Deep learning for instantaneous estimation of the out-of-field dose in external beam photon radiotherapy -A proof of concept

Nathan Benzazon, Mohammed El Aichi, Alexandre Carré, François de Kermenguy, Stéphane Niyoteka, Pauline Maury, Julie Colnot, Meissane M'Hamdi, Cristina Veres, Rodrigue Allodji, et al.

Deep learning for instantaneous estimation of the out-of-field dose in external beam photon radiotherapy – A proof of concept

Nathan Benzazon[1, 2], Mohammed El Aichi[1, 2], Alexandre Carré[1, 2], François de Kermenguy[1, 2], Stéphane Niyoteka[1, 2], Pauline Maury[1, 2], Julie Colnot[1, 2, 3], Meissane M'hamdi[1, 2], Cristina Veres[1, 2], Rodrigue Allodji[4], Florent de Vathaire[4], David Sarrut[5], Neige Journy[4], Claire Alapetite[6], Vincent Grégoire[7], Eric Deutsch[1, 2], Ibrahima Diallo[1, 2] and Charlotte Robert[1, 2]

[1]Unité Mixte de Recherche 1030, ImmunoRadAI, Université Paris-Saclay, Institut Gustave Roussy, Inserm, Villejuif, France
[2]Department of Radiation Oncology, Gustave Roussy, Villejuif, France
[3]THERYQ, PMB-Alcen, Peynier, France
[4]Unité Mixte de Recherche 1018, Radiation Epidemiology Team, Université Paris-Saclay, Institut Gustave Roussy, Inserm, Villejuif, France
[5]Université de Lyon; CREATIS; CNRS UMR5220; Inserm U1294; INSA-Lyon; Université Lyon 1, Léon Bérard cancer center, Lyon, France.
[6]Department of Radiotherapy, Institut Curie, Paris, France.
[7]Department of Radiation Oncology, Centre Léon-Bérard, Lyon, France.

Abstract

Doses deposited outside the treatment field during external photon-beam radiotherapy, known as out-of-field doses, appear to favor the appearance of radiation-induced cancer and hematological toxicities. These low doses are not currently estimated on a routine clinical basis, due to the lack of a suitable solution.
With this proof of concept, we demonstrate the value of deep learning for the development of a tool enabling rapid, routine-compatible out-of-field dose estimation. For this purpose, a 3D U-Net, considering as inputs the in-field dose, as computed by the treatment planning system, and the patient's anatomy, was trained to predict out-of-field dose maps, using a dataset of 3151 pediatric whole-body dose maps estimated with analytical methods as ground truth. Consistent results in line with the literature were obtained (RMSD of 0.28, 0.41 and 0.32 $cGy.Gy^{-1}$ for the training, validation and testing), arguing in favor of the use of neural networks for out-of-field dose estimation.

1 Introduction

In external beam radiotherapy (EBRT) with photons, non-zero doses are inevitably delivered outside the treatment field; often referred to as out-of-field or peripheral dose. The out-of-field doses are mostly low dose values (< 4 Gy), but the question of their potential impact on the probability of radiation-induced adverse events, such second cancers (1) and immunological dysfunctions like radiation-induced lymphopenia (2) is of the utmost importance. In particular, it has recently been shown that even very low doses resulting from computed tomography (CT) scans contribute to the development of radiation-induced cancers (3). In radiotherapy, this topic is currently experiencing renewed interest, particularly as modulated treatments (such as volumetric modulated arc therapy (VMAT) or intensity-modulated radiation therapy (IMRT)) are now routinely used in clinical care and are known to be associated with higher peripheral doses due to longer beam on times and larger irradiated volumes when compared to three-dimensional conformal radiotherapy (3D-CRT) (3,4)

Treatment planning systems (TPS) are used in clinical routine to estimate in-field dose distribution, but have been shown to systematically underestimate out-of-field dose for 3D conformational radiotherapy treatments, for intensity modulated treatments and for CyberKnife devices (Accuray, Sunnyvale, USA) (5–8). Thus, despite clear clinical potential, the out-of-field dose computation is currently not available in clinical practice. Two methods are currently used in the literature for out-of-field dose estimation for research purposes: Monte Carlo simulations and analytical approaches. These methods can provide accurate out-of-field dose estimation (7) but are inappropriate for clinical routine implementation, because of their lack of versatility and their cumbersome application (9).

