WILEY

Boosting radiotherapy dose calculation accuracy with deep learning

Yixun Xing | You Zhang | Dan Nguyen | Mu-Han Lin | Weiguo Lu | Steve Jiang

Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX,75390,USA

Authors to whom correspondence should be addressed. You Zhang and Steve Jiang
E-mails: you.zhang@utsouthwestern.edu, steve.jiang@utsouthwestern.edu;
Telephones: 214-645-2699, 214-645-8510.

Abstract

In radiotherapy, a trade-off exists between computational workload/speed and dose calculation accuracy. Calculation methods like pencil-beam convolution can be much faster than Monte-Carlo methods, but less accurate. The dose difference, mostly caused by inhomogeneities and electronic disequilibrium, is highly correlated with the dose distribution and the underlying anatomical tissue density. We hypothesize that a conversion scheme can be established to boost low-accuracy doses to high-accuracy, using intensity information obtained from computed tomography (CT) images. A deep learning-driven framework was developed to test the hypothesis by converting between two commercially available dose calculation methods: Anisotropic analytic algorithm (AAA) and Acuros XB (AXB). A hierarchically dense U-Net model was developed to boost the accuracy of AAA dose toward the AXB level. The network contained multiple layers of varying feature sizes to learn their dose differences, in relationship to CT, both locally and globally. Anisotropic analytic algorithm and AXB doses were calculated in pairs for 120 lung radiotherapy plans covering various treatment techniques, beam energies, tumor locations, and dose levels. For each case, the CT and the AAA dose were used as the input and the AXB dose as the "ground-truth" output, to train and test the model. The mean squared errors (MSEs) and gamma passing rates (2 mm/2% & 1 mm/1%) were calculated between the boosted AAA doses and the "ground-truth" AXB doses. The boosted AAA doses demonstrated substantially improved match to the "ground-truth" AXB doses, with average (± s.d.) gamma passing rate (1 mm/1%) 97.6% (±2.4%) compared to 87.8% (±9.0%) of the original AAA doses. The corresponding average MSE was 0.11(±0.05) vs 0.31(±0.21). Deep learning is able to capture the differences between dose calculation algorithms to boost the low-accuracy algorithms. By combining a less accurate dose calculation algorithm with a trained deep learning model, dose calculation can potentially achieve both high accuracy and efficiency.

KEY WORDS
AAA, AXB, CT, deep learning, dose calculation, inhomogeneous regions

1 | INTRODUCTION

In radiation therapy, the radiation doses to the tumor and surrounding normal tissues directly determine treatment efficacy and safety.[1,2] It is pivotal for the radiation therapy treatment planning systems (TPSs) to accurately calculate the dose distributions to aid physician decisions. Accurate dose calculation is also key to a reliable and reproducible model between dose distributions and clinical outcomes to guide future treatments.[3] The dose calculation algorithms have seen generations of development. The early generations of algorithms, usually referred to as *correction-based* methods,[4–7] are barely physics principle driven. Their accuracy is highly unreliable in heterogeneous regions (for instance, areas with lung tissue surrounding tumors), where the loss of electronic equilibrium occurs.[8] To better model the physics in these regions, *model-based* techniques were developed.[9–12] These techniques model the radiation energy transport via dose kernels and convolutions. To account for the tissue heterogeneity, the dose kernels are scaled based on the equivalent electron density path lengths encountered by radiation beams, leading to the superposition–convolution type of algorithm, which is widely used in today's clinic. Analytical anisotropic algorithm (AAA), one of such algorithms, is commercially implemented in the Eclipse TPS[12] (Varian Medical Systems, Palo Alto, CA). However, a discrepancy over 5% from measurements can still be observed for AAA calculations in inhomogeneous regions,[13] which can be clinically significant.[14,15] The discrepancy is due to the fact that kernel scaling does not explicitly and realistically model the energy transport through physical interactions either. The Monte Carlo algorithm represents a third type of dose calculation algorithms. It models the transport and energy deposition of each particle (photons, electrons, etc.) via explicit physics principles, which are modeled with measured data or proven formula, and provides the highest accuracy.[16,17] However, Monte Carlo needs to simulate the transport of each particle individually, which requires substantial computational power and may significantly prolong the dose calculation time. Besides Monte Carlo, recently a new dose calculation technique using the linear Boltzmann transport equation was implemented as the Acuros XB (AXB) algorithm in Eclipse.[18,19] It is proven that AXB would theoretically converge to the same solution as the Monte Carlo algorithm.[20] The accuracy of AXB has been extensively validated.[17,21,22] The efficiency of AXB is plan dependent, which improves with increasing beam numbers (relatively) and favors volumetric-modulated arc therapy plans. In some scenarios, however, AXB can be 10 times slower than AAA.[23]

In general, a trade-off exists on dose calculation: more accurate dose calculation requires more computational power, and is generally more time consuming and resource demanding. Due to this trade-off, current radiotherapy TPSs may have to use less accurate methods for dose calculation, in order to improve efficiency, especially during plan optimization.[24] Such an adoption is less ideal, since these low-accuracy dose calculations used during optimization may potentially trap the optimization into a local optimum, and yield a suboptimal final plan. As the radiotherapy society is pursuing more precise, individual-tailored treatments, the use of low-accuracy dose calculation algorithms for plan optimization may fail to generate high-

quality plans within a tight time frame, especially for online adaptive radiotherapy.[25] A technique enabling dose calculation with both high accuracy and efficiency is thus much desired for plan optimization. In addition, recent advancements of real-time imaging techniques also call for such a technique,[26] which will make possible on-the-fly dose monitoring and intervention through real-time plan re-optimization and adaptation.

