Radiotherapy and Oncology 176 (2022) 101–107

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Automated clinical decision support system with deep learning dose prediction and NTCP models to evaluate treatment complications in patients with esophageal cancer

Camille Draguet [a,b,*], Ana M. Barragán-Montero [a], Macarena Chocan Vera [a], Melissa Thomas [b,c], Pieter Populaire [b,c], Gilles Defraene [b], Karin Haustermans [b,c], John A. Lee [a], Edmond Sterpin [a,b]

[a] UCLouvain, Institut de Recherche Expérimentale et Clinique, Molecular Imaging Radiotherapy and Oncology (MIRO), Brussels; [b] KU Leuven, Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven; and [c] University Hospitals Leuven, Department of Radiation Oncology, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 8 April 2022
Received in revised form 3 August 2022
Accepted 28 August 2022
Available online 24 September 2022

Keywords:
Deep learning
NTCP model
Clinical decision tool
Patient referral
Esophageal cancer

ABSTRACT

Background and purpose: This study aims to investigate how accurate our deep learning (DL) dose prediction models for intensity modulated radiotherapy (IMRT) and pencil beam scanning (PBS) treatments, when chained with normal tissue complication probability (NTCP) models, are at identifying esophageal cancer patients who are at high risk of toxicity and should be switched to proton therapy (PT).
Materials and methods: Two U-Net were created, for photon (XT) and proton (PT) plans, respectively. To estimate the dose distribution for each patient, they were trained on a database of 40 uniformly planned patients using cross validation and a circulating test set. These models were combined with a NTCP model for postoperative pulmonary complications. The NTCP model used the mean lung dose, age, histology type, and body mass index as predicting variables. The treatment choice is then done by using a ΔNTCP threshold between XT and PT plans. Patients with ΔNTCP $\geq$ 10% were referred to PT.
Results: Our DL models succeed in predicting dose distributions with a mean error on the mean dose to the lungs (MLD) of 1.14 ± 0.93% for XT and 0.66 ± 0.48% for PT. The complete automated workflow (DL chained with NTCP) achieved 100% accuracy in patient referral. The average residual (ΔNTCP ground truth - ΔNTCP predicted) is 1.43 ± 1.49%.
Conclusion: This study evaluates our DL dose prediction models in a broader patient referral context and demonstrates their ability to support clinical decisions.
 Radiotherapy and Oncology 176 (2022) 101–107

Esophageal cancer (EC) is the seventh most common cancer and the sixth deadliest [1]. Nowadays, preoperative chemoradiotherapy is the standard treatment for locally advanced EC [2]. Unfortunately, this treatment is associated with severe risks of mortality and morbidity, as well as with a risk of developing complications in the heart, lungs, and gastrointestinal tract [3–5]. In particular, high radiation doses to the organs at risk (OAR) are important predictors of the development of complications, as reported in studies about normal tissue complications probabilities (NTCP) [6–9].

Specific NTCP models for EC are few, but they all suggest that higher OARs sparing reduces treatment complications [5,10–12]. Hence, proton therapy (PT) could potentially improve the patient's quality of life, since it better spares OARs compared to conventional photon-based radiotherapy (XT). In EC, PT has demonstrated a better sparing of relevant OARs, like the lungs and heart [13–15].

However, treating every patient with PT is not optimal, especially given its higher cost compared to XT [16–18]. Besides the level of dosimetric gain (e.g., little differences may have no clinical impact), the treatment outcome in EC is influenced by other factors, like age, cardiovascular risk profile, or body mass index (BMI). Clinical protocols that efficiently evaluate most relevant factors are needed to properly select patients who better benefit from PT. In the Netherlands, the model-based approach has been introduced for this purpose [18]. It requires XT and PT plans to be generated for each patient, which are later used, together with the aforementioned factors, to compute the difference in normal tissue complications probabilities between both modalities (ΔNTCP). The level of ΔNTCP determines the final patient referral to XT or PT [18–21].

The bottleneck of the model-based approach is the generation of high-quality XT and PT plans, which takes time and requires extra human resources. In addition, the planners' variable expertise may be a confounding factor and compromise the final treatment choice, when plans of suboptimal quality are generated. To

* Corresponding author at: Molecular Imaging, Radiotherapy and Oncology, Avenue Hippocrate 55/B1.55.01, B-1200 Bruxelles, Belgium.
E-mail address: camille.draguet@uclouvain.be (C. Draguet).

https://doi.org/10.1016/j.radonc.2022.08.031
0167-8140/

address this issue, different strategies for automatic planning have been proposed, including deep learning (DL) models trained to directly predict the optimal dose distribution for a given patient [22–27]. The combination of these DL models with NTCP models captures clinical knowledge, resulting in a treatment choice that ensures an optimal care for each patient, while removing human intervention and mitigating interobserver variability.

This work aims to develop a fully automated clinical decision support tool to refer a patient to either XT or PT. This tool harnesses recent advances in artificial intelligence applied to radiotherapy and combines them with state-of-the-art NTCP models in EC to recommend a specific treatment for each patient (Fig. 1).

Materials and methods

This study complied with the delineation and planning protocols of the PROTECT clinical trial [28], which will recruit in several European countries. This trial aims to determine whether PT will result in lower rates of postoperative pulmonary complications than XT in EC patients.