Artificial intelligence and in particular technologies based on deep learning (DL) have drastically changed clinical practice in radiotherapy in recent years, allowing notably the automation of several time-consuming tasks, including segmentation and treatment planning, or the generation of virtual images (synthetic images), demonstrating their ability to identify complex hierarchical features from spatially structured data (10). Three main components contribute to the out-of-field dose: patient scatter, head scatter and head leakage. Apart from leakage, which is a signature specific to each accelerator and whose signal is minimal in the irradiation field, the other components exhibit a valuable in-field signal for assessing the out-of-field dose. Based on the assumption that the in-field dose map associated with the patient's anatomy contains most of the information needed to calculate the out-of-field dose, our objective was to evaluate the ability of a unique deep learning-based out-of-field dose estimation algorithm to adapt to a wide variety of configurations, bringing an answer to the problems of computation time, information extraction, and versatility, making it compatible with clinical use.

2 Materials and Methods

Three thousand one hundred fifty-one whole-body dose maps, from the French childhood cancer survivor study (FCCSS) cohort were used as ground truth for the development of our neural network. With the goal to develop a deep learning model for out-of-field dose estimation suitable for megavoltage photon irradiations, we only selected patients treated with a photon beam and using linear accelerators with a high voltage (> 1 MV) or ^{60}Co irradiators. Whole-body dose maps were estimated using an analytical method originally developed for epidemiological purposes (11,12). Twenty-five irradiation devices were represented in our dataset and were grouped into 3 categories: standard linacs, cobalt units and betatron units. Our dataset included 38 different pathologies, of which the most represented were nephroblastoma and other nonepithelial renal tumors (695 patients), Hodgkin lymphomas (449 patients) and astrocytomas (235 patients).

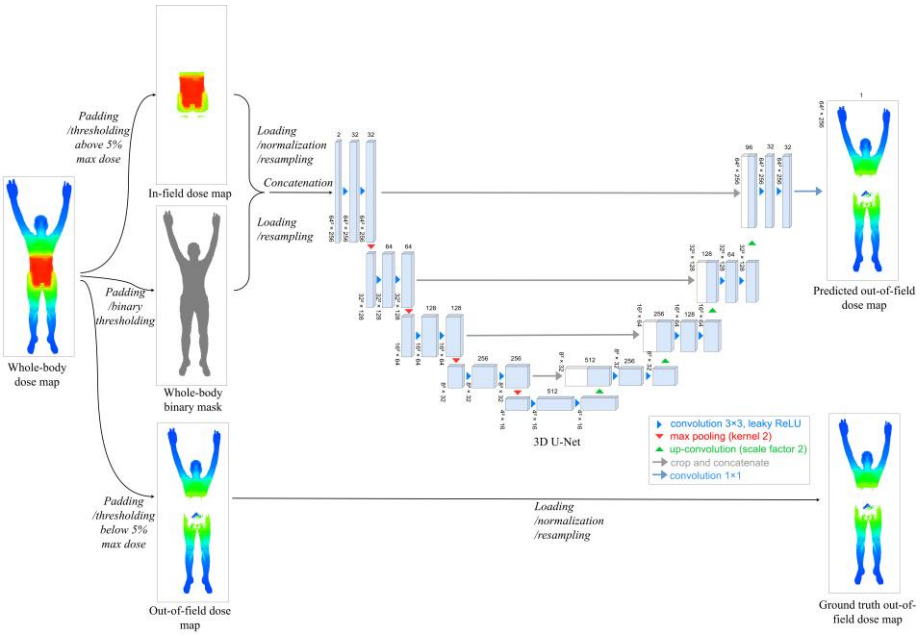

Figure 1 Dose map preprocessing pipeline and comprehensive design of the implemented 3D U-Net. Dose maps are displayed using logarithmic scale

