximum doses of the clinical plans could be reduced by 2 Gy or more, without compromising target coverage and sparing of other structures. Among those 35 OARs, the reoptimized plan was always able to achieve reduced doses, but in varied magnitudes, for all occasions except for one — the maximum esophagus dose for Patient 3.

It is worth noting that the reductions in mean dose by re-optimization for different OARs were mostly modest, that is, no more than 2 Gy in difference of mean or maximum dose. However, there were several cases with substantial dosimetric improvements. In one case (Patient 4), the maximum dose to the esophagus was reduced by over 5 Gy, and in two cases (Patient 1 and 14), the maximum dose to the spinal cord was reduced by over 10 Gy. This implies that dose prediction could identify suboptimal dosimetric indications and lead to improved plan quality. However, the predicted sparing gains were not always achievable. For example, the re-optimized plans had nearly the same mean lung dose as the clinical plans for Patients 14 and 15, where the predicted reductions were more than 2 Gy, that is, over prediction on lung sparing for these cases. There were also 7 OARs among all patients where the predicted doses had higher dose than the clinical plans with a difference over 2 Gy, which highlights the uncertainty in dose prediction and the replanning found those doses could not be changed (e.g., the mean esophagus dose for Patient 18). Additionally, comparisons on V20 of lungs and V30 of heart exhibited a similar trend as shown in comparisons on mean dose to lungs and heart, respectively. In other words, for example, patients showing benefits from re-optimization in reducing mean lung dose also had benefits in reducing V20.

Figure 7 shows the dose distributions and DVHs of the original plan, re-optimized plan, and prediction in an example case (i.e., Patient 1 in Figure 6) treated with an SIB plan, with a boosted dose prescribed to the GTV. Compared to the original clinical plan, the reoptimized plan significantly reduced the spinal cord maximum dose while maintaining the target coverage and sparing of other OARs. With marked difference in spinal cord sparing shown in the predicted dose distribution, the re-optimized plan was able to further reduce the maximum dose close to the predicted value. This

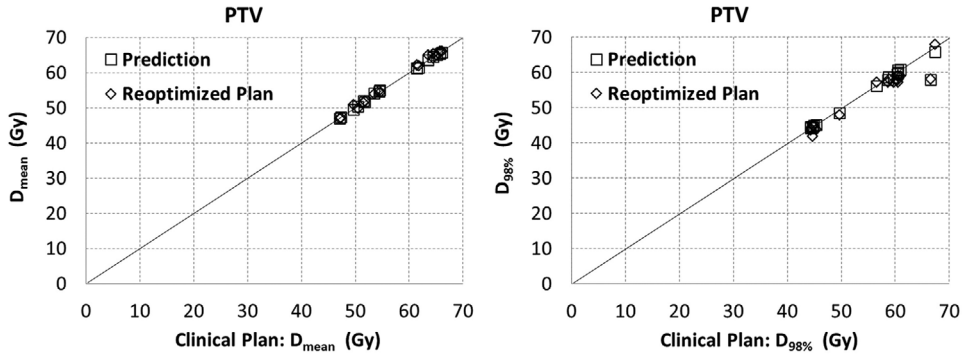

FIGURE 5 Comparison of mean dose and D98% of PTVs for the clinical, predicted, and reoptimized dose distributions for the 18 patients in the test set. PTV, planning target volume.

