Deep-Learning for Rapid Estimation of the Out-of-Field Dose in External Beam Photon Radiation Therapy – A Proof of Concept

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

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

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

1 : Unité Mixte de Recherche (UMR) 1030 Radiothérapie Moléculaire et Innovation Thérapeutique, 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 (UMR) 1018 Centre de Recherche en épidémiologie et Santé des Populations (CESP), 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.

† : Shared last authorship

Abstract

Background and purpose

The dose deposited outside of the treatment field during external photon beam radiotherapy treatment, also known as out-of-field dose, is the subject of extensive study as it may be associated with a higher risk of developing a second cancer, and could have deleterious effects on the immune system which compromise the efficiency of combined radio-immunotherapy treatments. Out-of-field dose estimation tools developed today in research, including Monte Carlo simulations and analytical methods, are not suited to the requirements of clinical implementation because of their lack of versatility and their cumbersome application. We propose a proof of concept based on deep learning for out-of-field dose map estimation that addresses the above limitations.

Materials and methods

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. The cohort used for learning and performance evaluation included 3151 pediatric patients from the FCCSS database, treated in 5 clinical centers, whose whole-body dose maps were previously estimated with an empirical analytical method. The test set, composed of 433 patients, was split into 5 subdatasets, each containing patients treated with devices unseen during the training phase. Root mean square deviation (RMSD) evaluated only on non-zero voxels located in the out-of-field areas was computed as performance metric.

Results

RMSD of 0.28 and 0.41 $cGy.Gy^{-1}$ were obtained for the training and validation datasets, respectively. Values of 0.27, 0.26, 0.28, 0.30 and 0.45 $cGy.Gy^{-1}$ were achieved for the 6 MV linac, 16 MV linac, Alcyon cobalt irradiator, Mobiletron cobalt irradiator, and betatron devices test sets, respectively.

Conclusion

This proof-of-concept approach using a convolutional neural network has demonstrated unprecedented generalizability for this task, although it remains limited, and brings us closer to an implementation compatible with clinical routine.

Introduction

In external beam radiotherapy with photons, non-zero doses are inevitably delivered outside the treatment field; this is often referred to as out-of-field or peripheral dose. The 5% isodose of the prescribed dose is the most commonly used threshold to differentiate between in-field and out-of-field regions [1–6]. While the out-of-field dose is mostly less than 4 Gy, questions remain about the potential association between this low dose exposure and adverse events such as second malignant neoplasms [7], immunological dysfunction [8,9], cardiovascular disease [10] and neurological effects [11]. In particular, it has

recently been shown that even very low dose exposures from computed tomography (CT) scans have been associated with the development of subsequent cancer [12]. 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 tend to result in higher peripheral doses due to longer beam on times and larger irradiated volumes when compared to three-dimensional conformal radiotherapy (3D-CRT) [13–16]. Recent studies have yielded inconsistent results, with varying effects observed across different populations, and were largely limited by too short duration of follow-up to detect any impact of advances in RT on the risk of radiation-related second cancers which are typically reported after decades of latency times [17–22]. In a very different context, the assessment of out-of-field doses seems increasingly crucial for optimizing radiotherapy treatments in the near future. The recent awareness of the immunomodulatory role of radiotherapy, coupled with the observation of a link between radiation-induced lymphopenia and patient response to treatment in several solid tumor sites, suggests the need to spare lymphocyte-rich structures as much as possible [23,24]. Indeed, the fact that the lethal doses reported by several independent groups are of the order of a few gray [25–27] is a strong argument for the need to characterize the dose in the periphery to the primary field.

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-CRT, IMRT, VMAT, and CyberKnife treatments (Accuray, Sunnyvale, USA) [2,13,28–30]. 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 (MC) simulations, which are based on a stochastic approach and aim to estimate the average dose per voxel and its associated variance by simulating the tracking path of millions of incident particles knowing cross-sections of particle-matter interaction, and analytical approaches, which mathematically model the out-of-field dose, either based on physical or empirical models [31]. MC simulations and analytical methods can provide accurate out-of-field dose estimation [29], but are today inappropriate for clinical routine implementation. On the one hand, MC simulations are hardware- and time-intensive,

especially when it concerns the evaluation of low doses, and require detailed modeling of the irradiator , which can be tedious when the technical drawing of the device is not available. On the other hand, analytical models, which most of the time require new experimental measurements to adjust intrinsic parameters, are not suitable for large retrospective studies, which often legacy linear accelerators (linacs) or cobalt irradiators that are no longer available for experimental measurements.