Several preprocessing steps were applied to the analytical dose maps to make them deep-learning compliant. These included: 1) padding into [370, 242, 1131] matrix sizes, 2) resampling of padded files (originally in Statistical Analysis System (SAS) format) to [128, 128, 512] sizes, 3) extraction of in-field and out-of-field dose maps from whole-body dose maps (a 5% isodose threshold was chosen in this goal considering the maximum dose per patient as the reference dose), 4) creation of binary masks from the whole-body dose maps by a thresholding method separating the background from the foreground. An on-the-fly preprocessing pipeline was then applied using Medical Open Network for Artificial Intelligence (MONAI 0.8.0) (13), including in the following order: loading, normalization, resampling, and concatenation. During the normalization step, 3D dose maps intensities were normalized by dividing by 100 Gy, to provide the neural network with values within [0,1]. The on-the-fly resampling step was implemented to test the impact of batch size as a function of available VRAM (video random-access memory). A nearest neighbor interpolation strategy was used for the resampling step. Finally, we transformed the data to fit a multi-dimensional structure denoted as B×2×H×W×D. In this expression, B refers to the batch size used during the training phase of the process. C indicates the number of channels present, while H, W, and D correspond to the height, width, and depth of the matrix, respectively. Figure 1 summarizes the preprocessing pipeline.

A conventional 3D U-Net (14,15), composed of four down-sampling blocks followed by four up-sampling blocks, was implemented (Figure 1). The Mean Square Error (MSE) evaluated only on the foreground voxels outside the 5% isodose, i.e. only in the region considered as the out-of-field dose, was selected as loss function. Learning rate and weight decay were considered in the ranges $[1e-7, 1e-3]$ and $[1e-8, 1e-4]$, respectively, with 1e-4 and 1e-6 providing the optimal results. The Adam optimizer was used to update the network parameters. Instance normalization was preferred. Batch size of 20 corresponding to a resampled size of $64 \times 64 \times 256$ was selected. The 3D U-Net was trained for 600 epochs (~50 hours) on a Nvidia RTX A6000. No early stopping was used. Weight based on best performance on validation loss were saved.

The dataset of 3151 patients was conventionally split into training (N = 2213), validation (N = 505), and test cohorts (N = 433). The test set was divided into 5 sub-cohorts, including two classic linear accelerators (names of accelerator model unavailable) operating at 6 MV and 16 MV, two cobalt 60 units (Alcyon and Mobiletron), and finally a betatron operating at 1.25 MV. All patients treated with betatron devices were voluntarily kept into the test set, because of its very specific design compared with a conventional linear accelerator or a cobalt unit. None of the devices in the test set were included in the training and validation sets. This was made in order to test the generalizability of the trained network on unknown devices.

We finally performed 2 additional experiments. The first aimed at evaluating if the inclusion of betatron-treated patients in the training set could improve performance for a subgroup of 46 patients here considered as a new sub-test set. The second additional experiment aimed to study the

benefits of a custom model, compared with the use of a generalized model. In this aim, we selected 100 of the 128 patients treated with a ^{60}Co-Alcyon, and trained the 3D U-Net from scratch considering only these data. Finally, as a last experiment, we fine-tuned the original model using the 100 ^{60}Co-Alcyon training patients for 2000 epochs and tested the performance of this fine-tuned model on the 28 ^{60}Co-Alcyon patients included in the test set. We used the root mean square deviation (RMSD) (corresponding to the root of the MSE) as performance evaluation metric.

