Physics and Imaging in Radiation Oncology 17 (2021) 65–70

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Development and evaluation of radiotherapy deep learning dose prediction models for breast cancer

Nienke Bakx, Hanneke Bluemink, Els Hagelaar, Maurice van der Sangen, Jacqueline Theuws, Coen Hurkmans *

Catharina Hospital, Radiation Oncology, Eindhoven, The Netherlands

ARTICLE INFO

Keywords:
Breast cancer
Convolutional neural networks
Dose prediction
Intensity-modulated radiotherapy
Machine learning

ABSTRACT

Background and purpose: Treatment planning of radiotherapy is a time-consuming and planner dependent process that can be automated by dose prediction models. The purpose of this study was to evaluate the performance of two machine learning models for breast cancer radiotherapy before possible clinical implementation.
Materials and methods: An in-house developed model, based on U-net architecture, and a contextual atlas regression forest (cARF) model integrated in the treatment planning software were trained. Obtained dose distributions were mimicked to create clinically deliverable plans. For training and validation, 90 patients were used, 15 patients were used for testing. Treatment plans were scored on predefined evaluation criteria and percent errors with respect to clinical dose were calculated for doses to planning target volume (PTV) and organs at risk (OARs).
Results: The U-net plans before mimicking met all criteria for all patients, both models failed one evaluation criterion in three patients after mimicking. No significant differences ($p < 0.05$) were found between clinical and predicted U-net plans before mimicking. Doses to OARs in plans of both models differed significantly from clinical plans, but no clinically relevant differences were found. After mimicking, both models had a mean percent error within 1.5% for the average dose to PTV and OARs. The mean errors for maximum doses were higher, within 6.6%.
Conclusions: Differences between predicted doses to OARs of the models were small when compared to clinical plans, and not found to be clinically relevant. Both models show potential in automated treatment planning for breast cancer.

1. Introduction

Breast cancer is the most common type of cancer in women in Europe, with 523,000 new cases diagnosed in 2018 [1]. Depending on the stage of the disease, different treatments are recommended. Breast-conserving surgery (BCS) followed by radiotherapy (RT) to the breast is a widely accepted local treatment for early breast cancer. RT after BCS for node-negative cancer compared to BCS only, halves the risk on recurrence after 10 years, while the breast cancer death rate is reduced by about a sixth [2]. However, this needs to be balanced against the dose delivered to healthy tissue which can lead to different side effects. For major coronary events, Darby et al. found an excess relative risk (ERR) of 7.4% per Gy mean heart dose (MHD) [3], whereas Taylor et al. found an ERR of 4.1% per Gy MHD for cardiac mortality alone and an ERR of 11% per Gy mean lung dose (MLD) for mortality resulting from RT induced lung cancer [4]. In order to minimise side effects as much as possible, low doses to the heart and lung should be pursued.

To create a treatment plan with acceptable doses to organs at risk (OARs) while still achieving adequate coverage of the target tissue, planners go through an iterative plan optimisation process. This process can be time-consuming, and the plan quality is dependent on the experience of the planner [5,6]. Recent years, several studies have been conducted on the development of dose prediction methods to automate this process, by using convolutional neural networks (CNNs) [7–11]. The treatment planning system (TPS) used in this study also implemented machine learning (ML) to offer automated planning. This module was already tested for head and neck RT and showed promising results [12]. However, most of the aforementioned studies on dose prediction focus on the treatment sites prostate or head and neck, and little is known about the performance of these models for breast cancer.

* Corresponding author.
E-mail address: coen.hurkmans@catharinaziekenhuis.nl (C. Hurkmans).

https://doi.org/10.1016/j.phro.2021.01.006
Received 28 August 2020; Received in revised form 21 January 2021; Accepted 21 January 2021
Available online 30 January 2021

Therefore, the purpose of this study was to evaluate the performance of two ML models for whole breast RT with regard to plan quality. The first model was an in-house developed model using CNNs, based on the often used U-net structure [13]. The second model was the implemented model of the TPS. Results of these methods were compared with clinical deliverable plans. The potential of these models for future clinical implementation was thereby explored, aiming to reduce planner dependence and time.