Patient database

The database contains 40 patients with locally advanced EC treated with neoadjuvant chemoradiation therapy followed by surgery, at the University Hospital of Leuven (UZLeuven) between 2016 and 2018. The use of patient data for the study was approved by the Institutional Ethical Review Board of the UZLeuven (S59667).

Following a data harvesting approach to increase consistency and quality, all patients were planned for PT and XT according to PROTECT guidelines (prescription of 50.4 Gy, 28 fractions). The plans were done in RayStation 10B (RaySearch Laboratories), by an experienced dosimetrist. The XT plan was designed for intensity modulated radiotherapy (IMRT), using a 7 mm iPTV. The PT plan was designed for intensity modulated proton therapy (IMPT) with pencil beam scanning, using robust optimization on iCTV (7 mm setup error and 2.6% range error). See Supplementary Materials for more information.

Deep learning model

We trained two DL models inspired by the popular U-Net architecture [29], one for each modality, to predict an optimal three-dimensional dose distribution. The U-Net encoder-decoder architecture was adapted by adding dense connection [30], similarly to previous works [24,25]. The model processes the CT scan and the contours of the body, 9 OARs and the target (PTV for XT and CTV for PT) delineated on the average CT, and outputs a 3D dose distribution map. The learning rate was 1e-3 and we used a mean square error as loss function. See Supplementary Materials for further details. As illustrated in Fig. 2, the database is repeatedly split into circulating groups for training (25 patients), validation (5 patients), and testing (10 patients). We employ a two-level ensembling strategy with four groups (outer level) of 5-fold cross-validation models (inner level), which results in a total of 20 models (4x5). This enables us to get predictions for all 40 patients (i.e. 10 test patients times 4 groups), while testing each of the 5 inner models in 10 totally unseen patients for each model. For each test patient, the outputs of the 5 inner models are averaged. The models were trained on a GPU TeslaA100 with 40 Gb of RAM. Training a fold with 250 epochs takes about 1,65 hours (8,25 hours for the 5-folds), and inference time for a test patient is about 8.3 seconds. Overall performance of the model in terms of the accuracy of the spatial dose distribution, using Dice coefficients for different isodose volumes and MSE on the test patients' predictions, are detailed in the Supplementary Materials.

Prior knowledge

A DL model should potentially yield more accurate results if it has additional prior information about the output it should produce. Therefore, we tried to improve the results obtained with the models presented in the previous subsection by passing as input to the network a first estimate of the dose distribution (D_0) for each patient, similar to the works of Barragan et al. [25] and

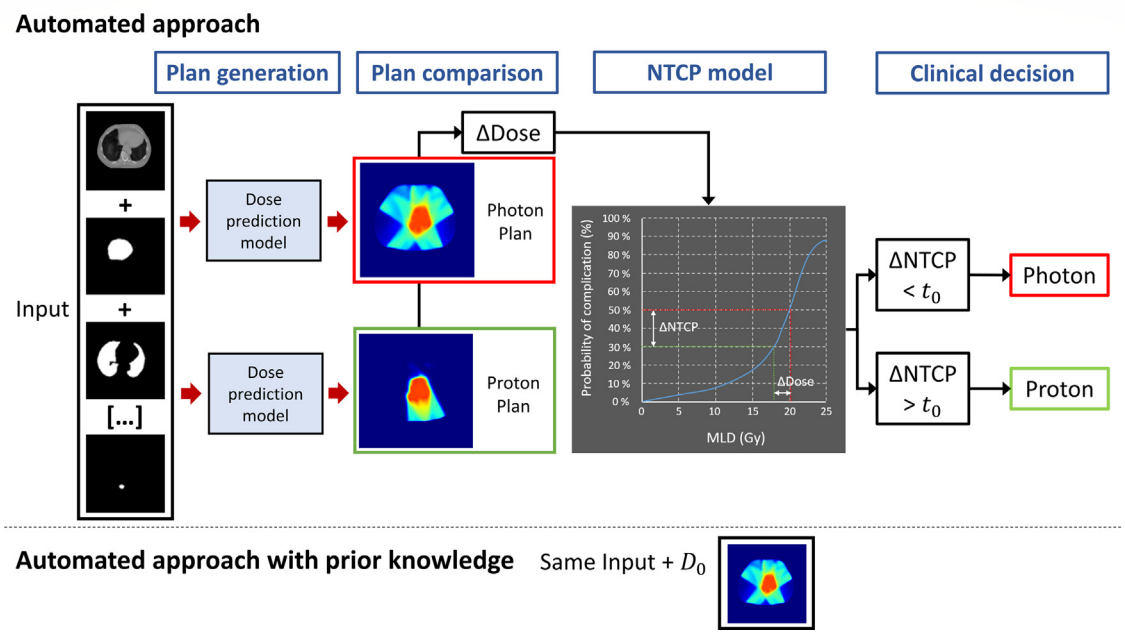

Fig. 1. Whole process workflow. The inputs are the CT scan and the contours of OARs (fundus, heart, both kidneys, left ventricle, liver, lungs, spinal canal and extended spinal canal) and of the CTV. D_0 = first estimate of the dose distribution.