In recent years, artificial intelligence, especially technology rooted in deep learning, has drastically changed the clinical practice of radiotherapy through the automation of several time-consuming tasks such as segmentation [32], treatment planning [33], and generation of virtual images (synthetic CT [34]). These examples demonstrate that deep learning can identify complex hierarchical features from spatially structured data [35], a capability which might also be exploited for estimation of out-of-field dose. Out-of-field dose consists of three main components: the patient scatter component, which corresponds to secondary photons resulting from a Compton interaction in the treatment field of primary photons, depositing their energy in the area outside the treatment field, the collimator scatter (or head scatter) component which results in doses deposited outside the treatment field by particles that have interacted with the collimator or other parts of the irradiating device head, and the leakage component which is made up of primary particles that have not been intercepted by collimation parts. The patient scatter component depends mainly on the size of the irradiated volume and the beam spectrum, and is the largest component close to the field [36], the two other components depend mostly on the geometry of the irradiation device and its configuration during the treatment. The leakage component appears to be the dominant component far from the field [36], and its amplitude is strongly dependent on the distance from the isocenter, as it depends on the angular shielding properties of the machine in a general way, i.e. integrating the attenuation properties of the jaws and the multi-leaf collimator. While it is obvious that the information relating to the patient scattering component is included in the in-field area, the absorbed dose generated by the head scattering component also generates signal in the field as the photons resulting from the Compton interaction are scattered throughout the volume [37]. The leakage component is ultimately a signature of the irradiation device, making the task undoubtedly more complex for a neural network. 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 single algorithm based on deep learning to adapt to a wide variety of configurations, bringing an answer to the problems of computation time, information extraction, and versatility. Even if major publications providing out-of-field dose data for specific treatment modalities are available and have been of great interest for this research topic [38], the interest of artificial intelligence via its generalization capabilities seems to us to be a promising field for. On the one hand, it can propose a solution that is not limited to the treatment configurations presented in the data tables; and, on the other hand, making it possible to predict the out-of-field dose map specific to each linac and to the patient's anatomy thanks to information extracted from the in-field dose map and the use of the 3D representation of the patient. In this paper, we present a proof of concept applied to high-energy photon irradiations.

Material and methods

1. Description of the dataset

The French Childhood Cancer Survivor Study (FCCSS) cohort, whose primary aim is to study long-term effects of treatments for cancer on children and adolescents, has been used in this work. More than 7000 patients under 21 years of age treated in 5 French centers between 1945 and 2000 for solid cancer or lymphoma composed this cohort. With the goal to develop a deep learning model for out-of-field dose estimation suitable for high-energy photon irradiation, the following inclusion criteria were considered: 1) patients treated with a photon beam 2) treatments using linacs with a high voltage > 1 MV or cobalt irradiators. A total of 3310 patients were selected at this stage. For all cohort members, whole body dose map was reconstructed using an analytical method originally developed for bone marrow dose analysis [39]. This empirical method concatenated the 3D dose map as estimated by the Isogray TPS (Dosisoft, Cachan, France) in in-field areas; each treatment plan having being re-simulated by an experienced operator based on the treatment details ; to an out-of-field dose estimation obtained device-wise. Out-of-field dose were estimated using reference percentage of depth dose values (known in function of depth in tissue, field size and source to skin distance) [40] extended to off-axis-ratio area by