2. Materials and methods

2.1. Patient database

Treatment plans for 105 patients diagnosed with left-sided node-negative breast cancer were included. Ethical approval was granted by the local ethics committee. This research is conducted on anonymised patient data and according to Dutch law this falls under the so-called non-WMO legislation (medical research law waiver). The patients were treated in 15 fractions with a prescribed total dose of 40.05 Gy, with a beam energy of either 6 or 10 MV. A tangential Intensity-modulated radiation therapy (IMRT) technique was used with one open segment for each tangential beam and a maximum of eight total segments, which should be at least 9 cm^2. On average, 336 monitor units (MU) (range 304–409) were needed to deliver one fraction. Computed Tomography (CT) data and clinical treatment plans were available for all patients. The resolution of the CT data was 2.5 × 2.5 × 2.5 mm^3. RayStation TPS was used to create new clinically deliverable plans using the standard approach, described below. Version 9B was used to generate new plans with the ML module. Furthermore, predicted dose distributions resulting from the in-house developed method were loaded into RayStation 9B and dose mimicking was applied to generate clinically deliverable plans. A dose calculation grid spacing of 3 mm was used for all plans.

The breast planning target volume (PTV) was generated by expanding the clinical target volume (CTV) by 5 mm and then cropping it 5 mm under the skin. During optimisation, beam segments were opened outside the skin region to account for setup errors, intra-fraction movements and possible increase of breast volume during treatment. For this study, a distance of 1–2 cm from the skin was used. Opening of the segments was executed in the last part of optimisation. The automatic beam angle function available in the 3D-CRT module of RayStation was used to determine the optimal beam angles.

Predefined clinical goals were used to evaluate the quality of a treatment plan, assuring an adequate dose coverage of the PTV and sparing of the OARs. These evaluation criteria were based on the Dutch national consensus criteria for breast treatment planning from October 2019 and summarized in Supplementary Table S1.

2.2. Machine learning models and model training

Two ML models were used for dose prediction. The first model was in-house developed, based on the U-net architecture, a widely used method in biomedical image processing [13]. For this study, the adapted U-net architecture of Nguyen et al. [8] was used. See the Supplementary material for a more in depth overview of the model.

The second model was developed and integrated by RaySearch, based on a contextual atlas regression forest (cARF) [12], consisting of two phases. In the atlas-to-image mapping phase, atlas regression forests (ARFs) are trained to model the relationship between image features and the dose distribution. In the atlas-selection phase, density estimation over observed image features is used to train a model to select the most relevant ARF for a new patient. After these phases, a conditional random field model is used to find the most probable dose distribution. The full method was described by McIntosh and Purdie [14,15].

For training of both models, treatment plans of 90 patients were used, with a training set containing plans of 72 patients and a validation set with plans of 18 patients. To independently test the model after training, 15 additional patients were used. The composition of these datasets can be found in Table 1.

The in-house developed U-net model was trained for 800 epochs, and no cross-validation was performed. Each batch contained three patients, of which eight slices were re-selected at each epoch using a sampling scheme. Central slices are more important for dose prediction as they always contain the PTV, so a Gaussian scheme was used with a standard deviation equal to one-third of the distance from the central to the end slice.

2.3. Dose mimicking

The two models both result in a dose prediction per voxel, which are not directly clinical applicable. To obtain clinically deliverable plans, dose mimicking can be used. In this study, mimicking was realized by direct machine parameter optimization to approximate the predicted dose distribution, based on the method of Petersson et al. [16]. Three intermediate collapsed cone convolution dose calculations were performed over the course of multiple dose mimicking iterations, ending with a final collapsed cone convolution dose calculation resulting in the final dose distribution. The beam energy and gantry angles were copied from the clinical plan of the corresponding patient. The same segmentation settings, mentioned in Section 2.1, were used for all plans. In addition, for all patients the same set of objectives and weights was used. After mimicking, created segments were visually inspected. If needed, leafs were opened manually outside the skin region, at similar distances from the skin as the clinical plans.