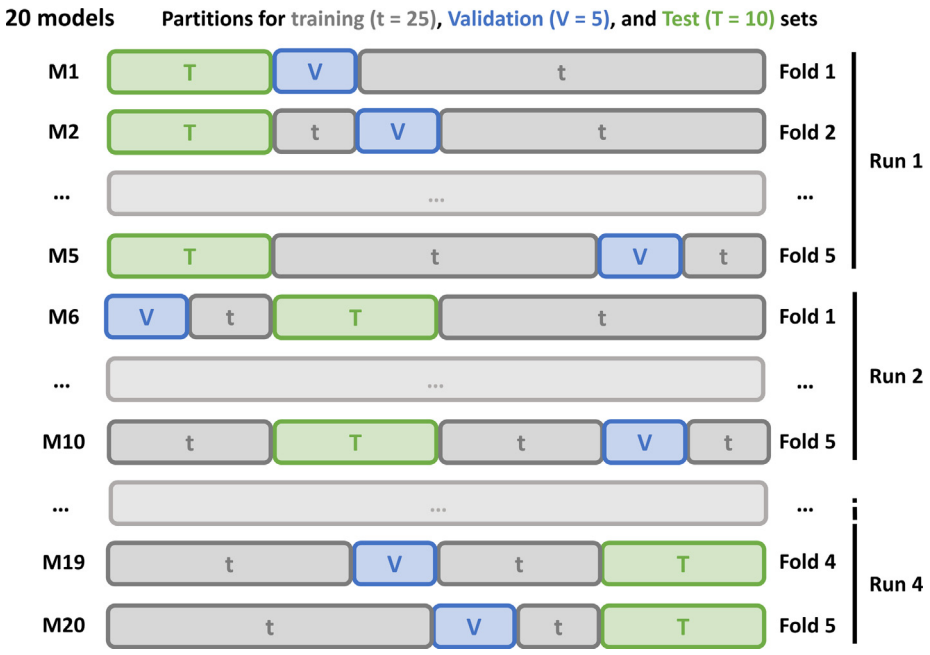

Fig. 2. Partitions of the training, validation and test sets for one modality. The dataset is divided into 4 groups of 5 models, each of these 5 models trained on 25 patients (represented in gray) and validated on 5 (blue). This 25–5 distribution is different for each of the 5 models. However, for each of the 4 groups of 5 models, we have 10 test patients (green) out of the 40 who were not seen by any of the 5 models. We observe that the test set (green) is the same for each group of 5 models and that the training (gray) and validation (blue) patients cannot be chosen in this test set. However, by proceeding in this way, we could only evaluate the performance of our models on 10 patients. To improve the significance of our statistical values and obtain a prediction for each of our 40 patients, we decided to repeat this 5-fold cross-validation process 4 times, taking each time a test set of 10 different patients.

Wang et al. [31]. It corresponds to the output of the first automatic optimization steps on the RayStation 10B software. This first estimate only results from running the plan optimization with a set of dose volume objectives imposed by a generic template (i.e., same penalties for all patients), without any manual operation. The process took 3 minutes 15 seconds on average for the XT plans, and 2 minutes 55 seconds for the PT plans. D_0 is then passed to the network input, in an additional channel separated from the CT scan and the contours. Improvement of the prediction with this model was verified with statistical tests.

Probability of complications

Pulmonary complication is an important factor to estimate the treatment outcome [14,32]. Thus, a NTCP model for pulmonary complications, developed at UZLeuven, was used to determine the treatment modality [5]. Pulmonary complications were defined as the occurrence of pneumonia, respiratory failure, or acute respiratory distress syndrome. The model was built with logistic regression and considers three clinical parameters (age, BMI, and histology type) and a dosimetric parameter, the mean dose to the lungs (MLD). Only two histologies were considered: adenocarcinoma (AC) and squamous cell carcinoma (SCC). Patients with other histologies were not used for the model building. Statistics about clinical parameters can be found in the Supplementary Materials. Table 1 reports the coefficients and general form of the model. The dataset contained patients treated with XT as well as PT. Unlike most of the literature, these NTCP model predictions are thus reliable for both modalities. In Fig. 3, we visualize the range of MLD in the manual plans used together with the NTCP predictions for these plans. The AC and SCC patients tend to cluster around their corresponding model curve. The outliers can be explained by an age that deviates significantly from the mean.

Table 1
Parameters and their coefficients of the pulmonary NTCP model [5]. AC = adenocarcinoma, SCC = squamous cell carcinoma.

Pulmonary complications parameters	Model coefficients
Intercept	−6.227
Age (year)	0.034
Histology (AC = 0, SCC = 1)	1.168
BMI	0.038
MLD (Gy)	0.144
Model general form	$p=\frac{\exp(b_0+b_1X_1+\cdots+b_nX_n)}{1+\exp(b_0+b_1X_1+\cdots+b_nX_n)}$

Model-based approach

The NTCP model described above was used to estimate the probability of pulmonary complications for both the XT and PT plans [5]. Following the Landelijk Platform Protonen Therapie (LPPT) protocol [33], we refer a patient to PT when the gain in NTCP with respect to XT is 10% or larger (ΔNTCP = 10%). The ΔNTCP thresholds defined in the LPPT protocol depend on the severity of the considered complications, which range from grade 1 (low) to grade 5 (high, death) according to the Common Terminology Criteria for Adverse Events (CTCAEv5.0) [34]. A ΔNTCP of 10% is suggested for complications of grade ≥ 2, and it decreases to 5% and 2% for grade ≥ 3 and ≥ 4 complications, respectively. In our case, the complications were of grade ≥ 1 (pneumonia), grade ≥ 3 (respiratory distress syndrome) and grade ≥ 4 (respiratory failure). Despite having complications of grade ≥ 3 and 4, we kept a rather conservative ΔNTCP = 10% to ensure PT selection only for patients for which the largest clinical benefit is expected.