in-water phantom measurements performed and gathered since the mid-1980s [41,42], and the absorbed dose on the beam central axis at maximum dose depth for reference source to skin distance and field size. Comprehensive description with mathematical support of the model can be found in [39]. Experimental measurement campaigns have been carried out since the mid-1980s on exhaustively all the devices listed in the cohort (see Supplemental Table 1). Each measure was used to fit the analytical model under different specific characteristics of the device studied [41–43]. In the present work, we did not carry out new measurements to validate the analytical dose maps, as the data have already been validated and used extensively in a large number of previous publications [44–48]. The process of dose reconstructions, considered various factors such as patient gender, age, height, weight, thickness at the target volume level, treatment position, treatment machine type, beam energy, irradiation technique, field size and shape, gantry angle, collimator angle, presence of accessories like wedges, target volume location, and the actual dose administered to the target volume. All 3D whole body dose maps had voxels dimensions of $2 \times 2 \times 2$ mm^3. The exclusion criteria applied to these whole-body dose maps were as follows: 1) outlier dose maps, i.e. maps presenting local doses higher than 100 Gy, which had no physical justification in view of the doses prescribed (N = 104) were deleted [49], 2) corrupted dose maps (N = 2) were removed. At the end, data from 3151 patients were kept for this study. Table 1 summarizes the distribution by center of these patients. Twenty-five irradiation devices were grouped into 3 categories: standard linacs, cobalt units and betatron units. The subset 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).

Center		Gustave Roussy, *Villejuif*	Institut Curie, *Paris*	Institut Claudius-Regaud, *Toulouse*	Institut Godinot, *Reims*	Centre Antoine Lacassagne, *Nice*	Total
Number of patients		2676	237	199	68	24	3204
Male **Female**		1447 1229	135 102	121 78	43 25	17 7	1763 1441
Mean age ± std [years]		7.3 ± 4.9	6.3 ± 5.0	7.2 ± 4.7	7.9 ± 5.0	6.8 ± 4.7	7.2 ± 4.9
Irradiation device (Number of patients/Number of devices/High voltage range or Mean energy)	**Linac**	1040 / 12 4-25 MV	110 / 4 6-25 MV	25 / 2 20-25 MV	32 / 2 16-25 MV	0 / 0 NA	1207/14 4-25 MV
	Cobalt irradiator	1532 / 5 1.25 MeV	86 / 7 1.25 MeV	174 / 4 1.25 MeV	36 / 2 1.25 MeV	24 / 3 1.25 MeV	1852/10 1.25 MeV
	Betatron	90 / 1 1.25 MV	2 / 1 1.25 MV	0 / 0 NA	0 / 0 NA	0 / 0 NA	92/1 1.25 MV

Table 1: Characteristics of the patients selected in this work from the FCCSS cohort described by center.

2. Data preprocessing

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) [50], including in the following order: loading, normalization, resampling, and concatenation. During the normalization step, 3D dose maps intensities were normalized 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, an adjustment of data was implemented to align with the input parameters required by the neural network. This adjustment involved transforming 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. The channel, distinguished by its two-dimensional

structure, encompasses the conjoined data from the whole-body binary masks and the in-field dose maps. Figure 1 summarizes the preprocessing pipeline.

3. Neural network training

A conventional 3D U-Net [51–53], composed of four down-sampling blocks followed by four up-sampling blocks, was implemented (Figure 1). The Mean Square Error (MSE, see Equation 1) evaluated only on the foreground voxels outside the 5% isodose, i.e. only in the region considered in this paper as the out-of-field dose, was selected as loss function.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(D_{nn,i} - D_{gt,i})^2 \qquad (1)$$

With $D_{nn,i}$ and $D_{gt,i}$ the normalized doses to the voxel i estimated respectively with the neural network and from the ground truth analytical method, and n the number of voxels considered.