As mentioned, the differences between low-accuracy and high-accuracy doses are mostly within inhomogeneous regions. The inhomogeneity, however, is fully captured in CT images that are used for dose calculation. From this observation, we hypothesize that the differences between dose algorithms can be learned and correlated with the dose distribution and the CT intensity information. With this learnt correlation, we can then quickly boost the low-accuracy doses to high-accuracy, to overcome the trade-off between dose calculation accuracy and efficiency. Recently, the developments and applications of artificial intelligence (AI) in radiation therapy have seen tremendous growth.[27–32] Sophisticated convolutional neural networks can handle intensive tasks including medical image de-noising, segmentation, treatment plan optimization/evaluation, and clinical outcome prediction. Some networks, including the U-Net,[33] can perform voxel-wise prediction and mapping, which allows a potential voxel-to-voxel dose map conversion to boost the accuracy of low-accuracy doses. Additionally, the U-Net can extract both global and local features from dose distributions and CT images, which can be directly correlated with dose differences between algorithms, as the energy transport is essentially determined by both long-range (global) photon transport and short-range (local) electron transport. With dedicated graphics processing units (GPUs), the inference of U-Net can also be executed within seconds, meeting the efficiency requirement. Driven by our hypothesis, in this study we introduced an AI-based framework to achieve rapid, direct 3D dose map conversion from low-accuracy doses to high-accuracy doses. We trained and evaluated the whole framework on AAA ("low-accuracy") and AXB ("high-accuracy") dose maps to demonstrate its effectiveness, since these two algorithms are well studied, widely available and can be easily evaluated by other groups. Note that the "low-accuracy" and "high-accuracy" here were defined relatively, since under different context and scenarios, AAA can also be high accuracy (for instance, when compared with pencil-beam convolution) and AXB may be low accuracy when compared with a full-fledged Monte-Carlo package. The "high" and "low" here, thus, were determined relatively between the two algorithms under study. We derived and tested the dose conversion model using a large lung cancer patient database, aiming to improve the dose calculation of lung cancer treatment, whose accuracy is the most susceptible to tissue inhomogeneity.[28]

2 | MATERIALS AND METHODS

2.A | Data preparation

In this study, we retrospectively collected a total of 120 lung cancer patient cases in our institution treated between 06/2017 and 03/

2018. The retrospective study was approved under an institutional review board umbrella protocol. All patients were planned in Eclipse V15.5, by techniques ranging from 3D non-coplanar conformal static beams, intensity-modulated static beams, and 3D conformal arcs to volumetric-modulated arcs. The total prescription doses ranged from 24 to 60 Gy, covering both conventional and stereotactic body radiation therapy treatments. The tumors were distributed across both central and peripheral lung regions. Both primary lung tumors and metastatic tumors from breast, liver, kidney, and prostate were included. The treatment plans used beam energies ranging from 6 MV, 10 MV, 6 MV FFF, to 10 MV FFF. All treatments were designed and successfully delivered on an Elekta VersaHD LINAC with a 160-leaf Agility multi-leaf collimator head.[34] All cases were planned and treated using AAA as the dose calculation engine, with heterogeneity correction turned on. The dose grid was 2.5 mm × 2.5 mm × 2.0 mm in resolution.

For each AAA dose distribution, we calculated the corresponding AXB dose distribution using exactly the same plan. The AXB doses were reported in the form of dose to medium to account for the elemental composition of different tissues. In Eclipse, the tissue designation is based on the densities determined from CT Hounsfield units. The corresponding elemental composition of each tissue is then determined on the basis of the International Commission on Radiological Protection Report 23.[35] The dose grid of AXB was 2.5 mm × 2.5 mm × 2.0 mm in resolution, same as AAA. For each patient, we also exported the planning CT volume. The planning CT volumes were of varying voxel resolutions and volumetric dimensions for different patient cases. We exported the CT and dose files as DICOM-RT files from Eclipse, registered them with DICOM coordinates, and converted them into numeric arrays for the training, validation, and testing purposes. Prior to feeding them into the neural network, we rescaled and interpolated both the AAA and AXB doses, as well as the patient-specific CT volumes to a uniform resolution of 1.37 mm × 1.37 mm × 2.00 mm.

2.B | Network structure selection

For efficient and accurate dose boosting, we employed a Hierarchically Dense U-Net (HD U-Net) structure.[30] Hierarchically Dense U-Net is a combined version of U-Net and DenseNet.[36] Compared with U-Net, HD U-Net uses densely connected layers within each hierarchical level of U-Net, which helps with feature propagation and reuse, and reduces the vanishing gradient issue. Compared with DenseNet, HD U-Net preserves the pooling and upsampling procedures of U-Net, which are able to capture the global features from the input. Once the HD U-Net structure is set, it trains the same way as U-Net. Quantitative comparisons between the three type of networks have been reported and well-documented in a previous publication,[30] showing the advantage of the HD U-Net. As reported, HD U-Net was able to achieve high accuracy with much fewer parameters in the network than U-Net, which reduced the chance of overfitting. In contrast, DenseNet provided the overall worst results due to its lack of ability in capturing global features. For the

supervised training, input channels for the HD U-Net include the Eclipse-calculated AAA dose distribution and the CT volume, and the "ground-truth" output is the Eclipse-calculated AXB dose distribution. For testing, the output will be the boosted dose. We used patch-based training[37] to balance the size of the training data and the computational resources. We separated the full dose volume (512 × 512 × 128) into patches (patch size: 512 × 512 × 16), and feed each of them individually into the network for training/testing. We then merged the output dose maps into a single volume as the final output. The overall training and testing framework is illustrated in Fig. 1.

2.C | Training and testing

Of the 120 paired, patient-specific AAA-AXB dose maps, we randomly selected 72 sets for training, 18 for validation during training, and another 30 for testing. The deep learning model was trained with its hyperparameters tuned using the validation data set. The HD U-Net contained a hierarchy of five levels to reduce the feature size down to 8 × 8 × 1 at the bottom layer with 2 × 2 × 2 interlayer max pooling, to learn both local and global features. Within each layer, the convolutional kernel of size 3 × 3 × 3 was implemented with zero padding to maintain the feature size. On the first half of the U-Net, 16 feature maps (filters) were generated in each convolution step. On the remaining half, the number of feature maps of convolution in each layer, except for the very last convolution step, increased by 16 features from the bottom to the top. The last convolution step generated one channel as the final output. Batch normalization was applied after convolution with rectified linear unit (ReLU) operations. The learning rate was set at 10^{-4} and the Adam algorithm[38] was selected as the optimizer to minimize the loss defined using the mean squared error (MSE). A total of 200 epochs were used for the training in our study, with 100 iterations per epoch. The deep learning model was trained on one NVIDIA Tesla V100 GPU card with 32 GB dedicated memory.

To assess the accuracy of the boosted AAA dose map, we visually evaluated its difference with the AXB dose map directly calculated from Eclipse. MSEs were computed between the boosted AAA and the AXB dose maps to evaluate their differences. 3D gamma analysis[39] based on both 1%/1 mm and 2%/2 mm criteria was also performed to quantitatively assess the match between the boosted AAA and the AXB dose distributions. We also compared the dose volume histograms (DVHs) of the planning target volume (PTV) and lungs between the boosted AAA and the AXB doses to evaluate the accuracy of dose conversion. Quantitative dosimetric endpoints, including the D_{95} and V_{100} of the PTV and V_{20Gy} and D_{mean} of the lungs, were also assessed. The corresponding results between the original AAA doses (prior to boosting) and the AXB doses were also computed for comparison.