3 Results

Best performances were achieved at epoch 467/600. For visual purposes, Figure 2 shows three examples of out-of-field dose maps generated by our trained network from test sets for patients with good (RMSD of 0.16 cGy.Gy^{-1}), median (RMSD of 0.29 cGy.Gy^{-1}) and poor performance (RMSD of 1.00 cGy.Gy^{-1}). RMSD of 0.28 ± 0.08 and 0.41 ± 0.26 cGy.Gy^{-1} were obtained for the training and validation datasets, respectively. Values of 0.27 ± 0.06, 0.26 ± 0.07, 0.28 ± 0.06, 0.30 ± 0.12 and 0.45 ± 0.25 cGy.Gy^{-1} were achieved for the 6 MV linac, 16 MV linac, ^{60}Co-Alcyon, ^{60}Co-Mobiletron, and betatron devices test sets, respectively, demonstrating overall performance similar to or better than that of the validation set, except for the fifth test set, corresponding to the betatron device. A RMSD threshold value of 0.6 cGy.Gy^{-1} was considered to separate good from poor out-of-field dose reconstructions. 87 out of 505 patients showed weaker performances in the validation set; 85 of whom being treated with a single device: a Sagittaire linear accelerator operating at 25 MV. This value was equal to 24 (out of 433 patients) in the test set, with 21 of the 24 patients identified having been treated with the betatron accelerator.

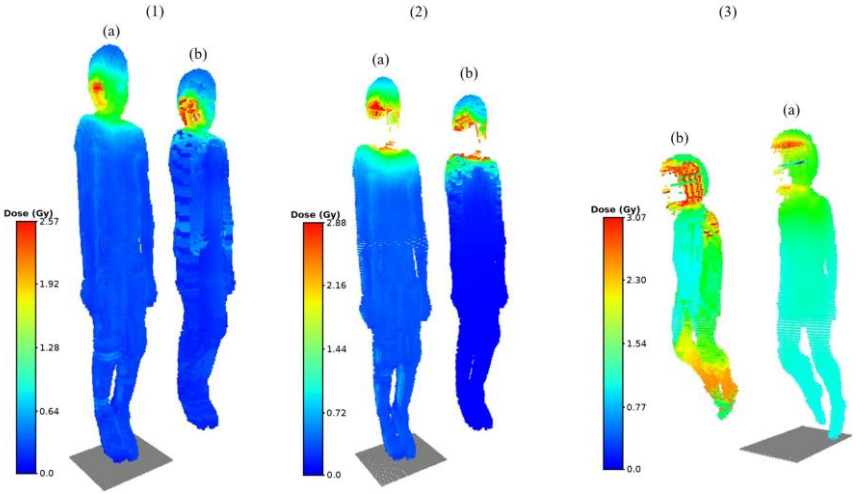

Figure 2 Comparison of out-of-field dose maps predicted by the 3D U-Net (a) and computed by the analytical model, the ground truth (b), for 3 patients. Patient 1 (RMSD=0.16 cGy.Gy^{-1}, into the 5% percentile) corresponds to a 15-year-old girl diagnosed for a primary pathology of retinoblastoma in 1982 and treated with a Neptune 6 MV device. Patient 2 (RMSD=0.29 cGy.Gy^{-1}, in ± 1% of median) is also a 15-year-old girl who was diagnosed and treated in 1982 for an intracranial neoplasm using an Alcyon 60Co device. Patient 3 (RMSD=1.00 cGy.Gy^{-1}, below 95% percentile) is a 5-year-old girl diagnosed in 1982 treated the same year with a 1.25 MV betatron device for an

Retraining the neural network using the original patient training set plus the 46 betatron patients kept aside resulted in a mean RMSD of 0.64 ± 0.41 cGy.Gy^{-1}, to be compared with 0.26 ± 0.11 cGy.Gy^{-1} for the same 46 patients as a benchmark value (no betatron patients included at the training stage). For the second additional experiment, mean RMSDs of 0.16 ± 0.01 cGy.Gy^{-1}, 0.53 ± 0.22 cGy.Gy^{-1} and 0.13 ± 0.05 cGy.Gy^{-1} were obtained for the benchmark configuration, configuration considering training with only the 100 ^{60}Co-Alcyon training set patients, and configuration considering the original model fine-tuned with the 100 ^{60}Co-Alcyon training set patients, respectively.