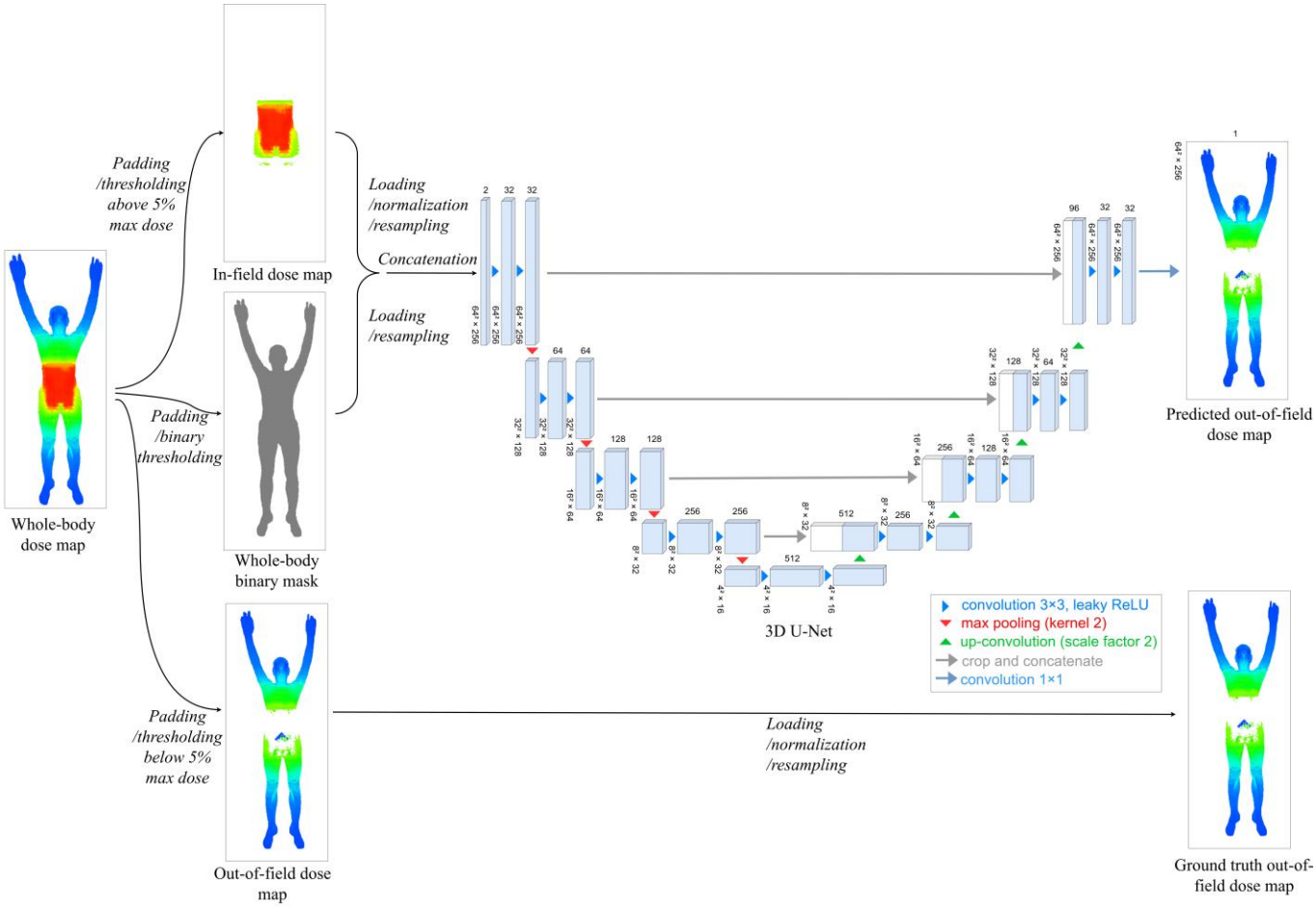

Fig 1. Dose map preprocessing pipeline and comprehensive design of the implemented 3D U-Net. Dose maps are displayed using logarithmic scale.

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) as shown in Table 2. As one of our main objectives was to test the hypothesis of generalization of the trained network to unseen machines, data splitting was performed in a controlled manner. Thus, the data were stratified so that 18 different irradiation devices formed the training set (including classic linacs and cobalt irradiators), while 2 unseen machines were part of the validation set exclusively (one cobalt irradiator, and one Sagittaire linac operating at 25 MV). For the validation, we chose to select two devices: a conventional linac and a cobalt irradiator, each associated with more than 100 patients treated, with the aim of monitoring the performance of the neural network on a large selection of patients for 2 different technologies in the training set. On the basis of these observations, we chose the Sagittaire linac as it is the accelerator operating at the highest energy in our database, and enables us to test the generalizability of our approach in this energy range, and arbitrarily a cobalt irradiator. The test set was divided into 5 subcohorts, including two classic linacs (names of accelerator model unavailable) operating at 6 MV and 16 MV, two cobalt irradiators (called 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 linac or a cobalt unit. No stratification on clinical data was applied. Supplemental Table 1 provides a full description of the dataset.