In addition to the proposed network with both original AAA dose and CT as input, we also evaluated a second network using only the original AAA dose as input. The second network was evaluated to assess the potential of directly learning intensity, texture, and
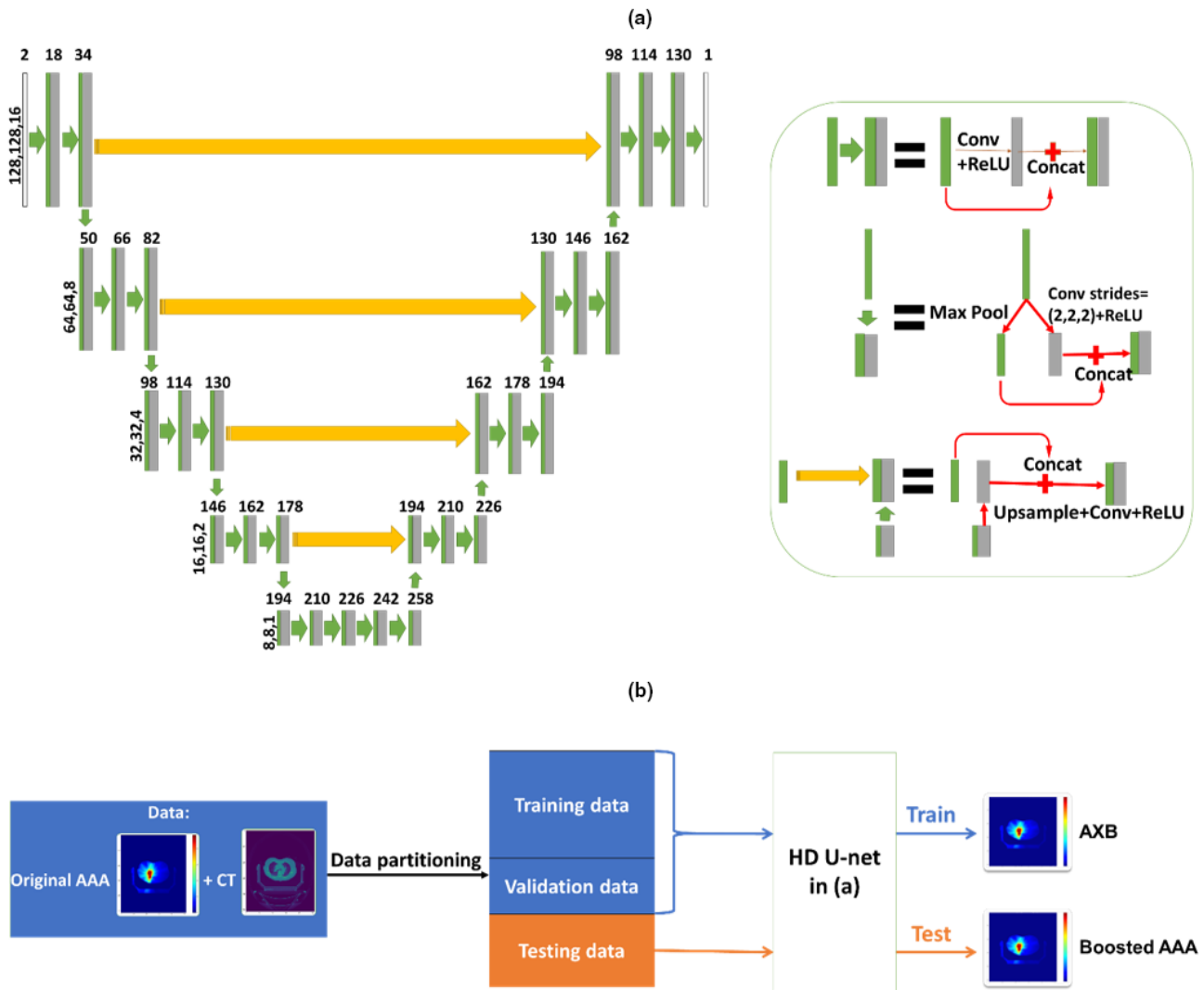
Fɪɢ. 1. (a) General framework of the Hierarchically Dense U-Net (HD U-Net) model for the proposed dose accuracy boosting technique. (b) General training and testing processes where the patient-specific computed tomography (CT) and low-accuracy anisotropic-analytic-algorithm (AAA) doses serve as the input into the HD U-Net structure, and the high-accuracy Acuros XB (AXB) doses serve as the 'ground-truth' output for supervised training/validation. Using the trained framework, a new patient-specific CT and low-accuracy AAA dose can be input to obtain a high-accuracy, boosted AAA dose as the output, with its accuracy matching the AXB dose level.

structural information from the low-accuracy dose maps, to correlate with high-accuracy dose maps for dose boosting. The second network was trained and tested with the same AAA-AXB dose map pairs.

3 | RESULTS

Figure 2 shows the relative differences between the original AAA and the AXB doses [Figs. 2(b) and 2(e)], and between the boosted AAA and the AXB doses [Figs. 2(c) and 2(f)], on three views (coronal, sagittal, and axial). Figures 2(a)–2(c) present a representative non-coplanar static beam plan case, and Figs. 2(d)–2(f) present a volumetric-modulated arc plan case. For both cases, the original AAA dose maps show prominent deviations from the reference AXB doses.

These dose deviations were substantially reduced in the boosted AAA dose maps.

In Fig. 3 we showed the gamma index maps between the original AAA and the AXB dose distributions, and between the boosted AAA and the AXB dose distributions for a 3D non-coplanar static beam plan [Fig. 3(a)] and a volumetric-modulated arc plan [Fig. 3(b)], respectively. A stringent criterion (1%/1 mm) was used to fully demonstrate the differences between dose maps. The red regions in the map indicated failed gamma index (>1). Large dose discrepancies can be observed on the original AAA gamma index maps, especially around the tumor region. As a comparison, the gamma index maps of the boosted AAA dose have these discrepancies largely removed.