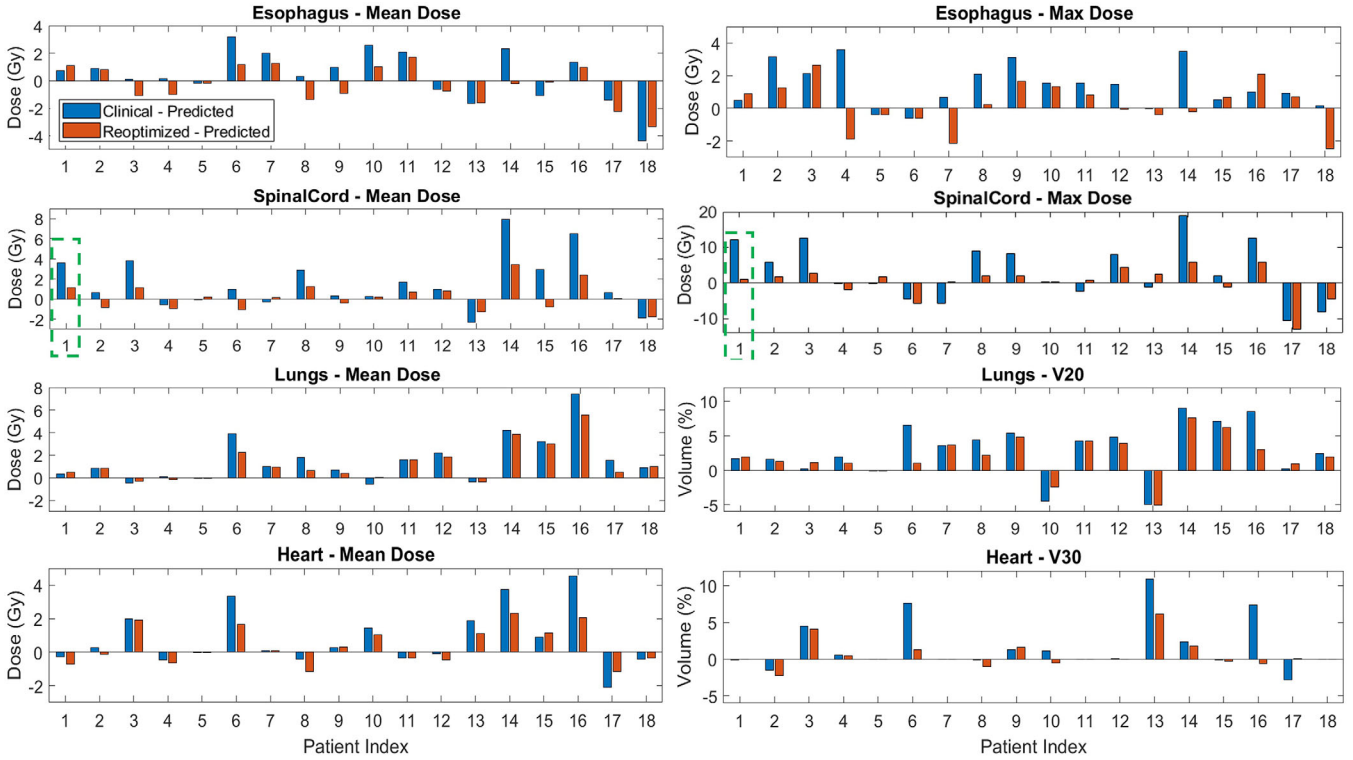

FIGURE 6 Differences in selected dose volume indices of esophagus, spinal cord, lungs, and heart between the clinical or reoptimized plans and the predicted doses for the 18 patients in the test set. A positive difference means the clinical (blue bar) or reoptimized plan (red bar) has higher mean or maximum dose than the predicted dose. In addition, a larger positive difference between the clinical plan and predicted dose (blue bar) indicates a greater reduction in mean or maximum dose might be possible according to prediction; whereas a smaller difference between the reoptimized plan and the predicted dose (red bar) indicates the prediction was achieved by plan re-optimization within a smaller margin. For example, the highlighted Patient 1 (outlined in green dashed lines) shows that the predicted reductions in the mean and maximum doses of the clinical plan were 4 and 12 Gy for the spinal cord, while the reoptimized plan reduced the mean and maximum doses by 3 and 10 Gy compared to the clinical plan.

suggests that the original treatment planning for this case did not minimize the spinal cord maximum dose as low as possible, given the large margin between the achieved dose (35 Gy) and the prescribed limit (45 Gy), while dose prediction based on a population of plans indicated a lower maximum dose could be achievable.