2.4. Model evaluation

For the U-net model, the predicted dose distributions before and after mimicking were evaluated. Of the cARF model predictions, only mimicked dose distributions were available. To compare the predictions, dose distributions of the clinical, predicted and mimicked plans were scaled such that 98% of the PTV volume receives 95% of the dose (38.05 Gy). The voxel-wise predictions were evaluated with the clinical goals, based on the national consensus aforementioned. As all plans were scaled to have a dose of 38.05 Gy at 98% volume, this goal was not scored. In addition to the average and maximum doses of the regions of interest (ROIs), the differences of the mean average dose and maximum dose to the PTV with respect to the prescribed dose (40.05 Gy) were calculated. The average dose and maximum dose (D2%) to the ROIs were then further investigated.

The performance of the models was further evaluated by a percent error relative to the prescribed dose, calculated as $\frac{\text{predicted dose} - \text{clinical dose}}{\text{prescribed dose}} \times 100\%$, with a prescribed dose of 40.05 Gy. Significance between the average and maximum doses of the clinical plan and the predicted and mimicked plans of each of the models were investigated with the Wilcoxon signed rank test. A p-value <0.05 was considered significant.

3. Results

The clinical plans and the U-net predicted plans met all clinical goals

Table 1
Characteristics of the patient sets used for training and testing of the two models.

Model Stage	Patient Set	Number of Patients	Beam Energy (number of patients)		PTV Volume (mean and range; cm^3)
			6 MV	10 MV	
Training	Training	72	62	10	818 (196–2864)
	Validation	18	11	7	879 (299–1168)
Testing	Test	15	10	5	831 (458–1811)

Table 2
Number of patients of the test set (n = 15) achieving the clinical goals, for the clinical plans, predicted (pred) and mimicked (mim) plans of the U-net model and conditional Atlas Regression Forest (cARF) plans. Numbers printed in italic indicate which goals are not achieved for all patients.

Clinical Goal	Clinical	U-net		cARF
		Predicted	Mimicked	
PTV: $D_{avg} \geq 40.05$ Gy	15	15	15	14
PTV: $D_{2\%} \leq 42.85$ Gy	15	15	12	15
Heart: $D_{avg} \leq 3$ Gy	15	15	15	15
Lungs: $D_{avg} \leq 6$ Gy	15	15	15	15
External – PTV: V107% ≤ 10.00 cm^3	15	15	15	14
Heart: $D_{avg} \leq 2$ Gy	15	15	15	14
Lungs: $D_{avg} \leq 4$ Gy	15	15	15	15

(Table 2). However, after mimicking, three patient plans exceeded the allowed volume of 2% receiving >42.85 Gy, with volumes of 2.6%, 2.9% and 3.0%. For the cARF model, three different patient cases also failed to meet all clinical goals. The first patient case failing a clinical goal had an average PTV dose of 39.96 Gy. This dose was still within 1% of the prescribed dose, which is clinical acceptable according to the guidelines of the national consensus. The second case had a volume of 12.6 cm^3 of the external receiving >107% of the prescribed dose. The last case had an average heart dose of 2.15 Gy, thereby exceeding the desired limit of an average dose of 2 Gy, but still achieving an average dose below 3 Gy.

Average and maximum doses of the ROIs in all plans are summarized in Table 3. For the U-net model, both average and maximum doses increased after mimicking for all ROIs, thereby significantly exceeding the clinical doses. The average PTV dose of the cARF was lower than the clinical dose, but both heart and lung doses significantly exceeded the clinical doses.