We also carried out two additional deep learning 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 46 patients were randomly selected from the test set 5 comprising 92 patients). Inference of the previously trained neural network on these 46 identified patients was considered as a benchmark. We then performed network retraining from scratch using the already identified hyperparameters on the original training data, to which were added the other 46 betatron patients not used for the benchmark, and tested the performance of this new network on the 46 patients kept aside from the training. The second additional deep learning 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 an Alcyon cobalt irradiator (Test set 3), and trained the 3D U-Net from scratch considering only these data. Training was carried out during 2000 epochs, using the same hyperparameters as for training the original generalized model, with the exception of a batch of 4. We then tested the performance of this custom-model on the 28 remaining Alcyon patients. We compared these results with the inference performance of the original generalized model on the same 28 patients (considered as our benchmark here). Finally, as a last deep learning experiment, we fine-tuned the original model using the 100 Alcyon cobalt irradiator training patients for 2000 epochs and tested the performance of this fine-tuned model on the 28 Alcyon cobalt irradiator patients included in the test set.

Because the root mean square deviation ($RMSD$, Equation 2) is used as performance evaluation metrics by a lot of research teams developing analytical models for out-of-field dose estimation [16,31,54–58], this measure was selected for the reporting of the results. To study the impact of the field size on the neural network performance, the results were also analyzed in subgroups, for which the threshold corresponded to the median size of the in-field volumes in the training set (3767 cm^3). Similarly, RMSD values were computed in two different zones: we differentiated the area outside the radiation field into a near-field area and a far-field area. The 1% isodose was chosen to distinguish these two zones. Finally, the mean absolute dose differences between the predictions and the ground truth were estimated per patient. Medians and min-max ranges were computed for the validation and test sets, also distinguishing between the near-field zone and the far-field zone.

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(D_{nn,i} - D_{gt,i})^2} \qquad (2)$$

All manipulations on data have been implemented in Python 3.7.7.

	Total	**Training**	**Validation**	**Test**				
				Subtest 1 6 MV Neptune	**Subtest 2** 16 MV	**Subtest 3** Alcyon cobalt irradiator	**Subtest 4** Mobiletron cobalt irradiator	**Subtest 5** Betatron
Linac	1207	683	365	103	56	0	0	0
Cobalt irradiator	1852	1530	140	0	0	128	54	0
Betatron	92	0	0	0	0	0	0	92
Subtotal	3151	2213	505	433				

Table 2: Distribution of patients and devices included into training, validation and testing processes.

Results

Best performances were achieved at epoch 467/600. Learning curves for the training and validation sets are presented in Supplemental Figure 1. Table 3 presents the $RMSD$ results obtained on training, validation and test sets for hyperparameters presented in the material and methods section. Table 4 summarizes the median of absolute dose differences per patient on validation and test sets.

		Training	**Validation**	**Test**				
				Subtest 1 6 MV Neptune	**Subtest 2** 16 MV	**Subtest 3** Alcyon cobalt irradiator	**Subtest 4** Mobiletron cobalt irradiator	**Subtest 5** Betatron
RMSD ± Standard deviation (cGy.Gy^{-1})	Out-of-field area	0.28 ± 0.08	0.41 ± 0.26	0.32 ± 0.15				
				0.27 ± 0.06	0.26 ± 0.07	0.28 ± 0.06	0.30 ± 0.12	0.45 ± 0.25
	Close from the field area (>1 % isodose)		0.49 ± 0.29	0.40 ± 0.20				
				0.41 ± 0.13	0.39 ± 0.13	0.33 ± 0.13	0.31 ± 0.16	0.54 ± 0.31
	Far from the field area (<1 % isodose)		0.22 ± 0.12	0.25 ± 0.08				
				0.24 ± 0.06	0.23 ± 0.07	0.27 ± 0.06	0.30 ± 0.14	0.22 ± 0.07
	Large irradiation field		0.37 ± 0.23 (N = 262)	0.28 ± 0.11				
				0.25 ± 0.08 (N = 36)	0.26 ± 0.08 (N = 23)	0.26 ± 0.07 (N = 48)	0.31 ± 0.15 (N = 27)	0.39 ± 0.13 (N = 18)
	Small irradiation field		0.43 ± 0.28 (N = 242)	0.33 ± 0.11				
				0.29 ± 0.06 (N = 67)	0.26 ± 0.07 (N = 33)	0.29 ± 0.06 (N = 80)	0.29 ± 0.10 (N = 27)	0.46 ± 0.26 (N = 74)

Table 3: Performances of the 3D U-Net for the out-of-field dose map estimation task.