To assess the original clinical plans and the re-optimized plans for all test patients, Figure 8 summarizes the mean and maximum doses to spinal cord, lungs heart and esophagus for the two plans, as well as predicted doses for all patients in the test set. It shows the medians of both mean and max doses were reduced by the re-optimized plan for the spinal cord and esophagus.

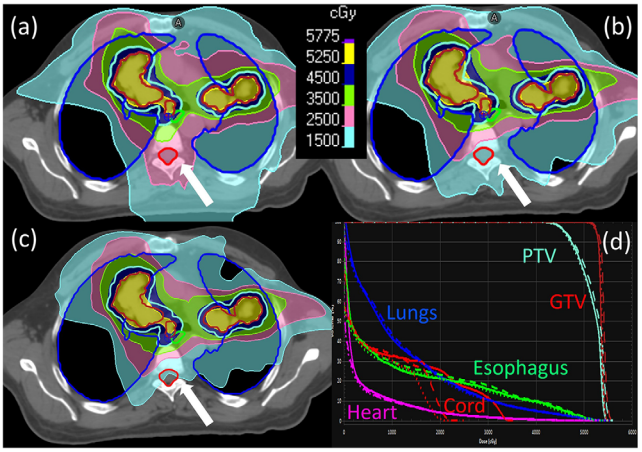

FIGURE 7 Comparison among the original plan, reoptimized plan, and predicted dose distribution for an example case. Shown are dose distributions on an axial slice from the original (a) and reoptimized (b) plans, as well as prediction (c). The white arrows point to the region where the spinal cord (outlined in red) is covered by the 25 Gy isodose line (dark pink) in the original plan while entirely spared in the reoptimized plan. In the DVHs (d), the original plan is shown by solid lines while the reoptimized plan is shown by dashed lines; the dose prediction is shown by dotted lines. DVH, dose-volume histogram.

The median of the max (mean) doses of the spinal cord for clinical, reoptimized, and predicted doses was 33.2 (9.3) Gy, 26.4 (8.6) Gy, and 27.8 (7.7) Gy, respectively. The median of the max (mean) doses of the esophagus for clinical, reoptimized, and predicted doses was 61.5 (19.1) Gy, 59.5 (18.4) Gy, and 60.0 (18.7) Gy, respectively. The upper thresholds of mean doses were also reduced by the re-optimized plan for lungs and heart. However, the differences in median of mean or max doses for the lungs and heart between clinical and optimized plans were negligible (less than 0.2 Gy).

3.3 | Prospective planning study

We conducted a prospective study first to evaluate the dose prediction model, aiming to determine its effectiveness in identifying suboptimal plans and compare with a typical clinical workflow. The comparison of potentially suboptimal initial plans, final clinical plans and predicted doses for the prospective set of 10 patients is summarized in Figure 9. Data for individual patients can be found in Appendix (Table S1a–c). For PTVs, dose predictions indicated that increase of mean dose for the initial plans was possible. The final plans showed consistent reduction in its differences to predicted doses, that is, increase of PTV mean dose, including one case where more than an increase of 5 Gy was found by both the final plan and dose prediction. For OARs, dose predictions indicated better sparing in terms of mean or maximum doses was possible for majority of the initial plans. The medians of the differential doses in the initial plans were reduced in the final plans for all OARs, indicating improved OAR sparing in final plans as predicted by dose prediction. Large variation of these differential doses for the final plans was observed for spinal cord compared to other OARs. For example, one case had a maximum spinal cord dose in the final plan 9 Gy higher than the predicted dose. We found, for this case, the final plan reduced mean doses to the lung, esophagus, and heart compared to the initial plan as predicted, but had an increase in maximum spinal cord dose. For heart and esophagus mean doses, the majority of the final plans achieved greater improvements on the initial plan than predicted. For liver mean doses, there were outliers that the final plans had marked difference over the predicted doses. We found, in the most extreme case where mean liver dose in the final plan was 9 Gy higher than the predicted dose, that the mean liver dose could be reduced through further plan optimization without compromising other dosimetric measures. The final plan did not minimize liver dose as low as possible may be because it had already achieved low dose, for example, well below the dose requirements of V30 $\leq$ 40% and mean dose $<$ 30 Gy.