Furthermore, the average doses for all models fell within 2.0% of the prescribed dose. For the maximum dose, the means of the different plans differed between 5.0% and 6.0% from the prescribed dose.

The percent errors on the average and maximum dose are shown in Fig. 1. The U-net model predicted a dose distribution which did not differ significantly from the clinical dose distribution, but after mimicking it did for all ROIs. The cARF model predicted a significantly different dose for the heart and lungs, compared to the clinical plans. In addition, the differences between the mimicked U-net predictions and cARF predictions were also significant for all ROIs, except for the maximum dose to the heart. cARF predicted lower average and maximum doses to the PTV, while the U-net predicted lower average and maximum doses to the OARs. Overall, the percent error was lower for the average doses compared to the maximum doses. In addition, the spread of the error was decreased after mimicking. Both models tend to over-predict the doses for the OARs, while predicting a lower dose to PTV, although this dose increases after mimicking.

The percent errors on average doses (mean ± SD) to PTV, heart and lungs, were 0.7 ± 0.2%, 0.1 ± 0.1% and 0.2 ± 0.5% for the mimicked U-net model predictions. For the cARF model, these errors were −0.3 ± 0.6%, 0.3 ± 0.3% and 0.5 ± 0.2%. For the maximum doses to PTV, heart and lungs, errors of 0.9 ± 0.5%, 1.2 ± 1.6% and 3.3 ± 2.0% were found for the mimicked U-net predictions. The cARF model resulted in errors to these ROIs of 0.04% ± 1.0, 3.8% ± 6.0 and 6.6% ± 3.3.

Fig. 2 shows two axial slices of two typical patient cases. For both patients, the U-net predicted a higher dose for PTV, resulting in more high dose regions than the clinical and cARF plans.

4. Discussion

Two ML models to predict dose distributions were evaluated and compared to clinical plans. Both models resulted in average doses to ROIs within a range of 1.0% to 1.5% deviation compared to clinical plans. Maximum doses to heart and lungs deviated more, with range of 6.6% and 3.3% error. In our opinion, the found differences between the models were not clinically relevant, since the clinical accepted average doses were not exceeded. Furthermore, the plans were also physically acceptable, with similar segment shape and MU per fraction. These results indicate comparable performance of both models, showing the capability of creating clinically acceptable plans.

There are a few studies which used other models than CNN models for dose prediction in breast cancer patients. Fan et al. [17] predicted dose volume histograms (DVHs) using kernel density estimation, which were then transferred to objectives. With these objectives, treatment plans were created with the auto-planning module of the TPS Pinnacle. No significant differences were found when comparing the automatic generated plans with manually generated plans by experienced planners. In another study, van Duren-Koopman et al. [18] created an automatic workflow, using the RapidPlan knowledge-based planning of the Eclipse TPS to generate treatment plans for breast irradiation, including locoregional lymph nodes. For 14 out of the 15 test patients, the automated plans were preferred or equal to the manually generated plans.

This is to our knowledge the first study focusing on dose prediction for breast cancer, with the use of CNNs. The study of Nguyen et al. [8], where the U-net model architecture was based on, applied it to treatment plans for prostate cancer. They report mean absolute percent errors on average doses of 1.1% for PTV to a maximum of 4.2% for an OAR. The mean absolute percent errors for maximum doses range from 1.8% for PTV to 5.1% for an OAR. The dose distributions in the study of Nguyen et al. were not mimicked, meaning that comparison to the U-net predictions would be more fair. The results in this study showed no significant different errors. So although direct comparison is not possible due to different treatment sites, it can be stated that the model from this study is comparable in performance or even outperforms the model of Nguyen et al. The study of McIntosh et al. [12] is the only one evaluating the cARF model. They predicted head and neck dose distributions, but no predicted average and maximum dose values were reported, so no comparisons can be made.