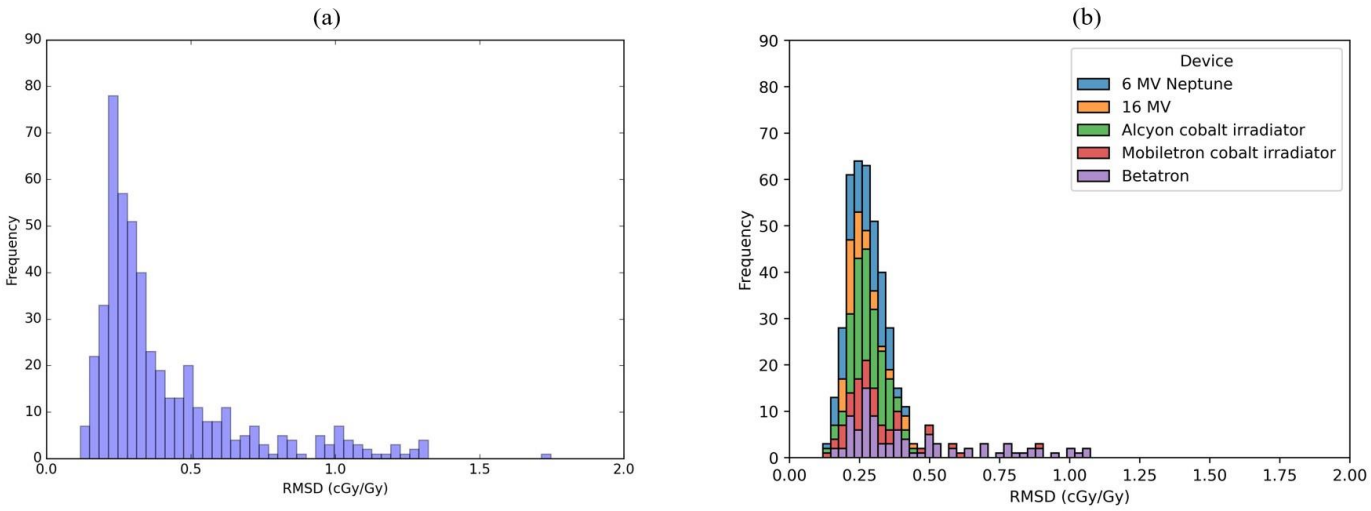

Fig 2. Histograms of the $RMSD$ metric computed patient-wise obtained on the validation dataset (a) and the test dataset detailed for each machine (b).

$RMSD$ 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, Alcyon cobalt irradiator, Mobiletron cobalt irradiator, 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. The same observations are obtained by analyzing the medians of the dose differences per patient with values of 0.31, 0.27, 0.23, 0.27, 0.30 and 0.33 Gy obtained for the validation sets, linac 6 MV, linac 16 MV, Alcyon cobalt irradiator, Mobiletron cobalt irradiator and betatron devices, respectively (Table 4). Similarly, betatron is associated with the highest maximum dose difference of the test sets (maximum value equal to 1.07 Gy). The results observed in the validation and test sets as a function of distance from the irradiation field show better $RMSD$ values far from the field than close to it, except for the fourth (corresponding to the Mobiletron cobalt irradiator). For example, far from the field $RMSD$ values of 0.22 ± 0.12 and 0.25 ± 0.08 cGy.Gy^{-1} are reported for respectively the validation and the test sets, while 0.49 ± 0.29 and 0.40 ± 0.20 cGy.Gy^{-1} are reported for area close from the field. Finally, the results of the validation set comparing dose maps for large fields and small fields suggest that the neural

network performs better for large irradiation fields (0.37 ± 0.23 cGy.Gy^{-1} to be compared with 0.43 ± 0.28 cGy.Gy^{-1}), but this difference is less pronounced than the differences observed in the previous results between areas close to the field and areas far from the field, especially when taking into account standard deviation of $RMSD$ results. This trend is also observed in the fifth test set (0.39 ± 0.13 cGy.Gy^{-1} to be compared with 0.46 ± 0.26 cGy.Gy^{-1}), while the other four test sets showed fairly similar results between large and small irradiation fields. The same observation can be drawn for absolute dose differences, with higher values obtained for the near-field zone for both the validation and test sets (median dose differences equal to 0.33 Gy and 0.24 Gy for the test set in the near-field and far-field zones, respectively). Based on this metric, the subset of patients treated with the Mobiletron cobalt irradiator obtained the poorest results in the far-field zone and was associated with the highest maximum dose difference (0.85 Gy).