In the second planning study where dose prediction guidance was incorporated in plan generation, we aimed to evaluate the dose prediction model's potential for clinical implementation. Here we found the DL-guided plans consistently outperformed the standard plans in terms of OAR sparing. It should be noted that both plans were optimized to meet all clinically prescribed dose limits for each patient. Figure 10 shows the difference between the standard plan or DL-guided plan and predicted dose for PTV and spinal cord, lungs, heart, esophagus, and liver for the 10 prospective patients. Data for individual patients can be found in Appendix (Table S2a–c). For majority of those patients, OAR doses (mean or maximum) in the standard plans were higher than predicted doses, except heart. In the DL-guided plans, all OAR doses were lower than those in the standard plans; some were lower than the predicted doses, such as spinal cord, heart, and esophagus. Within all OARs, spinal cord was seen with most improvement in sparing, where the median of maximum doses was reduced by 5 Gy by the DL-guided plans compared to the standard plans.

Figure 11 compares the DL-guided plan, standard plan, and predicted dose for an example case from the prospectively assessed dataset. While showing similar target coverage, the DL-guided plan demonstrated better sparing of medium doses to the esophagus, as well as a smaller volume of 25 Gy to normal tissues, compared to the standard plan. The DVH also shows dose exposure to right lung was also reduced by the DL-guided plan, which correlates with the predicted dose.

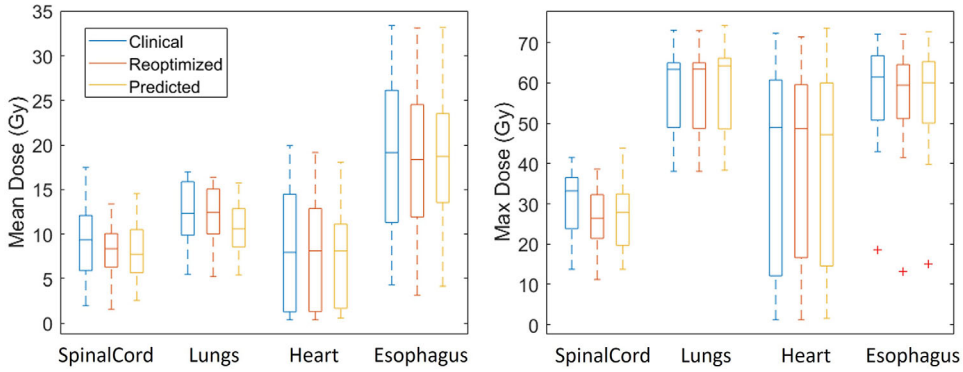

FIGURE 8 Boxplots comparing mean and maximum dose for selected OARs among clinical plans, re-optimized plans, and predicted dose distributions. OAR, organ at risk.

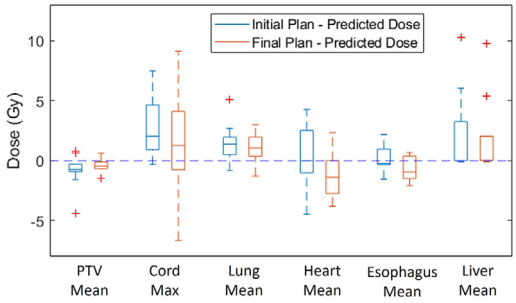

FIGURE 9 Boxplots of differences in mean or maximum doses of PTV and selected OARs between initial or final plans and predicted dose for 10 prospective patients. Data points above 0 mean that the dose of the initial plan (blue) or the final plan (red) is higher than the predicted dose, whereas data points below 0 mean that the dose of the initial plan (blue) or the final plan (red) is lower than the predicted dose. For PTVs, one outlier of lower mean dose than the predicted dose in the initial plan was improved in the final plan. For OARs, almost all medians and upper limits of the differential dose in the initial plans were reduced in the final plans. Larger variation of the dose differences was seen for spinal cord compared to other OARs. OAR, organ at risk.