Figure 4 compared the DVH curves between the original AAA, the AXB, and the boosted AAA dose distributions, for both PTV and lungs. The DVH curves of the boosted AAA doses matched well with those
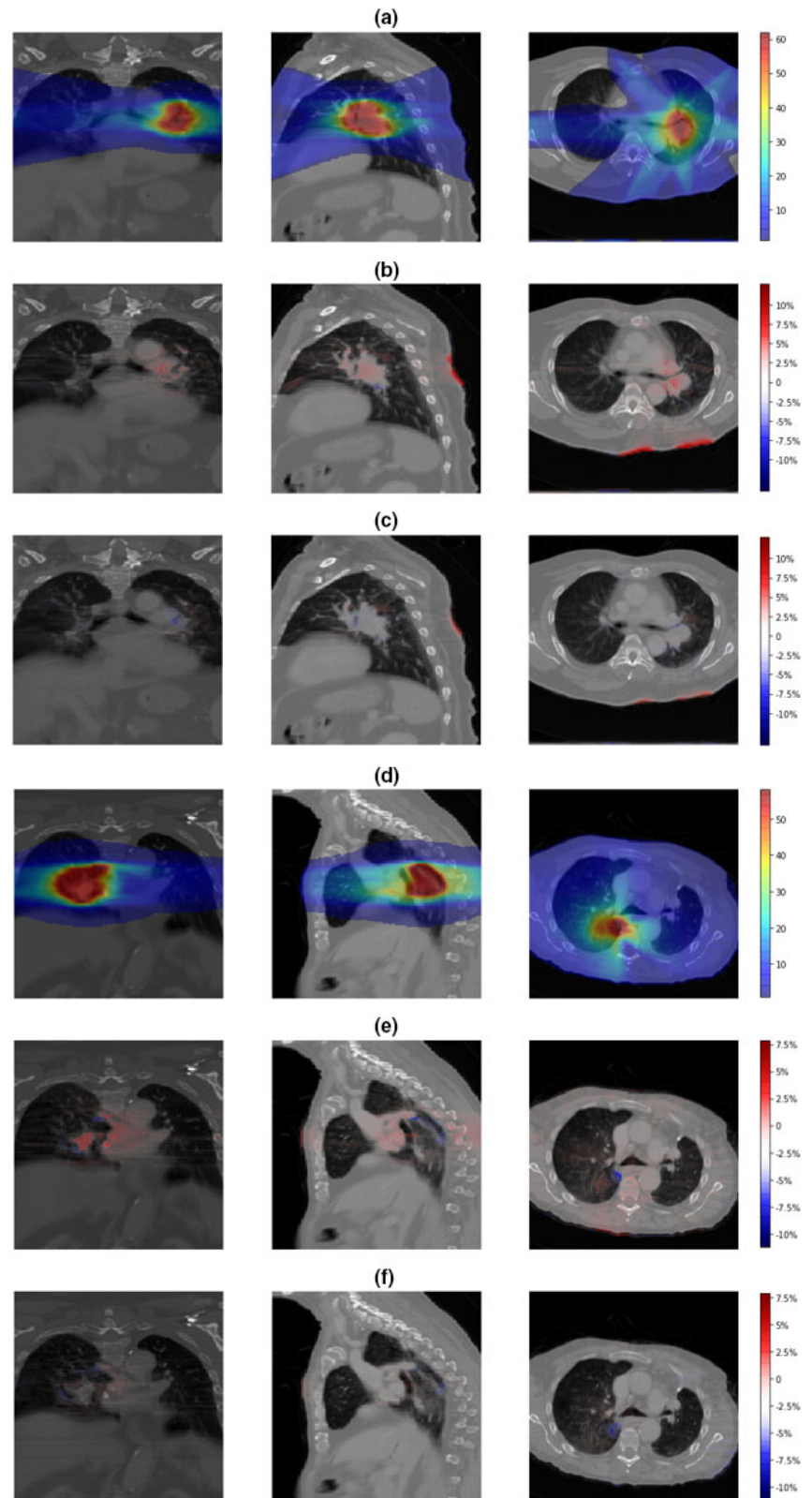
FIG. 2. (a) The "ground-truth" Acuros XB (AXB) dose maps and relative differences between (b) the original anisotropic-analytic-algorithm (AAA) and the AXB dose maps; (c) the boosted AAA and the AXB dose maps, for a three-dimensional non-coplanar static beam plan. (d) The "ground-truth" AXB dose maps and relative differences between (e) the original AAA and the AXB dose maps; (f) the boosted AAA and the AXB dose maps, for a volumetric-modulated arc plan. The "ground-truth" AXB doses were shown in absolute quantities (Gy). The dose differences were normalized to the plan prescription dose (%).

of the AXB doses, while substantial discrepancy could be observed between DVH curves of the original AAA and the AXB doses.

In Table 1, the boosted AAA doses demonstrated substantially improved match to the AXB doses, with average (± s.d.) gamma passing rate (1 mm/1%) 97.6% ± 2.4%, compared to 87.8% ± 9.0%

for the original AAA doses. Using a less strict criterion (2 mm/2%) yielded 99.8% ± 0.4% for the boosted AAA doses, compared to 98.4% ± 1.5% for the original AAA doses. The corresponding average MSE was 0.11 ± 0.05 between the boosted AAA and the AXB doses, compared to 0.31 ± 0.21 between the original AAA and the

AXB doses. The boosted AAA doses (w/o CT as input) were of accuracy in between the original AAA doses and the boosted AAA doses. Note that since this study focuses on developing a network using CT as one of the input, if not specifically mentioned, boosted AAA doses by default refer to those obtained from this network.

In Table 2, dosimetric comparison results were reported in terms of D_{95} of PTV, V_{100} of PTV, D_{mean} of lungs, and V_{20Gy} of lungs. Wilcoxon signed-rank tests with the Bonferroni corrections revealed that the AXB was significantly distinct from the original AAA in D_{mean} of lungs ($P < 10^{-5}$), D_{95} of PTV ($P = 0.001$), and V_{100} of PTV ($P = 0.002$). In contrast, the boosted AAA did not statistically differ from the AXB ($P = 0.094$ for D_{mean} of lungs, $P = 0.642$ for D_{95} of PTV, and $P = 0.417$ for V_{100} of PTV). No significant differences were found for V_{20Gy} of lungs, either for the boosted AAA and the AXB pair, or the original AAA and the AXB pair. Therefore, the boosted AAA matched better with AXB than the original AAA doses, which provided more accurate target coverage and OAR sparing information.