Histograms of the $RMSD$ metric computed patient-wise on the validation set (Figure 2a) and on test sets (Figure 2b) were plotted, leading to heavy-tailed distributions, and more specifically to log-normal distributions. On the basis of these figures, an 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 linac 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.

		Validation	**Test**				
			Subtest 1 6 MV Neptune	**Subtest 2** 16 MV	**Subtest 3** Alcyon cobalt irradiator	**Subtest 4** Mobiletron cobalt irradiator	**Subtest 5** Betatron
Median dose difference (Mininum / Maximum) (Gy)	Out-of-field area	0.31 (0.12 / 1.75)	0.28 (0.12 / 1.07)				
			0.27 (0.13 / 0.41)	0.23 (0.18 / 0.43)	0.27 (0.12 / 0.43)	0.30 (0.14 / 0.89)	0.33 (0.16 / 1.07)
	Close from the field area (>1 % isodose)	0.38 (0.11 / 1.79)	0.33 (0.12 / 1.24)				
			0.37 (0.18 / 0.91)	0.37 (0.18 / 0.71)	0.30 (0.16 / 0.68)	0.27 (0.12 / 0.94)	0.40 (0.15 / 1.24)
	Far from the field area (<1 % isodose)	0.19 (0.04 / 0.87)	0.24 (0.07 / 0.85)				
			0.23 (0.12 / 0.43)	0.21 (0.17 / 0.42)	0.26 (0.07 / 0.41)	0.26 (0.14 / 0.85)	0.20 (0.13 / 0.58)

Table 4: Median and range of dose differences between predicted dose maps and ground truth.

Figure 3 shows the out-of-field dose maps obtained for 3 representative patients from the testing sets: one with good performance (Patient a, $RMSD = 0.16$ cGy.Gy^{-1}), one with median performance (Patient b, $RMSD = 0.29$ cGy.Gy^{-1}) and one with poor performance (Patient c, $RMSD = 1.00$ cGy.Gy^{-1}). Associated axial dose profiles are presented in Figure 4. Applying our entire pre-processing pipeline including data loading, neural network application and data saving takes an average computation time of 2.59 ± 0.09 seconds, 0.047 ± 0.008 seconds being necessary for the out-of-field dose calculation itself.

For Patients a and b, doses are well predicted in the near-field dose gradient, with mean relative dose differences, evaluated along the profiles (Figure 4), equal to 23.1 % and 26.3 % up to 20 cm from the edge of the field. For Patient a, the continuous component far from the field is correctly reconstructed (average dose of 0.23 Gy for the ground truth compared with 0.26 Gy for the prediction by the network between 20 and 155 cm), but with a jump in dose that is not correctly predicted. For patient b, the continuous component far from the field is not predicted as well (average dose of 0.06 Gy for ground truth compared with 0.30 Gy for network prediction between 20 and 153 cm). For patient c, treated on a betatron machine, we note that the neural network has weaknesses in dose prediction not only in areas close to the field, but also in more distant areas, where, for example, it fails to predict the local increase in dose on the patient's legs associated with this specific linac.