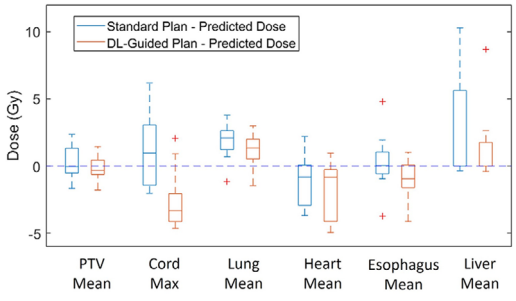

FIGURE 10 Boxplots of differences in mean or maximum doses between the standard plan or DL-guided plan and predicted dose for PTV and selected OARs for 10 patients. For OAR mean or maximum doses, a data point above 0 means that the dose of the standard plan or DL-guided plan is higher than the predicted dose. The medians of the differences for OARs were consistently reduced by the DL-guided plans, which indicates sparing of these OARs were improved in the DL-guided plans compared to the standard plans. DL-guided, deep learning-guided; OAR, organ at risk; PTV, planning target volume.

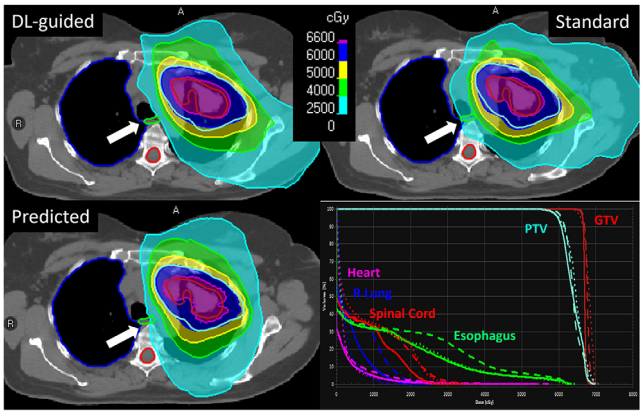

FIGURE 11 An example patient with DL-guided plan, standard plan (generated without guidance from dose prediction), and predicted dose. Dose distributions on an axial slice for the three plans are shown. The white arrows point to the region where the sparing of the esophagus (outlined in light green) by the 25 Gy isodose line was improved in the DL-guided plan and the predicted dose compared to the standard plan. The DVH curves of the targets and OARs compare the DL-guided plan (solid line), standard plan (dashed line), and predicted dose (dotted line). DL-guided, deep learning-guided; DVH, dose-volume histogram; OAR, organ at risk.

4 | DISCUSSION

In this study, we investigated the performance of deep learning-based 3D dose prediction for lung VMAT plans and the potential of its use in automated plan evaluation. To our knowledge, this is the first study focusing on VMAT dose prediction for lung cancer and incorporating multiple prescription doses including SIB. The dataset used in this study encompasses a wide range of prescription doses and more than half of plans are SIB. This suggests high flexibility of the dose prediction model, that is, using a single deep learning architecture to predict on multiple prescription doses either in different plans or in the same plan (e.g., SIB). Our results showed that the predicted and ground truth dose distributions were in good agreement. In addition, plan re-optimization according to dose prediction found

instances where sparing of the OARs could be further improved from the clinical plans.

Besides the help of using a dataset with various prescription doses in model training, the benefit of the present model in predicting multiple target dose levels is mainly attributed to the selected loss function. The majority of dose prediction methods in the literature have been focusing on loss functions based on the difference between the predicted and real dose distributions only, including various forms such as MSE, mean absolute error (MAE), DVH, moment loss, and so forth. In this work, the added loss between the predicted and prescribed dose to the target, in addition to the conventional MSE loss, enhanced the training to predict doses with multiple prescriptions.

Another unique aspect of this study is the evaluation of potentially suboptimal plans for prospective patients, collected from the clinical planning workflow at our institution. The results showed that the dose predictions were able to identify possible improvements to the initial plans (prior to manual fine-tuning and finalizing) similar to those made by the final clinical plans for most of the patients but in varied magnitudes (Figure 9). We should acknowledge here that the initial plans might have had mixed plan quality due to the variation in experience and preferences of the individual dosimetrists. Nevertheless, because the planning of the final plans was blinded to the predicted dose distributions, it was promising to observe that plan quality improvements made in the manual clinical planning process were also indicated by model prediction.