4 | DISCUSSION

Accurate dose calculation and reporting are key to effective and safe radiation therapy. A dose calculation algorithm with both high accuracy and high efficiency is much desired in today's clinic for treatment planning and dose reporting. Our study demonstrated that through a deep learning framework, we could achieve a high-accuracy dose map by boosting from a low-accuracy dose map with patient-specific anatomical CT information, to successfully overcome the trade-off between computational speed and accuracy. The training of our deep learning model takes around ~48 h for 200 epochs. Currently, the inference of our deep learning model takes on average 19 s, with a standard deviation of 3 s. The relatively long inference time is mostly due to the memory limit of our GPU hardware. Under the memory limit, we have to use a patch-based strategy to boost the AAA dose patch by patch, and then merge the results together for a full dose volume. With a larger GPU memory, we can boost the whole AAA dose volume by running the model only once, which
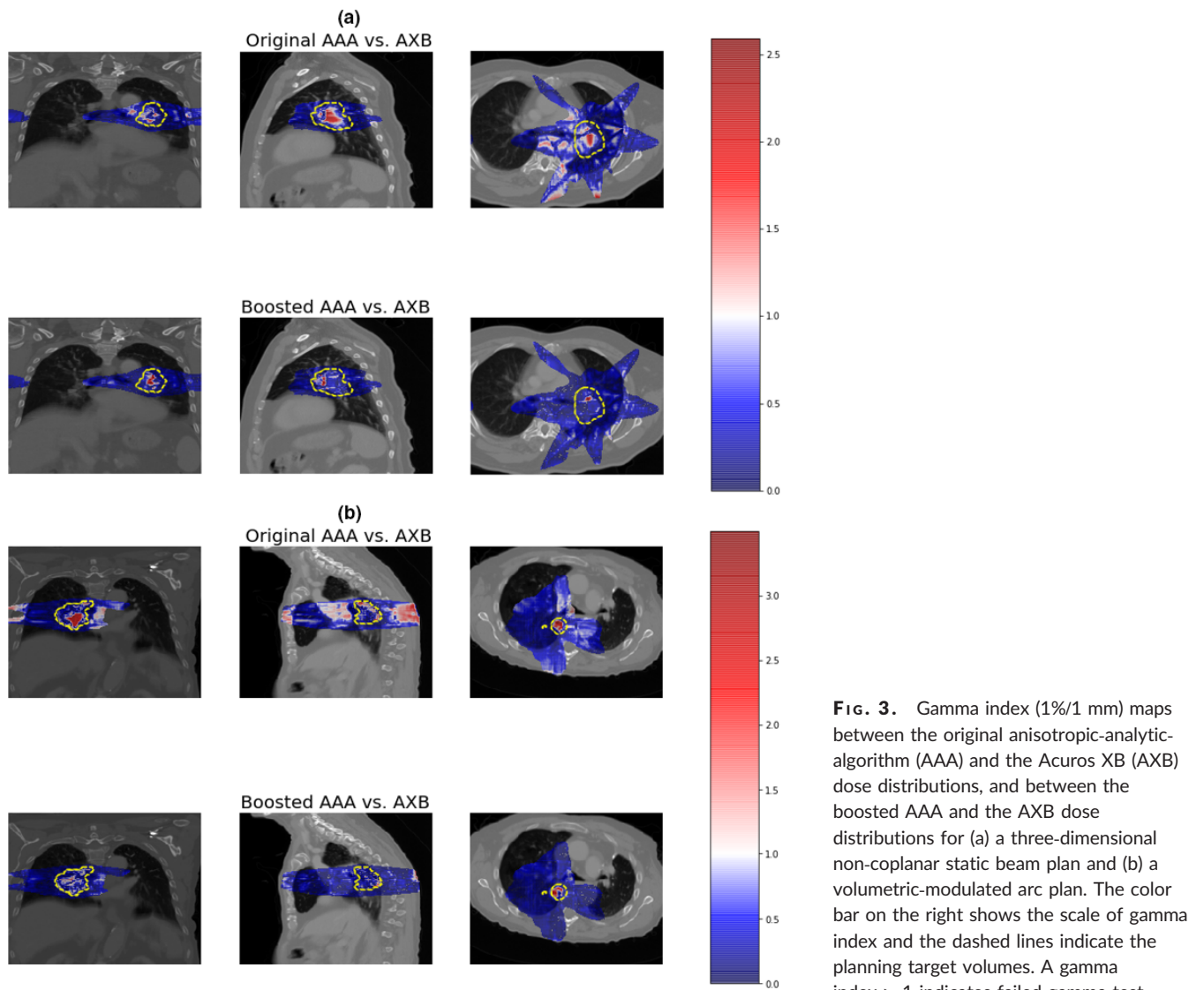

FIG. 3. Gamma index (1%/1 mm) maps between the original anisotropic-analytic-algorithm (AAA) and the Acuros XB (AXB) dose distributions, and between the boosted AAA and the AXB dose distributions for (a) a three-dimensional non-coplanar static beam plan and (b) a volumetric-modulated arc plan. The color bar on the right shows the scale of gamma index and the dashed lines indicate the planning target volumes. A gamma index > 1 indicates failed gamma test.
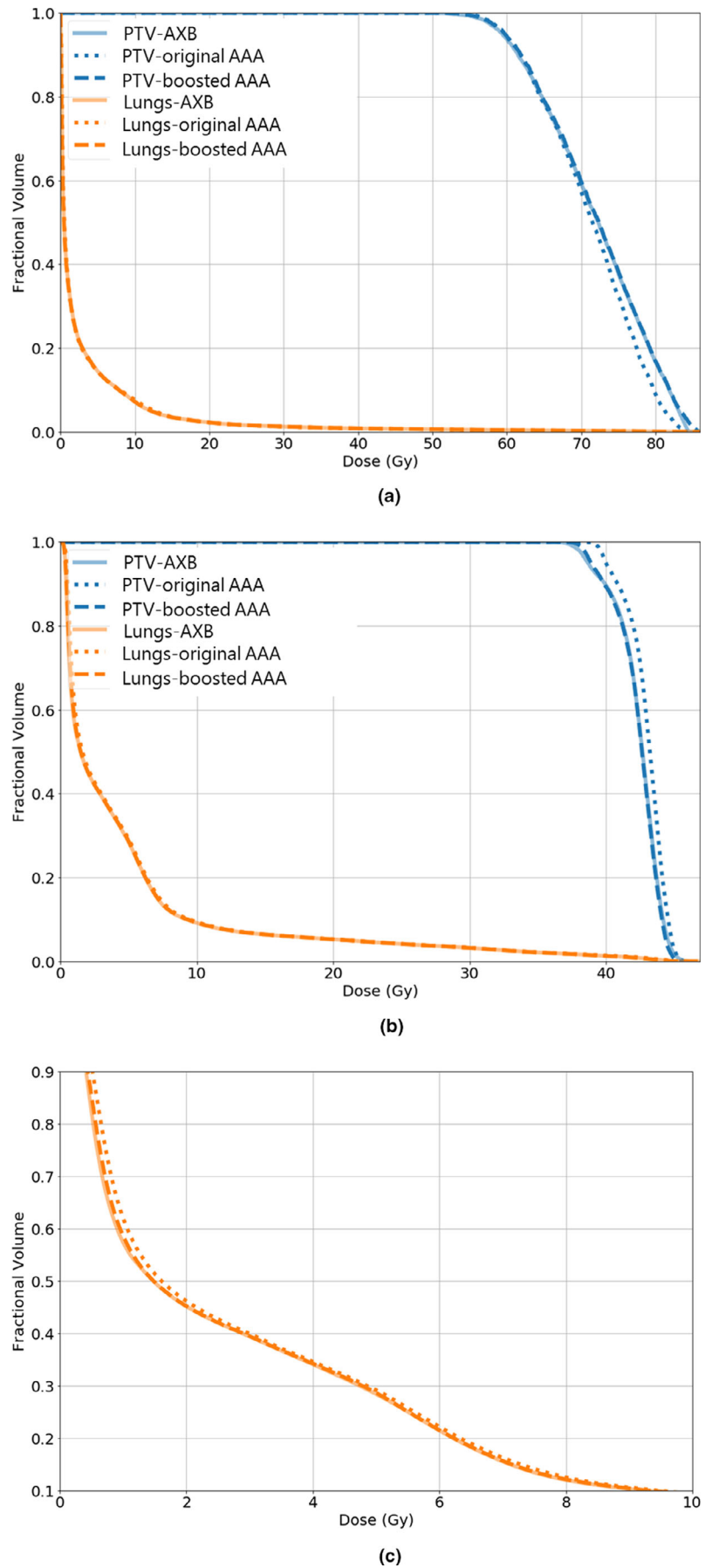
FIG. 4. The dose volume histograms (DVH) curves of planning target volume and lungs for (a) a three-dimensional non-coplanar static beam plan, and (b) a volumetric-modulated arc plan. (c) shows the zoomed-in lung DVH curves for the volumetric-modulated arc plan. The solid, dashed, and dotted lines correspond with the Acuros XB, the boosted anisotropic-analytic-algorithm (AAA), and the original AAA dose maps, respectively