4 Discussion

The results are rather encouraging, with RMSD values of the same order of magnitude in the test sets (mean value of 0.32 ± 0.15 cGy.Gy^{-1} in the test set) as in the validation set (0.41 ± 0.26 cGy.Gy^{-1}) or the training set (0.28 ± 0.08 cGy.Gy^{-1}), suggesting that the neural network has acquired a strong degree of robustness and generalization. However, we notice that significantly lower results were observed for test set corresponding to the betatron accelerator. In addition, most of the poorest results in the validation set concerned patients treated with the Sagittaire accelerator operating at 25 MV (the highest voltage of the linear accelerators considered in the training set was equal to 20 MV). These results logically highlight the fact that the generalization capabilities of the neural network cannot be extended to out-of-distribution cases including non-conventional or highly atypical linear accelerators, i.e. with different spectral characteristics, shielding properties, or head geometries. Another limitation of our proof of concept is the tendency by the neural network to systematically overestimated doses at locations farthest from the field. This discrepancy appears to stem from our choice of cost function (MSE), which minimizes absolute dose errors. Consequently, the greatest relative deviations are observed in these distant regions, where the lowest dose are expected.

The first additional experiment shows that adding betatron patients to the training set resulted in an unexpected decline in performance (0.26 ± 0.11 cGy.Gy^{-1} for the benchmark compared to 0.63 ± 0.41 cGy.Gy^{-1} when considering betatron patients in the training set). We hypothesize that the number of betatron cases in the training set is too small to allow the neural network to converge towards a solution tailored to this accelerator, and that this accelerator may be too unusual and confuses the network. The results obtained from the second additional experiment demonstrated that the model developed from 100 ^{60}Co-Alcyon patients showed poorer averaged performance on the 28 ^{60}Co-Alcyon test patients than the original model, trained

without any ^{60}Co-Alcyon data, but on more than 2000 patients (0.53 ± 0.22 cGy.Gy^{-1} for the custom model versus 0.16 ± 0.01 cGy.Gy^{-1} for the benchmark). This suggests that the inclusion of a wider variety of patient anatomies and irradiation configurations allows the network to better predict out-of-field dose distributions than focusing on a model that has been trained on data from a single accelerator, but with a smaller number of patients and diversity of treatment conditions. The results of the fine-tuned model nevertheless demonstrate the benefits of specializing the model to a certain extent. Indeed, the fine-tuned model makes it possible to achieve hitherto unequalled performance (0.13 ± 0.05 cGy.Gy^{-1}). This approach appears to be the most promising since it enables both optimizing performance on a particular machine of interest and using the vast diversity of treatment conditions found in the original database.

To address the identified limitations (overestimation of the lowest dose values far from the field and limited application for the most atypical linear accelerators), a strategy involving the incorporation of a relative or local component in the network's cost function, along with the inclusion of additional data in the training dataset, could offer remedies. Furthermore, a separate research team is collaboratively working on developing whole-body dose maps for contemporary linear accelerators and treatments, such as intensity modulation, using Monte Carlo simulation. Retraining the network with these new data will enhance its applicability to modern irradiation conditions. Finally, we carried out experimental measurements (and are still continuing this campaign) on a whole-body anthropomorphic phantom using radiophotoluminescent dosimeters with the aim of experimentally validating the whole-body dose maps generated by deep learning.