The other planning study with dose prediction guidance for the prospective patients showed how the final plan quality was affected if the predicted dose distributions were shown to the planner. We found that the DL-guided plans outperformed the standard plans in OAR sparing consistently. This shows the potential benefit of dose prediction in improving plan quality for a manual planning approach (e.g., Figure 2) very similar to current clinical practice. More importantly, our results suggest that dose prediction can help relatively inexperienced treatment planner to achieve plans in higher quality, as those DL-guided plans showed clear advantage over the standard plans generated without dose prediction guidance (Figure 10). This highlights the advantage of dose prediction-based planning in improving the consistency of plan quality and mitigating the impact of treatment planner variations. With regard to planning efficiency, although a formal timing study was not performed, we estimated that it took less than 30 min to re-optimize each case.

The development of 3D dose prediction offers a major enhancement to KBP approaches. Current commercial KBP systems are mostly based on models for predicting dose-volume parameters, such as RapidPlan and others.[29–31] Such models are limited by the inability to provide spatial distribution of dose in patient, which is particularly important in plan evaluation and for planners to obtain insights to further optimize the plans. 3D dose prediction not only addresses these limitations, but also can be translated into deliverable plans by automated planning techniques more effectively than DVH prediction methods.[13,32] The model developed in the present study will be used as an automated plan evaluation and quality assurance tool in our practice going forward. As DL-guided planning in this study was limited by using conventional manual input of dose-volume indices based on predicted dose, instead of more advanced techniques such as dose mimicking, our future development will focus on integrating the dose prediction model into an end-to-end automated planning workflow for lung VMAT.

Regarding prediction accuracy, one observation from our study was that dose prediction tended to overestimate normal tissue sparing. For example, as shown in Figure 1, the medians of mean dose differences of lungs, heart, spinal cord, and esophagus between clinical and predicted dose distributions among the test set patients were all positive (no more than 1 Gy), and the predicted mean doses were lower than clinical plan ones for the majority of these OARs and patients. This bias of dose prediction, also discussed by other works,[25] is possibly due to the multi-objective nature of treatment planning. Clinical planning can result in multiple, equally acceptable, dose distributions for one patient or multiple patients with similar geometry, with varied spatial distributions of dose in the OARs. However, deep learning models accumulate a large number of dose distributions over voxels by “averaging” errors, MSE in this study, and possibly lead to lower OAR doses than what is achievable. In lung radiotherapy, spatial variability in clinically acceptable dose distributions is particularly high, as tumor can locate anywhere in the lungs. Consequently, the impact of this dose prediction bias could be greater for lung than other cancer sites, such as prostate and breast. Nevertheless, the above conjecture needs to be tested by further studies possibly utilizing training data with stratified homogeneity.

It is possible that the performance of the dose prediction model could be further improved by filtering the quality of the training data. The current approach assumed all clinical plans collected were of high quality. However, our replanning study found that a small subset of clinical plans in the test set could have better spared the spinal cord and esophagus (Figure 6). As variations in the quality of manually optimized plans are likely to exist, an additional manual review and filtering of input data before it is entered into model training might enhance our dose prediction. With regard to the evaluation of model performance, it should be highlighted that analyses of results in average are not sufficient, and each individual patient needs to be studied (e.g., Figures 5–7 and 11). Even if the average comparison

shows good agreement, variation between patients could be nontrivial.