In our study, we can identify 3 decision processes:

1. The "baseline approach", built on the manual XT and PT plans (also referred as ground truth plans).
2. The "automated approach", built on the XT and PT predicted by the standard DL models.

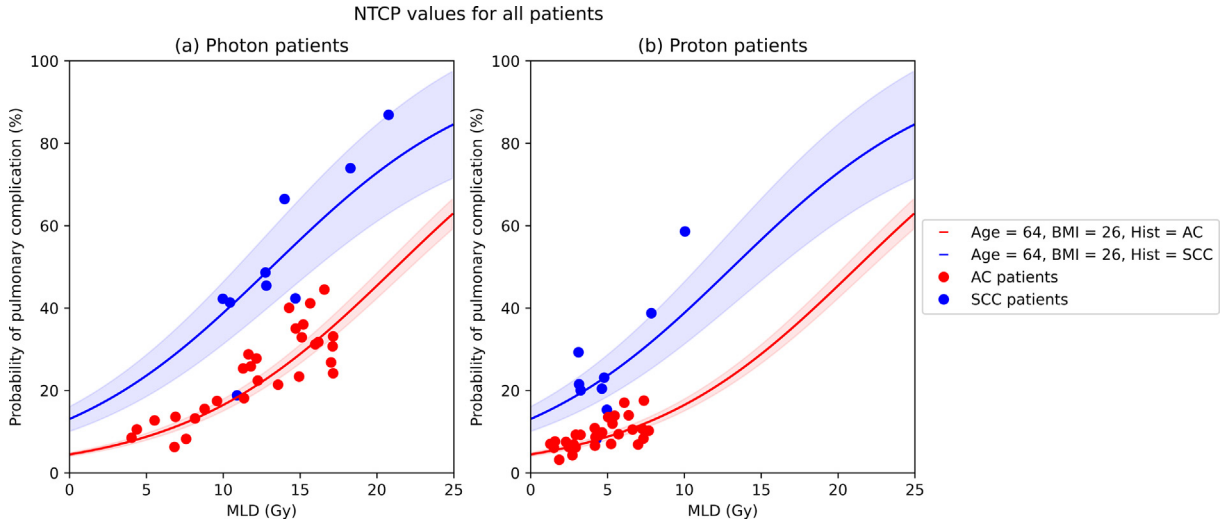

Fig. 3. Representation of the patients on the curves of the NTCP model together with their 95% confidence interval (lighter color shade). The red curve corresponds to adenocarcinoma histology while the blue curve corresponds to squamous cell carcinoma histology. Age = 64 and BMI = 26 correspond to the average age and BMI of the population. Patients planned for photon treatment (a) and proton treatment (b) are represented by the dots. AC = adenocarcinoma, SCC = squamous cell carcinoma.

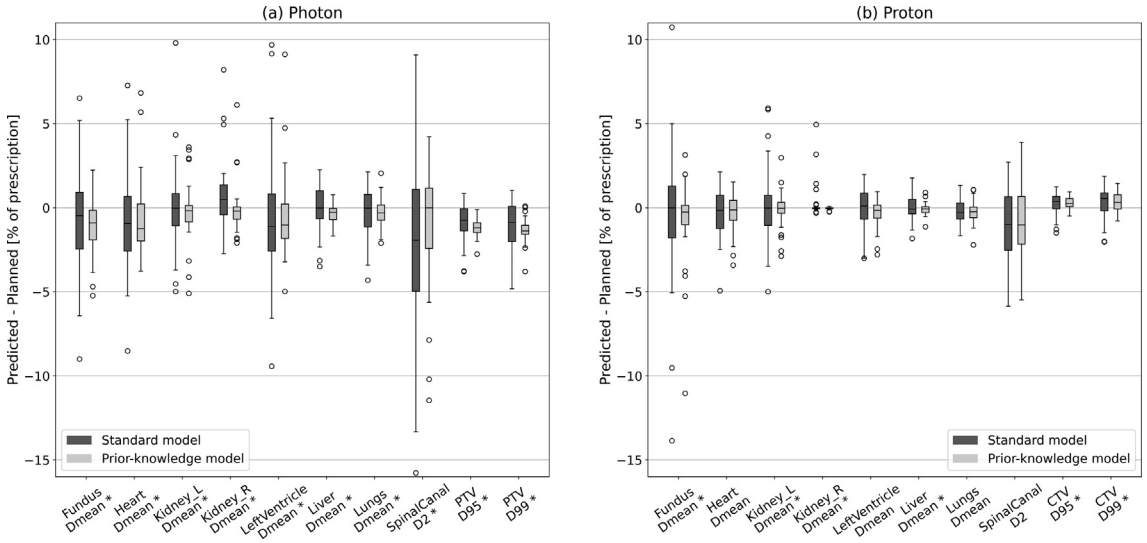

Fig. 4. Box plots of the prediction error on the DVH metrics (in % of the prescription = 50.4 Gy) for the photon model (a) and the proton model (b). For each structure, box plots for the standard model and for the prior-knowledge model are presented side by side. We were interested in the average dose for the organs in parallel and D2 for the organs in series. For the target, the D95 and D99 doses are represented. The star symbol (*) indicate structures for which there is a statistically significant difference between the variance on the error of the standard model and the variance on the error of the prior-knowledge model.