Furthermore, several studies used a three dimensional (3D) U-net architecture for predicting dose distributions, sometimes combined with

Table 3
Average and maximum doses in Gy to ROIs for the clinical plans, predicted and mimicked plans of the U-net model and mimicked plans generated by the cARF model (mean ± standard deviation). For PTV, the difference between mean average and maximum dose with respect to the prescribed dose (40.05 Gy) is shown. Doses differing significantly from clinical doses are indicated with an asterisk.

		PTV		Heart	Lungs
		Dose [Gy]	Difference with respect to prescribed dose [%]	Dose [Gy]	Dose [Gy]
Clinical	Average	40.5 ± 0.2	+1.2	1.0 ± 0.4	1.9 ± 0.5
	Maximum	42.1 ± 0.3	+5.0	4.4 ± 3.4	26.2 ± 4.7
U-net-predicted	Average	40.5 ± 0.1	+1.1	1.0 ± 0.3	2.0 ± 0.4
	Maximum	42.2 ± 0.2	+5.4	4.9 ± 3.1	36.8 ± 3.9
U-net-mimicked	Average	40.8 ± 0.3*	+1.9	1.0 ± 0.4*	2.0 ± 0.5*
	Maximum	42.4 ± 0.4*	+5.9	4.9 ± 3.8*	27.5 ± 4.3*
cARF	Average	40.4 ± 0.2	+1.0	1.1 ± 0.5*	2.1 ± 0.5*
	Maximum	42.1 ± 0.4	+5.0	5.9 ± 6.1*	28.8 ± 3.7*