Another method to further improve dose prediction performance in plan quality improvement could be transfer leaning, by utilizing plans re-optimized with guidance derived from dose prediction models as implemented in our study. Specifically, the re-optimized plans could serve as a dataset with improved dosimetric characteristics that, with the use of transfer learning, could help refine the model so that it predicts higher-quality dose distributions. In general, the size of datasets required for transfer learning is only a fraction (1/3 or smaller) than that needed to train a deep learning dose prediction model initialized with random weights.[18] Consequently, techniques like transfer learning have the potential to allow one to build a high-performance model with a small set of carefully selected high-quality plans. However, further research is needed to understand the interplay relationship between the re-optimized data and the data used to train the original model, and to develop evaluation strategies to properly compare the quality of models before and after transfer learning. We are currently exploring this research direction.

Accurate dose prediction for lung radiotherapy can provide significant value in improving treatment quality and safety for lung cancer patients. Two immediate extensions to our model would be applications in SBRT and proton therapy. As increasingly more SBRT treatments are adopted today, dose prediction allows fast dosimetric evaluation of treatment plans to assist decision-making for fractionation and adaptation. In proton therapy, dose prediction can expedite the robustness planning process and assist with generating scanning beam treatments that are less sensitive to patient breathing motion.

5 | CONCLUSION

We demonstrated that our deep learning model can consistently predict high-quality VMAT lung plans for a variety of prescription doses, including SIB techniques. In two planning studies with the predicted dose blinded or provided to the treatment planners on a prospective patient cohort, dose prediction was effective in identifying suboptimal plans and improving plan quality, suggesting the potential benefit of the model in automated treatment planning and its clinical application.

ACKNOWLEDGMENTS

The authors would like to thank Fernanda Gomez, Jinu James, Amyn Joseph, Carrie Lujano, Dana Martinez, and Joshua Solis from the dosimetrist team at MD Anderson Cancer Center for their assistance in generating radiotherapy treatment plans for this study. The manuscript was edited by Sarah Bronson, ELS, of the Research Medical Library at MD Anderson. We also acknowledge the support of the High-Performance Computing for Research Facility at MD Anderson for providing computational resources. This work was supported in part by Varian Medical Systems. It was also supported by the NIH/NCI under award number P30CA016672. M.G. would like to acknowledge support from the Cancer Prevention Research Institute of Texas under award number RP200395 and the National Center for Advancing Translational Sciences of the NIH under award numbers TL1TR003169 and UL1TR003167. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

REFERENCES

1. Jeter MD, Gomez D, Nguyen QN, et al. Simultaneous integrated boost for radiation dose escalation to the gross tumor volume with intensity modulated (photon) radiation therapy or intensity modulated proton therapy and concurrent chemotherapy for stage ii to iii non-small cell lung cancer: a phase 1 study. *Int J Radiat Oncol Biol Phys*. 2018;100(3):730-737.
2. Nelms BE, Robinson G, Markham J, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. *Pract Radiat Oncol*. 2012;2(4):296-305.
3. Berry SL, Boczkowski A, Ma R, et al. Interobserver variability in radiation therapy plan output: results of a single-institution study. *Pract Radiat Oncol*. 2016;6(6):442-449.
4. Kubo K, Monzen H, Ishii K, et al. Inter-planner variation in treatment-plan quality of plans created with a knowledge-based treatment planning system. *Phys Med*. 2019;67:132-140.
5. Fogliata A, Belosi F, Clivio A, et al. On the pre-clinical validation of a commercial model-based optimisation engine: application to volumetric modulated arc therapy for patients with lung or prostate cancer. *Radiother Oncol*. 2014;113(3):385-391.
6. Tol JP, Delaney AR, Dahele M, et al. Evaluation of a knowledge-based planning solution for head and neck cancer. *Int J Radiat Oncol Biol Phys*. 2015;91(3):612-620.
7. Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiother Oncol*. 2016;120(3):473-479.
8. Li N, Carmona R, Sirak I, et al. Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials. *Int J Radiat Oncol Biol Phys*. 2017;97(1):164-172.
9. Wu B, Ricchetti F, Sanguineti G, et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. *Med Phys*. 2009;36(12):5497-5505.
10. Appenzoller LM, Michalski JM, Thorstad WL, et al. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys*. 2012;39(12):7446-7461.
11. Tol JP, Dahele M, Delaney AR, et al. Can knowledge-based DVH predictions be used for automated, individualized quality assurance of radiotherapy treatment plans?. *Radiat Oncol*. 2015;10:234.
12. Nguyen D, Jia X, Sher D, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Phys Med Biol*. 2019;64(6):065020.