3. The "automated approach with prior-knowledge", built on the XT and PT predicted by the prior-knowledge DL models.

In each approach, ΔNTCP is calculated from the corresponding plans. We compared the baseline approach with the two automated approaches to see if a DL model could replace manual optimization in a clinical decision workflow.

Results

In total, 20 models for each modality were trained. A dose distribution could be predicted for each of the 40 patients, for both treatments. In order to evaluate the prediction accuracy, we considered metrics derived from dose-volume histograms (DVH metrics). These metrics reflect the clinical goals and ensure that damage to healthy tissue will be limited, as well as the patient's chances of developing a complication, while maintaining target coverage. Fig. 4 shows the error made on the prediction by the DL models, for both modalities. Most of the organs present an error on the predicted dose-volume metric that is below 7.5% of the prescribed dose, except the D2 on the spinal canal in XT (standard model). For the MLD, the error was 1.14 ± 0.93% (mean ± std) for the XT standard model, 0.72 ± 0.55% for the XT prior-knowledge model, 0.66 ± 0.48% for the PT standard model and 0.51 ± 0.42% for the PT prior-knowledge model. For the MHD, the error was 2.34 ± 1.91% for the XT standard model, 1.81 ± 1.44% for the XT prior-knowledge model, 1.1 ± 0.92% for the PT standard model and 0.77 ± 0.74% for the PT prior-knowledge model. Narrower box plots can be observed for the prior-knowledge models, indicating a more precise model. This precision was verified with an upper-tailed F-test on the variance of the error made on the different structures. We obtained statistically significant differences

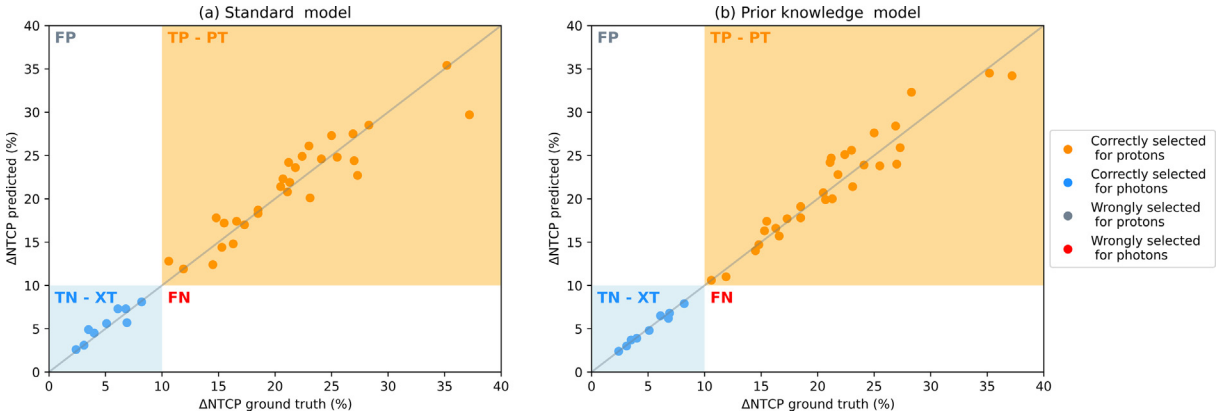

Fig. 5. Comparison between different approaches for clinical decision. (a) Comparison between the baseline approach (abscissa) and the standard automated approach (ordinate). (b) Comparison between the baseline approach (abscissa) and the automated approach in prior knowledge (ordinate).

($p = 0.05$) between the variances for all structures and all modalities except for the proton variances on the mean dose to the heart, left ventricle, lungs and on the D2 spinal canal. The statistically significant differences are emphasized in Fig. 4 with a star symbol (*) above the structure's names.

Fig. 5 depicts the comparison between the two automated approaches and the baseline approach. The authors from [5] specified that patients with histologies other than AC or SCC were excluded when the pulmonary NTCP model was built. The ΔNTCPs were only computed for 38 of the 40 patients because two patients did not meet these histology criteria. Based on the manual plans, 20/29 (69%) patients with AC were redirected to PT whereas 9/9 patients with SCC were redirected to PT. The clinical decision for each of these 38 patients is identical with the baseline and automated approaches, both in Fig. 5a and Fig. 5b, where the benefit of using the U-Net network with prior-knowledge rather than the standard network may be difficult to quantify. Therefore, we have also computed the residuals (ΔNTCP ground truth - ΔNTCP predicted) between the automated approaches and the baseline approach. The ΔNTCPs obtained with the automated approach with prior-knowledge were closer to the ground truth than the ΔNTCPs obtained with the standard automated approach, with a lower mean residual and a lower standard deviation (1.17 ± 1.11 % vs 1.43 ± 1.49%) but those differences were not statistically significant (T-test on the means and F-test on the variances).