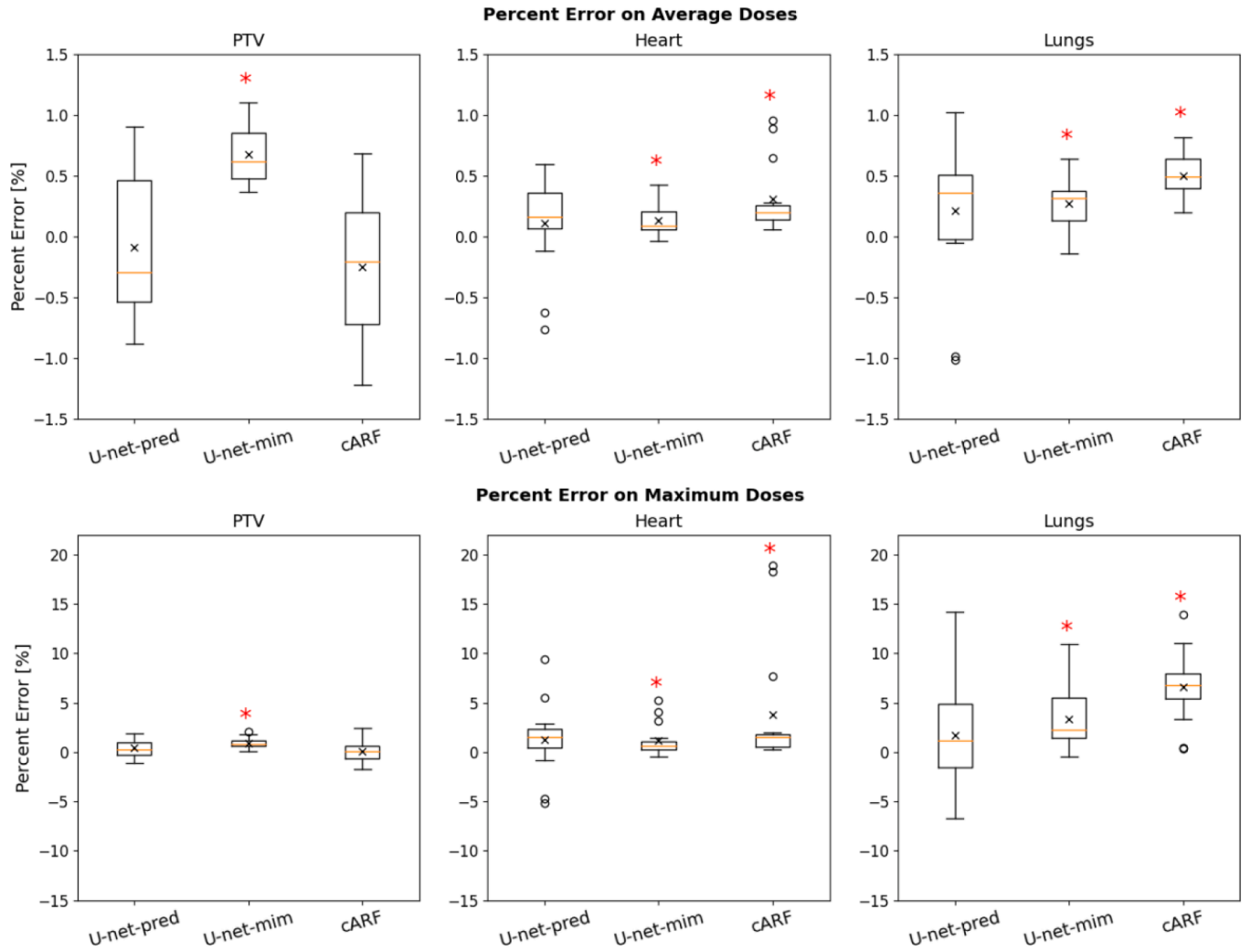

Fig. 1. Percent error of predicted and mimicked dose distributions of the U-net model and mimicked does distribution of the cARF model, when compared to clinical plans, with respect to the prescribed doses. Average (upper row) and maximum (lower row) doses of different ROIs are evaluated. Horizontal lines in boxes are medians, crosses are means, dots are outliers. Statistically significant differences with respect to clinical dose are marked with a red asterisk ($p < 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

other well-known CNN architectures such as DenseNet [7], or extended with beam configuration information [9,19]. Both studies that included beam configuration, report an improvement in prediction compared to only using anatomical information. Barragán et al. [9] predicted dose distributions for lung IMRT patients and reported a decrease of the error of the mean dose, averaged over all ROIs, from 2.3% ± 2.0 to 1.4% ± 1.3, for a model without and with beam configuration information, respectively. The error of the maximum dose decreased from 4.0% ± 4.7 to 2.9% ± 3.1. Similar to the 2D U-net study, dose distributions in these studies were not mimicked to clinical deliverable plans. Although direct comparison is again difficult, the prediction errors of the in-house developed 2D U-net were in the same range or even better as the 3D U-net without any beam information. The 3D U-net including beam configuration seemed to outperform the mimicked U-net and cARF predictions with regard to maximum doses.

The patient set used in this study contains patients treated with a beam energy of 6 or 10 MV, where the latter is used for patients with larger target volumes. By using a large dataset with a variety of patient anatomies and target volumes, it was assumed that there would be no difference in the results of 6 or 10 MV predictions. During evaluation, small differences between predictions for 6 and 10 MV plans were observed in the percent errors on the average dose to PTV of 0.23% and 0.12% for the mimicked U-net and cARF models, respectively. However, due to a small test set and no proof of significance, these differences were assumed to be negligible.

For dose mimicking, initial set-up settings were used for the two models. No adjustments to further improve mimicking were made. In addition, the scaling could result in more hot dose regions and could also be prevented by specific mimicking settings. Future research could involve improving these settings.

During mimicking several settings were copied from clinical plans, such as the beam angle. To further automate the planning process and be less planner dependent, a method could be developed to automatically determine beam angles, fitting the predicted dose distribution.