TABLE 1 Quantitative comparisons between the original anisotropic-analytic-algorithm (AAA) and the Acuros XB (AXB) doses, and between the boosted AAA (w/ and w/o computed tomography (CT) as input) and the AXB doses. MSE: mean squared error. RX: prescription. The results of the 30 testing patient cases were included in the analysis.

Dose maps	Gamma passing rates (%)		MSE of voxels with dose > 5% RX dose	% of voxels with dose deviations > 3% RX dose
	2 mm/2%	1 mm/1%		
Original AAA vs AXB	98.4 ± 1.5	87.8 ± 9.0	0.31 ± 0.21	2.01 ± 1.19
Boosted AAA w/o CT vs AXB	99.3 ± 0.7	94.6 ± 2.8	0.15 ± 0.10	0.85 ± 0.53
Boosted AAA vs AXB	99.8 ± 0.4	97.6 ± 2.4	0.11 ± 0.05	0.46 ± 0.46

takes roughly 1–2 s. In addition, in this study we performed dose boosting on an AAA dose volume of $512 \times 512 \times 128$ voxels, while the meaningful dose cloud only occupies a much smaller region within. We may trim the AAA dose volume to remove the irrelevant regions of minimal doses, and apply the deep learning model to the remaining dose cloud only. By this way, we can also remove the need of patch-based model inference and accelerate dose boosting to ~1 s. To put the dose boosting time into context, for our AAA dose and AXB dose calculations, we found the AXB calculation was around 1.5 to 5 times slower than the AAA calculation, depending on the treatment technique and beam arrangement of each plan. In terms of seconds/minutes, AXB can be ~30 s to more than 2 min slower than AAA. In general, AXB tends to be much less efficient for plans with static gantry beams, which are frequently used in our clinic for lung treatments. Our dose boosting scheme, which can potentially be executed within 1 s, will significantly improve the dose calculation efficiency.

In this study, we trained, validated, and tested a voxel-wise dose boosting model on 120 in-house lung cancer patient cases using a Hierarchically Dense U-Net. Visual comparisons of dose differences showed major improvements in the boosted AAA doses, for areas both within the tissues and along tissue interfaces (Fig. 2). In comparison, the original AAA lacks accuracy in calculating doses at multiple regions, where electron densities are changing rapidly and invalidate the kernel scaling approach it applied to account for tissue inhomogeneity. Gamma index distribution maps shown in Fig. 3 also confirmed the improvement of accuracy in boosted AAA doses. The structural-specific DVH curves and dosimetric endpoints also demonstrated that the boosted AAA doses matched well with the AXB doses (Fig. 4; Table 2), and provided more accurate target coverage and OAR sparing information. In general, the conversion model yielded ~98% gamma passing rate between the boosted AAA and the AXB doses for 1%/1mm gamma analysis and ~100% gamma passing rate for 2%/2 mm gamma analysis, showing almost perfect match. In comparison, the corresponding results were only ~88% and ~98% for the original AAA doses (Table 1).

In our developed network, we used both CT images and a low-accuracy dose map as input to derive a high-accuracy dose map. We also evaluated the feasibility of directly using the low-accuracy dose map (w/o CT) to correlate with the high-accuracy dose map for dose boosting (Table 1). It can be observed that the network without using CT as input also helps to boost the AAA dose to match better with the AXB dose than the original AAA dose. However, it is also evident that the boosted AAA dose with CT as input has its accuracy best matched with the AXB dose. With the CT images providing electron density information, the HD U-Net will be better informed of potential inaccuracies in the original AAA dose maps through interpreting the CT density information, to further improve the accuracy of dose boosting.

In current clinical practice, convolution/superposition algorithms like AAA remain the main dose calculation engine for clinical TPS. To improve the dose calculation accuracy, AXB has been introduced into clinics as an alternative to full Monte-Carlo simulation, but its clinical use is still limited. Potential hurdles include the acquisition

TABLE 2 Comparisons between the original AAA, AXB, and boosted AAA doses in terms of D_{95} of PTV, V_{100} of PTV, D_{mean} of lungs, and V_{20Gy} of lungs. The results of the 30 testing patient cases were included in the analysis.