Discussion

We have developed a fully automated approach that determines whether a patient with locally advanced EC would benefit from PT based on a CT scan, OARs and tumor contours, and some patient-specific clinical information. We have demonstrated the efficiency of our approach to provide a decision identical to the one obtained from manually generated plans, based on a preselected NTCP model. However, the accuracy of our tool is highly dependent on the choice of the threshold value. To test the sensitivity of our model with respect to this threshold, we explored values in the range [1%–20%] and reached accuracies in the range [92.1%–100%]. This tool can be used as a decision support for clinicians and has the advantage of being efficient in time and modular. The clinician has access to the dose distribution predicted by the DL model, has control over the choice of the NTCP model and the threshold at which a patient should be redirected to PT. Dice coefficients and low error on the dose maps predicted for the test patients were used to assess the reliability of our dose prediction DL models (see Supplementary Materials). We have also demonstrated the increase in precision of an automated approach with prior-knowledge, when available, compared to a standard automated approach. The prior-knowledge model did not bring statistically significant improvement for the MLD metric for PT, but it did for other relevant metrics. In case of using NTCP models where other metrics are involved, using prior-knowledge is recommended to ensure better decision support.

In our data harvesting approach, the data was shaped for integration into a deep learning model. The plans were made by an experienced dosimetrist, following a class solution set of guidelines, thus mitigating inter-observer variability. Training models with this kind of well-curated database achieves better performance [35]. Furthermore, shaping a database for a particular application allows us to integrate newly gained clinical insights when building the database. In our case of dose prediction for EC, special attention was paid to lung and heart sparing since several studies have shown a direct link between lung and heart doses and patient survival [10,36–39].

Using models like those presented in this work would accelerate and standardize the elaboration of PT and XT plans in the PROTECT trial. The method developed in this work proposes objective criteria to refer a patient to PT, that could be integrated in clinical trials in a standardized way. For QA purposes, this framework could also be used to compare ΔNTCP obtained in different centers, to ensure the quality of the dose distribution in all centers as well as having a uniform clinical decision for a patient across all centers. This could be of particular interest for PROTECT. A dose mimicking step is still required to generate a deliverable plan. Specifically, proton plans should include robust dose mimicking to ensure robustness against treatment uncertainties.

Importantly, DL dose prediction models can be used widely to transfer knowledge from centers with strong expertise to less experienced centers. This would open up the clinical trial to smaller, less experienced centers, allowing data to be collected in a larger patient population. This knowledge transfer also matters in the field of PT and EC, where expertise is not readily available since the technique is still emerging for this location.

The models developed in this study are specific, and thus limited, to EC patients. However, similar decision support tools can be easily developed for other cancer locations if DL dose prediction models are trained with specific databases [23–27]. Eventually, transfer learning could be used to build a new model in a faster way and with less data for a new location [40] or for emerging modalities (e.g., proton arc therapy [41]).

Moreover, the NTCP model selected to apply the model-based approach is specific to a certain patient population. It needs to be previously validated in the population used for the study. The

model was created to account for postoperative complications after preoperative chemoradiotherapy followed by surgery. We are aware of the influence of chemotherapy and surgery on complications and patient survival. The final decision should be made by looking at the big picture of the treatment. We acknowledge that the chosen NTCP model is limited in that it assesses the incidence of postoperative complications and it cannot be predicted whether or not the patient will undergo surgery following a chemoradiotherapy treatment. However, our DL models can be used with any NTCP model that needs dose metrics of the delivered dose distribution. As already mentioned, the existing NTCP models are typically developed for just one complication, without considering the whole spectrum of late and acute complications. One could imagine a clinical decision gathering several NTCP models, provided that all of them are relevant for the considered population and cancer type, as it is more commonly done for head and neck cancers [42,43]. Such an approach would necessitate determining the best way to combine these NTCP models to obtain a global clinical decision.

To the best of our knowledge, the NTCP models developed to date are simple logistic regression models that only consider clinical and dosimetric parameters. Future research could investigate an end-to-end DL NTCP model, to directly integrate 3D dose distribution, anatomy of major organs at risk and clinical parameters. This would allow direct consideration of 3D information for the development of NTCP models.

In conclusion, our population of 40 patients with locally advanced EC allowed us to develop a fully automated approach for patient referral to PT. This approach includes the training of a DL dose prediction model and the use of a NTCP model calculating the probability of developing pulmonary complications. The developed tool can also be advantageously shared between various centers without extra training, as long as the new input data are similar to the training data.

Conflicts of interest

We declare that none of the authors have conflict of interest.

Acknowledgments

The DL models were trained on clusters hosted by the "Consortium des équipements de calculs intensifs" (CECI). Camille Draguet is funded by UCLouvain. Ana M. Barragán-Montero is funded by the Walloon region in Belgium (PROTHERWAL/CHARP, grant 7289). Gilles Defraene is postdoctoral fellow of the Research Foundation Flanders (FWO, project 1292021N). John A. Lee is a Senior Research Associate with the F.R.S.-FNRS.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2022.08.031.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49. https://doi.org/10.3322/caac.21660.