The biggest limitation of the U-net model used in this study was its 2D nature. Due to a lack of computational power and time, a 3D U-net was not feasible in this study. In the 2D U-net, slices were predicted independent of each other, disregarding any spatial information in the slice direction. Nevertheless, this spatial information was restored with dose mimicking, leading to deliverable plans. However, mimicking lead in almost all cases to higher average and maximum doses to the ROIs, while the dose without mimicking did not differ significantly from the clinical dose. Future research could include a 3D U-net to test its ability to predict dose distributions, compared to the 2D U-net.

Resulting plans from both models did not show clinically relevant differences for studied dose parameters. To further investigate potential

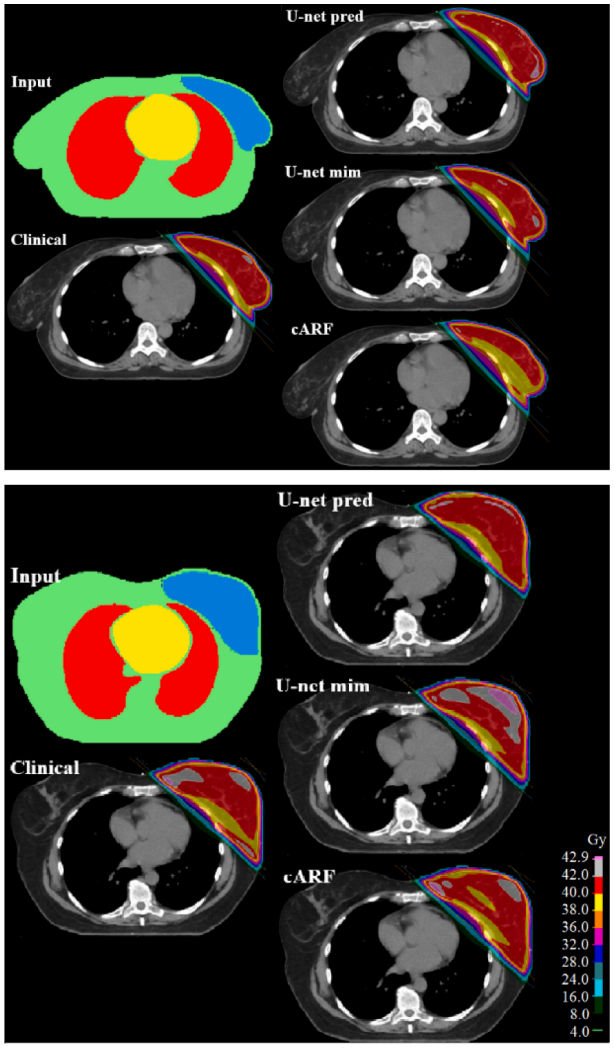

Fig. 2. Two axial slices of two different typical patient cases, with the corresponding input, consisting of 4 different masks for PTV, heart, lungs and external of the body. For the U-net model, corresponding predicted and mimicked dose distributions are shown, as is the mimicked dose distribution form the cARF model.

of clinical implementation, a new study will be performed which includes scoring by radiation oncologists of plans generated in clinic and by the models.

In this study, a U-net model was tested to predict dose distributions for whole breast RT. These dose predictions were mimicked to create clinical deliverable plans and evaluated together with the predictions of an already existing commercially available cARF model and compared with corresponding clinical plans. Differences between the two models in predicted doses to OARs compared to the clinical plans were small and not found to be clinically relevant. Results of both models are promising for automatic plan generation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Fredrik Löfman, Mats Holmström and Hanna Gruselius from the machine learning department of RaySearch for their contribution in preparing the data, training of the cARF model and integration of different steps into RayStation. Furthermore, we would like to thank Josien Pluim from the Technical University of Eindhoven for providing capacity to train the models. Lastly we would like to thank Anke Habraken, Fanny van Aarle, Jorien van der Leer, Simone van Barneveld and Therese van Nunen for assisting in creating the clinical treatment plans.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2021.01.006.

References

[1] Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 2018;103:356–87. https://doi.org/10.1002/mp.13597.