	D_{95} (Gy)			V_{100}		
	Original AAA	AXB	Boosted AAA	Original AAA	AXB	Boosted AAA
PTV						
Mean	45.00	44.43	44.40	94.48%	89.40%	89.71%
s.d.	7.82	7.94	7.84	0.98%	10.73%	8.50%
	D_{mean} (Gy)			V_{20Gy}		
	Original AAA	AXB	Boosted AAA	Original AAA	AXB	Boosted AAA
Lungs						
Mean	4.49	4.42	4.41	5.89%	5.84%	5.85%
s.d.	2.61	2.57	2.58	4.43%	4.37%	4.39%

cost, the lack of resources, the lack of experience to commission a new AXB engine, and, often, the physicians' familiarity and reliance with AAA doses upon which the clinical experience was accumulated. However, it is always highly desirable to use the most accurate dose distribution to correlate with clinical outcomes, in order to derive a reliable and reproducible dose–outcome relationship to benefit and guide future practices. The dose boosting model developed by us can contribute to this goal, as it could be easily adopted by clinics to prospectively/retrospectively convert the original AAA doses into AXB-quality doses. The potential fast-speed conversion (within seconds) introduces minimal interference toward current clinical workflow. The model could be incorporated into the TPS's application programming interface (API) (e.g., ESAPI),[40] to enable direct dose conversion with a one-click solution to allow physicians to evaluate and compare the original AAA and the boosted AAA doses side by side. Currently, our model is built and run on Python, an interpreted, general-purpose programming language. To run our model in Eclipse, one way is to compile our Python code into an executable, and call the executable from within Eclipse using scripts. There are also ongoing developments on a Python interface for ESAPI, which could make an alternative path to run our model within Eclipse. The model could also be run in batch in the background to accumulate boosted AAA data for further dose–clinical outcome analysis. The dose conversion model also does not incur costs associated with proprietary dose calculation algorithms, which can potentially benefit less resourceful, underserved cancer centers. We believe the accuracy and convenience offered by the dose map conversion framework will help physicians to make a more informed decision when evaluating treatment plans, and ultimately benefit the current radiotherapy practice with more accurate dose calculations especially in heterogeneous regions.[41]

In this study, all the evaluated patient data sets are acquired from our institution. The developed model needs to be further evaluated on patient data sets from other institutions to assess its transferability. Although no substantial model adjustments are expected, some slight model tweaking might be needed to accommodate the interinstitutional variations. Transfer learning, an artificial intelligence technique that allows easy model adaptation based on limited new data samples, may help in the case.[28] Similarly, all our data are currently trained on lung cancer patients, other clinical scenarios where the inhomogeneity might affect the dose accuracy, such as larynx and pelvis,[42] also warrant additional model evaluation and potential model fine-tuning.

In addition to our study, there are other groups working on projects to uncover the potential of deep learning in dose accuracy enhancement. A recent work by Ref. [43] uses parallel U-Net branches to boost Monte-Carlo dose calculation accuracy. However, this method is currently limited to boosting doses calculated by a single beamlet, as compared to the full 3D dose map boost achieved by our method. In addition, the network developed by Ref. [43] works by boosting low-accuracy Monte-Carlo doses (fewer events: faster calculation, more noise, and more uncertainty) to high-accuracy Monte-Carlo doses (more events: slower calculation, less noise,

and less uncertainty), essentially an intra-algorithm conversion (between same Monte-Carlo type algorithms). In comparison, our method can boost the doses from one type of algorithm to another (AAA to AXB in our study), allowing interalgorithm conversion. Another study by Ref. [44] is also converting between low-accuracy and high-accuracy Monte-Carlo doses for de-noising, which may not be readily applicable to non-Monte-Carlo–based dose calculation algorithms which are dominant in current clinical TPSs.

In summary, it is shown in this study that with the power of deep learning, we can uncover a mapping scheme between low-accuracy and high-accuracy dose maps, using patient anatomical structure maps and intensity distributions from CT as guidance. The "low" and "high" are defined only relatively in the context of the two algorithms under study, and should not be interpreted in an absolute fashion. We tested this hypothesis through developing and evaluating a dose boosting framework between AAA and AXB dose maps. This framework can be readily extended to other potential pairs of dose maps, the relative accuracy difference of which may be more pronounced. The relative accuracy levels of different algorithms can be determined through the four types of algorithms defined in the AAPM Task Report No. 85,[1] or the "a" to "c" categories stratified by Ref. [45]. It will be of interest to test the framework in boosting pencil-beam convolution doses to Monte-Carlo doses, which might have a positive impact on treatment plan optimization as less-accurate pencil-beam convolution-type calculations are usually used within optimization to promote efficiency. It also remains to be investigated how robust the current framework will be to boost a rudimentary algorithm (such as the correction-based dose calculation algorithm) to a high-accuracy algorithm like Monte Carlo. Additional information, like the treatment plan itself, could be potentially fed into the AI framework to improve the dose boosting accuracy if needed.

5 | CONCLUSIONS

A deep learning-based framework was developed to convert dose maps to boost their accuracy toward the AXB level in a time frame of seconds. The developed method allows dose computation with both high efficiency and high accuracy, potentially benefiting other applications including retrospective dose analysis (especially for heterogeneous regions), fast plan optimization, and secondary dose calculations and plan checks. Future studies involving other dose algorithms, anatomical sites, and treatment centers are warranted to further evaluate the efficacy and robustness of the developed framework.

CONFLICT OF INTEREST

No conflict of interest.