13. Fan J, Wang J, Chen Z, et al. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. *Med Phys*. 2019;46(1):370-381.
14. Chen X, Men K, Li Y, et al. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. *Med Phys*. 2019;46(1):56-64.
15. Liu S, Zhang J, Li T, et al. Technical note: a cascade 3D U-Net for dose prediction in radiotherapy. *Med Phys*. 2021;48(9):5574-5582.
16. Gronberg MP, Gay SS, Netherton TJ, et al. Technical Note: dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture. *Med Phys*. 2021;48(9):5567-5573.
17. Lempart M, Benedek H, Jamtheim Gustafsson C, et al. Volumetric modulated arc therapy dose prediction and deliverable treatment plan generation for prostate cancer patients using a densely connected deep learning model. *Phys Imaging Radiat Oncol*. 2021;19:112-119.
18. Kandalan RN, Nguyen D, Rezaeian NH, et al. Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices. *Radiother Oncol*. 2020;153:228-235.
19. Ahn SH, Kim E, Kim C, et al. Deep learning method for prediction of patient-specific dose distribution in breast cancer. *Radiat Oncol*. 2021;16(1):154.
20. Hedden N, Xu H. Radiation therapy dose prediction for left-sided breast cancers using two-dimensional and three-dimensional deep learning models. *Phys Med*. 2021;83:101-107.
21. Song Y, Hu J, Liu Y, et al. Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy. *Radiother Oncol*. 2020;149:111-116.
22. Zhang G, Jiang Z, Zhu J, et al. Dose prediction for cervical cancer VMAT patients with a full-scale 3D-cGAN-based model and the comparison of different input data on the prediction results. *Radiat Oncol*. 2022;17(1):179.
23. Gronberg MP, Beadle BM, Garden AS, et al. Deep learning-based dose prediction for automated, individualized quality assurance of head and neck radiation therapy plans. *Pract Radiat Oncol*. 2023;13(3):e282-e291.
24. Gronberg MP, Jhingran A, Netherton TJ, et al. Deep learning-based dose prediction to improve the plan quality of volumetric modulated arc therapy for gynecologic cancers. *Med Phys*. 2023;50(11):6639-6648.
25. McIntosh C, Conroy L, Tjong MC, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med*. 2021;27(6):999-1005.
26. Barragán-Montero AM, Nguyen D, Lu W, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. *Med Phys*. 2019;46(8):3679-3691.
27. Shao Y, Zhang X, Wu G, et al. Prediction of three-dimensional radiotherapy optimal dose distributions for lung cancer patients with asymmetric network. *IEEE J Biomed Health Inform*. 2021;25(4):1120-1127.
28. Jhanwar G, Dahiya N, Ghahremani P, Zarepisheh M, Nadeem S. Domain knowledge driven 3D dose prediction using moment-based loss function. *Phys Med Biol*. 2022;67(18):185017.
29. Cornell M, Kaderka R, Hild SJ, et al. Noninferiority study of automated knowledge-based planning versus human-driven optimization across multiple disease sites. *Int J Radiat Oncol Biol Phys*. 2020;106(2):430-439.
30. Krayenbuehl J, Zamburlini M, Ghandour S, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. *Radiat Oncol*. 2018;13(1):170.
31. Cilla S, Romano C, Morabito VE, et al. Personalized treatment planning automation in prostate cancer radiation oncology: a comprehensive dosimetric study. *Front Oncol*. 2021;11:636529.
32. McIntosh C, Welch M, McNiven A, et al. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. *Phys Med Biol*. 2017;62(15):5926-5944.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cao W, Gronberg M, Bilton S, et al. Dose prediction via deep learning to enhance treatment planning of lung radiotherapy including simultaneous integrated boost techniques. *Med Phys*. 2025;52:3336–3347. https://doi.org/10.1002/mp.17692