[2] Stahl M, Mariette C, Haustermans K, Cervantes A, Arnold D, ESMO Guidelines Working Group. Oesophageal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013;24:vi51.

[3] Bailey SH, Bull DA, Harpole DH, Rentz JJ, Neumayer LA, Pappas TN, et al. Outcomes after esophagectomy: A ten-year prospective cohort. Ann Thoracic Surgery 2003;75:217–22. https://doi.org/10.1016/S0003-4975(02)04368-0.

[4] Seely, Andrew JE, Ivanovic J, Threader J, Al-Hussaini A, Al-Shehab D, et al. Systematic classification of morbidity and mortality after thoracic surgery. Ann Thoracic Surg 2010;90:936–42. https://doi.org/10.1016/j.athoracsur.2010.05.014.

[5] Thomas M, Defraene G, Lambrecht M, Deng W, Moons J, Nafteux P, et al. NTCP model for postoperative complications and one-year mortality after trimodality treatment in oesophageal cancer. Radiother Oncol 2019;141:33–40. https://doi.org/10.1016/j.radonc.2019.09.015.

[6] Rancati T. Factors predicting radiation pneumonitis in lung cancer patients: a retrospective study. Radiother Oncol 2003;67:275–83. https://doi.org/10.1016/s0167-8140(03)0011.

[7] Chapet O, Kong F-M, Lee JS, Hayman JA, Ten Haken RK. Normal tissue complication probability modeling for acute esophagitis in patients treated with conformal radiation therapy for non-small cell lung cancer. Radiother Oncol 2005;77:176–81. https://doi.org/10.1016/j.radonc.2005.10.0.

[8] 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.

[9] Van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol 2017;122:185–91. https://doi.org/10.1016/j.radonc.2016.07.0.

[10] Beukema JC, Kawaguchi Y, Sijtsema NM, Zhai T, Langendijk JA, van Dijk LV, et al. Can we safely reduce the radiation dose to the heart while compromising the dose to the lungs in oesophageal cancer patients? Radiother Oncol 2020;149:222–7. https://doi.org/10.1016/j.radonc.2020.05.033.

[11] Warren S, Partridge M, Carrington R, Hurt C, Crosby T, Hawkins MA. Radiobiological determination of dose escalation and normal tissue toxicity in definitive chemoradiation therapy for esophageal cancer. Int J Radiat Oncol Biol Phys 2014;90:423–9. https://doi.org/10.1016/j.ijrobp.2014.06.028.

[12] Tucker SL, Liu HH, Wang S, Wei X, Liao Z, Komaki R, et al. Dose–volume modeling of the risk of postoperative pulmonary complications among esophageal cancer patients treated with concurrent chemoradiotherapy followed by surgery. Int J Radiat Oncol Biol Phys 2006;66:754–61. https://doi.org/10.1016/j.ijrobp.2006.06.002.

[13] Lin, Steven H, Komaki R, Liao Z, Wei C, Myles B, et al. Proton beam therapy and concurrent chemotherapy for esophageal cancer. Int J Radiat Oncol Biol Phys 2012;83:e345–51. https://doi.org/10.1016/j.ijrobp.2012.01.003.

[14] Wang J, Wei C, Tucker SL, Myles B, Palmer, Matthew CMD, et al. Predictors of postoperative complications after trimodality therapy for esophageal cancer. Int J Radiat Oncol Biol Phys 2013;86:885–91. https://doi.org/10.1016/j.ijrobp.2013.04.006.

[15] Lin SH, Merrell KW, Shen J, Verma V, Correa AM, Wang L, et al. Multi-institutional analysis of radiation modality use and postoperative outcomes of neoadjuvant chemoradiation for esophageal cancer. Radiother Oncol 2017;123:376–81. https://doi.org/10.1016/j.radonc.2017.04.013.

[16] Brada M, Pijls-Johannesma M, De Ruysscher D. Proton therapy in clinical practice: Current clinical evidence. J Clin Oncol 2007;25:965–70. https://doi.org/10.1200/JCO.2006.10.0131.

[17] Lievens Y, Pijls-Johannesma M. Health economic controversy and cost-effectiveness of proton therapy. Seminars Radiat Oncol 2013;23:134–41. https://doi.org/10.1016/j.semradonc.2012.11.005.

[18] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013;107:267–73. https://doi.org/10.1016/j.radonc.2013.05.007.

[19] Widder J, Van Der Schaaf A, Lambin P, Marijnen CA, Pignol JP, Rasch CR, et al. The quest for evidence for proton therapy: Model-based approach and precision medicine. Int J Radiat Oncol Biol Phys 2016;95:30–6. https://doi.org/10.1016/j.ijrobp.2015.10.004.

[20] Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van der Schaaf A, et al. Clinical trial strategies to compare protons with photons. Seminars Radiat Oncol 2018;28:79–87. https://doi.org/10.1016/j.semradonc.2017.11.008.

[21] Tambas M, Steenbakkers RJHM, van der Laan HP, Wolters AM, Kierkels RGJ, Scandurra D, et al. First experience with model-based selection of head and neck cancer patients for proton therapy. Radiother Oncol 2020;151:206–13. https://doi.org/10.1016/j.radonc.2020.07.056.