REFERENCES

1. Papanikolaou N, Battista JJ, Boyer A, Kappas C. Tissue inhomogeneity corrections for megavoltage photon beams. *AAPM Task Group*. 2004;65:1–142.
2. Chen WZ, Xiao Y, Li J. Impact of dose calculation algorithm on radiation therapy. *World J Radiol*. 2014;6:874–880.
3. Mukai Y, Matsuyama R, Koike I, et al Outcome of postoperative radiation therapy for cholangiocarcinoma and analysis of dose-volume histogram of remnant liver. *Medicine (Baltimore)*. 2019;98:e16673.
4. Clarkson J. A note on depth doses in fields of irregular shape. *Br J Radiol*. 1941;14:265–268.
5. Cunningham JR. Scatter-air ratios. *Phys Med Biol*. 1972;17:42–51.
6. Khan FM, Levitt SH, Moore VC, Jones TK. Computer and approximation methods of calculating depth dose in irregularly shaped fields. *Radiology*. 1973;106:433–436.
7. Biggs PJ, Doppke KP, Leong JC, Russell MD. Tissue phantom ratios for a Clinac 4/100. *Med Phys*. 1982;9:753–757.
8. Lu L. Dose calculation algorithms in external beam photon radiation therapy. *Int J Cancer Ther Oncol*. 2014;1:01025.
9. Mohan R, Chui C, Lidofsky L. Differential pencil beam dose computation model for photons. *Med Phys*. 1986;13:64–73.
10. Bourland JD, Chaney EL. A finite-size pencil beam model for photon dose calculations in three dimensions. *Med Phys*. 1992;19:1401–1412.
11. Irvine C, Morgan A, Crellin A, Nisbet A, Beange I. The clinical implications of the collapsed cone planning algorithm. *Clin Oncol (R Coll Radiol)*. 2004;16:148–154.
12. Sievinen J, Ulmer W, Kaissl W. *AAA Photon Dose Calculation Model in Eclipse*. Palo Alto (CA): Varian Medical Systems; 2005. Vol. 118:2894.
13. Van Esch A, Tillikainen L, Pyykkonen J, et al Testing of the analytical anisotropic algorithm for photon dose calculation. *Med Phys*. 2006;33:4130–4148.
14. Wulf J, Baier K, Mueller G, Flentje MP. Dose-response in stereotactic irradiation of lung tumors. *Radiother Oncol*. 2005;77:83–87.
15. Partridge M, Ramos M, Sardaro A, Brada M. Dose escalation for non-small cell lung cancer: analysis and modelling of published literature. *Radiother Oncol*. 2011;99:6–11.
16. Ma CM, Li JS, Pawlicki T, et al A Monte Carlo dose calculation tool for radiotherapy treatment planning. *Phys Med Biol*. 2002;47:1671–1689.
17. Bush K, Gagne M, Zavgorodni S, Ansbacher W, Beckham W. Dosimetric validation of Acuros XB with Monte Carlo methods for photon dose calculations. *Med Phys*. 2011;38:2208–2221.
18. Vassiliev ON, Wareing TA, McGhee J, et al Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. *Phys Med Biol*. 2010;55:581–598.
19. Failla GA, Wareing TA, Archambault Y, Thompson SL. *Acuros XB Advanced Dose Calculation for the Eclipse Treatment Planning System*. Palo Alto, CA: Varian Medical Systems; 2010.
20. Kan MW, Yu PK, Leung LH. A review on the use of grid-based Boltzmann equation solvers for dose calculation in external photon beam treatment planning. *Biomed Res Int*. 2013;2013:1–10.
21. Gifford KA, Horton JL, Wareing TA, Failla G, Mourtada F. Comparison of a finite-element multigroup discrete-ordinates code with Monte Carlo for radiotherapy calculations. *Phys Med Biol*. 2006;51:2253–2265.
22. Han T, Mikell JK, Salehpour M, et al Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. *Med Phys*. 2011;38:2651–2664.
23. Rana S, Rogers K. Dosimetric evaluation of Acuros XB dose calculation algorithm with measurements in predicting doses beyond different air gap thickness for smaller and larger field sizes. *J Med Phys*. 2013;38:9–14.
24. Li Y, Rodrigues A, Li T, et al Impact of dose calculation accuracy during optimization on lung IMRT plan quality. *J Appl Clin Med Phys*. 2015;16:5137.
25. Chen X, Ahunbay E, Paulson ES, Chen G, Li XA. A daily end-to-end quality assurance workflow for MR-guided online adaptive radiation therapy on MR-Linac. *J Appl Clin Med Phys*. 2019;21:205–212.
26. Keall PJ, Nguyen DT, O'Brien R, et al Review of real-time 3-dimensional image guided radiation therapy on standard-equipped Cancer radiation therapy systems: are we at the tipping point for the era of real-time radiation therapy? *Int J Radiat Oncol Biol Phys*. 2018;102:922–931.
27. Nguyen D, Long T, Jia X, Lu W. Dose Prediction with U-net: A Feasibility Study for Predicting Dose Distributions from Contours using Deep Learning on Prostate IMRT Patients. arXiv preprint arXiv:1709.09233, 2017.
28. Zhen X, Chen J, Zhong Z et al Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. *Phys Med Biol*. 2017;62:8246–8263.
29. Thompson RF, Valdes G, Fuller CD, et al Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? *Radiother Oncol*. 2018;129:421–426.
30. Nguyen D, Jia X, Sher D et al Three-Dimensional Radiotherapy Dose Prediction on Head and Neck Cancer Patients with a Hierarchically Densely Connected U-net Deep Learning Architecture. arXiv preprint arXiv:1805.10397; 2018.
31. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology*. 2018;288:177–185.
32. Zhang Y, Chen L, Li B et al Incorporating biomechanical modeling and deep learning into a deformation-driven liver CBCT reconstruction technique. In: Medical Imaging 2019: Physics of Medical Imaging. 2019. International Society for Optics and Photonics.
33. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer; 2015.
34. Zhang Y et al Modeling Elekta VersaHD using the Varian Eclipse treatment planning system for photon beams: A single-institution experience. *J Appl Clin Med Phys*. 2019;20:33–42.
35. Snyder W. Report of the task group on reference man. International Commission on Radiological Protection no. 23. Pergamon Press, Oxford22.
36. Huang G, Liu Z, van der Maaten L, et al Densely connected convolutional networks. in *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017.
37. Sharma A, Liu X, Yang X, Shi D. A patch-based convolutional neural network for remote sensing image classification. *Neural Netw*. 2017;95:19–28.

38. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980; 2014.

39. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25:656–661.

40. Janes LA, Anderson CJR, Matuszak MM, Roberts DA, Moran JM, Younge KC. SU-E-T-779: use of the Varian eclipse scripting API to collect volumetric modulated arc therapy plan data. *Med Phys*. 2015;42:3516.

41. Fogliata A, Nicolini G, Clivio A, Vanetti E, Cozzi L. Critical appraisal of Acuros XB and anisotropic analytic algorithm dose calculation in advanced non-small-cell lung cancer treatments. *Int J Radiat Oncol Biol Phys*. 2012;83:1587–1595.

42. Schwartz DL, Sosa A, Chun SG, et al SBRT for early-stage glottic larynx cancer-Initial clinical outcomes from a phase I clinical trial. *PLoS One*. 2017;12:e0172055.

43. Neph R, Huang Y, Yang Y, Sheng K. DeepMCDose: A Deep Learning Method for Efficient Monte Carlo Beamlet Dose Calculation by Predictive Denoising in MR-Guided Radiotherapy. in *Workshop on Artificial Intelligence in Radiation Therapy*; 2019. Springer.

44. Peng Z, Shan H, Liu T et al Deep learning for accelerating Monte Carlo radiation transport simulation in intensity-modulated radiation therapy. arXiv preprint arXiv:1910.07735; 2019.

45. Ojala JJ, Kapanen MK, Hyödynmaa SJ, Wigren TK, Pitkänen MA. Performance of dose calculation algorithms from three generations in lung SBRT: comparison with full Monte Carlo-based dose distributions. *J Appl Clin Med Phys*. 2014;15:4–18.