[2] Early Breast Cancer Trialists' Collaborative Group. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. Lancet 2011;378:1707–16. https://doi.org/10.1016/S0140-6736(11)61629-2.

[3] Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med 2013;368:987–98. https://doi.org/10.1056/NEJMoa1209825.

[4] Taylor C, Correa C, Duane FK, Aznar MC, Anderson SJ, Bergh J, et al. Estimating the risks of breast cancer radiotherapy: evidence from modern radiation doses to the lungs and heart and from previous randomized trials. J Clin Oncol 2017;35:1641–9. https://doi.org/10.1200/JCO.2016.72.0722.

[5] Batumalai V, Jameson MG, Forstner DF, Vial P, Holloway LC. How important is dosimetrist experience for intensity modulated radiation therapy? A comparative analysis of a head and neck case. Pract Radiat Oncol 2013;3:99–106. https://doi.org/10.1016/j.prro.2012.06.009.

[6] Berry SL, Boczkowski A, Ma R, Mechalakos J, Hunt M. Interobserver variability in radiation therapy plan output: results of a single-institution study. Pract Radiat Oncol 2016;6:442–9. https://doi.org/10.1016/j.prro.2016.04.005.

[7] Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, Jiang S. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture. Phys Med Biol 2018;64:065020. https://doi.org/10.1088/1361-6560/ab039b.

[8] Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, Jiang S. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. Sci Rep 2019;9:1–10. https://doi.org/10.1038/s41598-018-37741-x.

[9] Barragán-Montero AM, Nguyen D, Lu W, Lin M-H, Norouzi-Kandalan R, Geets X, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. Med Phys 2019;46:3679–91. https://doi.org/10.1002/mp.13597.

[10] Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: a volumetric dose prediction algorithm using 3D fully-convolutional neural networks. Phys Med Biol 2018;63:235022. https://doi.org/10.1088/1361-6560/aaef74.

[11] Kajikawa T, Kadoya N, Ito K, Takayama Y, Chiba T, Tomori S, et al. A convolutional neural network approach for IMRT dose distribution prediction in prostate cancer patients. J Radiat Res 2019;60:685–93. https://doi.org/10.1093/jrr/rrz051.

[12] McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Phys Med Biol 2017;62:5926–44. https://doi.org/10.1088/1361-6560/aa71f8.

[13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. MICCAI 2015:234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

[14] McIntosh C, Purdie TG. Contextual atlas regression forests: multiple-atlas-based automated dose prediction in radiation therapy. IEEE Trans Med Imaging 2016;35:1000–12. https://doi.org/10.1109/TMI.2015.2505188.

[15] McIntosh C, Purdie TG. Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning. Phys Med Biol 2017;62:415–31. https://doi.org/10.1088/1361-6560/62/2/415.

[16] Petersson K, Nilsson P, Engström P, Knöös T, Ceberg C. Evaluation of dual-arc VMAT radiotherapy treatment plans automatically generated via dose mimicking. Acta Oncol 2016;55:523–5. https://doi.org/10.3109/0284186X.2015.1080855.

[17] Fan J, Wang J, Zhang Z, Hu W. Iterative dataset optimization in automated planning: Implementation for breast and rectal cancer radiotherapy. Med Phys 2017;44:2515–31. https://doi.org/10.1002/mp.12232.

[18] van Duren-Koopman MJ, Tol JP, Dahele M, Bucko E, Meijnen P, Slotman BJ, et al. Personalized automated treatment planning for breast plus locoregional lymph nodes using Hybrid RapidArc. Pract Radiat Oncol 2018;8:332–41. https://doi.org/10.1016/j.prro.2018.03.008.

[19] Zhou J, Peng Z, Song Y, Chang Y, Pei X, Sheng L, et al. A method of using deep learning to predict three-dimensional dose distributions for intensity-modulated

radiotherapy of rectal cancer. J Appl Clin Med Phys 2020;21:26–37. https://doi.org/10.1002/acm2.12849.