[22] Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Medical Phys (Lancaster) 2019;46:370.

[23] Guerreiro F, Seravalli E, Janssens GO, Maduro JH, Knopf AC, Langendijk JA, et al. Deep learning prediction of proton and photon dose distributions for paediatric abdominal tumours. Radiother Oncol 2021;156:36–42. https://doi.org/10.1016/j.radonc.2020.11.026.

[24] Nguyen D, Jia X, Sher D, Lin M, Iqbal Z, Liu H, et al. Three-dimensional radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. 2018.

[25] Barragán-Montero AM, Nguyen D, Lu W, Lin M, 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. Medical Phys (Lancaster) 2019;46:3679–91. https://doi.org/10.1002/mp.13597.

[26] Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning 1076–1076. Sci Reports 2019;9. https://doi.org/10.1038/s41598-018-37741-x.

[27] Kandalan RN, Nguyen D, Rezaeian NH, Barragán-Montero AM, Breedveld S, Namuduri K, et al. Dose prediction with deep learning for prostate cancer radiation therapy: Model adaptation to different treatment planning practices. Radiother Oncol 2020;153:228–35. https://doi.org/10.1016/j.radonc.2020.10.027.
[28] National Library of Medicine. PROton Versus Photon Therapy for Esophageal Cancer – a Trimodality Strategy (PROTECT). Identifier NCT05055648. (2021, October -) https://clinicaltrials.gov/ct2/show/NCT05055648.
[29] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at the, 2015;9351 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
[30] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2017.243.
[31] Wang H, Zhang H, Hu J, Song Y, Bai S, Yi Z. DeepEC: An error correction framework for dose prediction and organ segmentation using deep neural networks. Int J Intell Syst 2020. https://doi.org/10.1002/int.22280.
[32] Law S, Wong K, Kwok K, Chu K, Wong J. Predictive factors for postoperative pulmonary complications and mortality after esophagectomy for cancer. Ann Surg 2004;240:791–800. https://doi.org/10.1097/01.sla.0000143123.24556.1c.
[33] Consensus document voor selection van patiënten met een model-based indicatie voor protonen therapie (Unpublished internal document), Landelijk Platform Protonentherapie (LPPT), 2015.
[34] Common Terminology Criteria for Adverse Events (CTCAE) Version 5. Published: November 27. US Department of Health and Human Services, National Institutes of Health, National Cancer Institute https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf.
[35] Barragán-Montero AM, Thomas M, Defraene G, Michiels S, Haustermans K, Lee JA, et al. Deep learning dose prediction for IMRT of esophageal cancer: The effect of data quality and quantity on model performance. Physica Medica 2021;83:52–63. https://doi.org/10.1016/j.ejmp.2021.02.026.
[36] Xu C, Guo L, Liao Z, Wang Y, Liu X, Zhao S, et al. Heart and lung doses are independent predictors of overall survival in esophageal cancer after chemoradiotherapy. Clin nd Translat Radiat Oncol 2019;17:17–23. https://doi.org/10.1016/j.ctro.2019.04.016.
[37] Frandsen J, Boothe D, Gaffney DK, Wilson BD, Lloyd S. Increased risk of death due to heart disease after radiotherapy for esophageal cancer. J Gastrointestinal Oncol 2015;6:516–23. https://doi.org/10.3978/j.issn.2078-6891.2015.040.
[38] Cai G, Li C, Yu J, Meng X. Heart dosimetric parameters were associated with cardiac events and overall survival for patients with locally advanced esophageal cancer receiving definitive radiotherapy 153–153. Front Oncol 2020;10. https://doi.org/10.3389/fonc.2020.00153.
[39] Addendum Slokdamcarcinoom bij Landelijk Indicatie Protocol Protonentherapie Longcarcinoom, Landelijk Platform Protonentherapue (LPPT), 2021 https://nvro.nl/images/Addendum_Slokdarmcarcinoom_bij_LIPP_longcarcinoom_goedgekeurd_ZiN_151021.pdf.
[40] Pratt L, Jennings B. A Survey of Transfer Between Connectionist Networks. Connect Sci 1996;8:163–84. https://doi.org/10.1080/095400996116866.
[41] Ding X, Li X, Zhang JM, Kabolizadeh P, Stevens C, Yan D. Spot-Scanning Proton Arc (SPArc) Therapy: The First Robust and Delivery-Efficient Spot- Scanning Proton Arc Therapy. Int J Radiat Oncol • Biol • Phys 2016;96:1107–16. https://doi.org/10.1016/j.ijrobp.2016.08.049.
[42] Van den Bosch L, van der Schaaf A, van der Laan HP, Hoebers FJ, Wijers OB, van den Hoek JG, et al. Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. Radiotherapy Oncol 2021;157:147–54. https://doi.org/10.1016/j.radonc.2021.01.024.
[43] Landelijk Indicatie Protocol Protonentherapie (versie 2.2) (LIPPv2.2) Hoofd-Halstumoren, andelijk Platform Protonentherapie (LPPT), 2019. https://nvro.nl/images/documenten/rapporten/2019-08-15__Landelijk_Indicatieprotocol_Protonentherapie_Hoofdhals_v2.2.pdf.