5 Conclusion

Based on this proof of concept, we have shown that deep learning is a relevant tool for addressing the limitations of analytical methods or MC simulations for out of field dose estimation. Thanks to its generalization capabilities and short inference times of just a few seconds, this tool should make it possible to move forward for routine clinical application and mass application in retrospective studies. We are convinced that a tool for estimating out-of-field dose in clinical routine would be a powerful aid for optimizing modern radiotherapy treatments.

References

1. Goy E, Tomezak M, Facchin C, Martin N, Bouchaert E, Benoit J, et al. The out-of-field dose in radiation therapy induces delayed tumorigenesis by senescence evasion. Elife. 2022 Mar 18;11:e67190.
2. de Kermenguy F, Meziani L, Mondini M, Clémenson C, Morel D, Deutsch E, et al. Radio-induced lymphopenia in the era of anti-cancer immunotherapy. In: International Review of Cell and Molecular Biology [Internet]. Academic Press; 2023 [cited 2023 Jun 30].
3. Colnot J, Zefkili S, Gschwind R, Huet C. Out-of-field doses from radiotherapy using photon beams: A comparative study for a pediatric renal treatment. J Appl Clin Med Phys. 2021 Feb 5;
4. Yoon J, Heins D, Zhao X, Sanders M, Zhang R. Measurement and modeling of out-of-field doses from various advanced post-mastectomy radiotherapy techniques. Phys Med Biol. 2017 Nov 13;62(23):9039–53.
5. Colnot J, Barraux V, Loiseau C, Berejny P, Batalla A, Gschwind R, et al. A new Monte Carlo model of a Cyberknife® system for the precise determination of out-of-field doses. Phys Med Biol. 2019 Oct 4;64(19):195008.
6. Howell RM, Scarboro SB, Kry SF, Yaldo DZ. Accuracy of out-of-field dose calculations by a commercial treatment planning system. Phys Med Biol. 2010 Dec 7;55(23):6999–7008.
7. Sánchez-Nieto B, Medina-Ascanio KN, Rodríguez-Mongua JL, Doerner E, Espinoza I. Study of out-of-field dose in photon radiotherapy: A commercial treatment planning system versus measurements and Monte Carlo simulations. Med Phys. 2020 Sep;47(9):4616–25.
8. Schneider U, Hälg RA, Hartmann M, Mack A, Storelli F, Joosten A, et al. Accuracy of out-of-field dose calculation of tomotherapy and cyberknife treatment planning systems: a dosimetric study. Z Med Phys. 2014 Sep;24(3):211–5.
9. Benzazon N, Colnot J, de Kermenguy F, Achkar S, de Vathaire F, Deutsch E, et al. Analytical models for external photon beam radiotherapy out-of-field dose calculation: a scoping review. Frontiers in Oncology [Internet]. 2023 [cited 2023 May 29];13.
10. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. Nat Rev Clin Oncol. 2020 Dec;17(12):771–81.
11. Diallo I, Lamon A, Shamsaldin A, Grimaud E, de Vathaire F, Chavaudra J. Estimation of the radiation dose delivered to any point outside the target volume per patient treated with external beam radiotherapy. Radiother Oncol. 1996 Mar;38(3):269–71.
12. Veres C, Allodji RS, Llanas D, Vu Bezin J, Chavaudra J, Mège JP, et al. Retrospective reconstructions of active bone marrow dose-volume histograms. Int J Radiat Oncol Biol Phys. 2014 Dec 1;90(5):1216–24.
13. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. MONAI: An open-source framework for deep learning in healthcare [Internet]. arXiv; 2022 [cited 2023 Apr 18]. Available from: http://arxiv.org/abs/2211.02701
14. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [Internet]. arXiv; 2016 [cited 2023 Jan 5]. Available from: http://arxiv.org/abs/1606.06650
15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 May 18 [cited 2020 Aug 11]; Available from: https://arxiv.org/abs/1505.04597v1