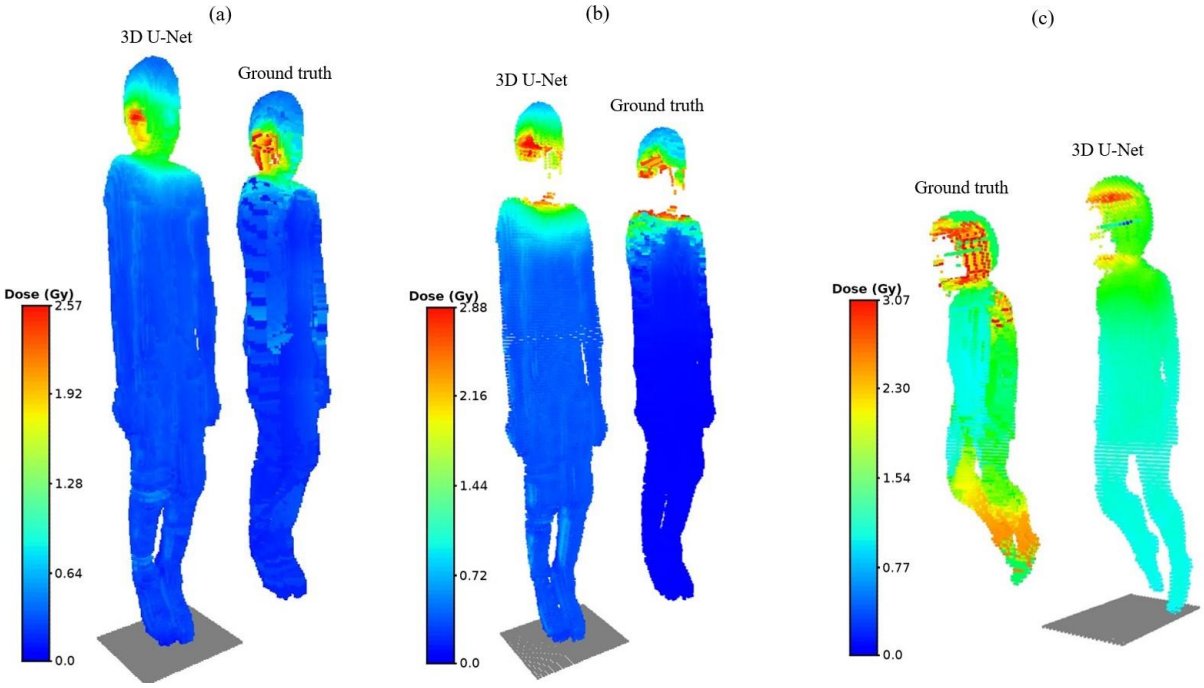

Fig 3. Comparison of out-of-field dose maps predicted by the 3D U-Net and computed by the analytical model, the ground truth, for 3 patients of varying performance (test set only). Patient a (Test set 1 – $RMSD = 0.16$ cGy.Gy^{-1}, into the 5% percentile of the distribution) corresponds to a 15-year-old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

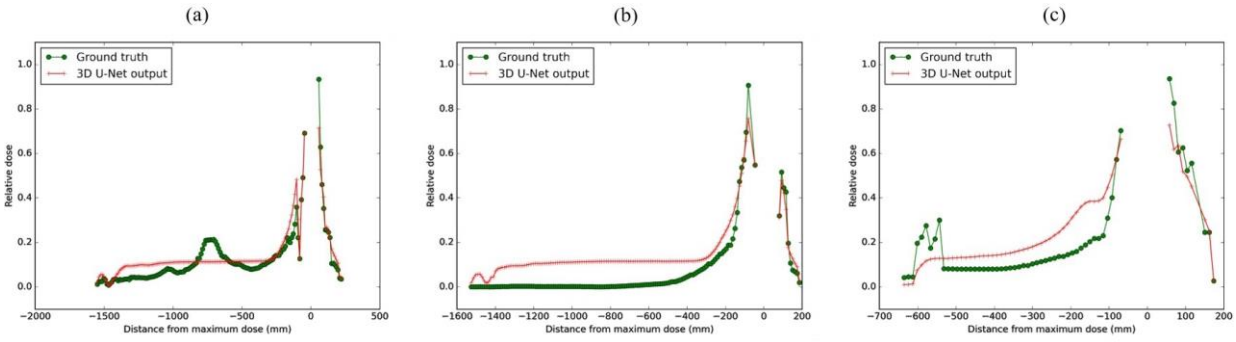

Fig 4. Head-foot dose profiles of the whole body dose maps predicted by the U-Net 3D network (in green) compared with the dose profile of the associated ground truth dose map (in red). Profiles were

plotted only in the out-of-field dose areas, for three same patients as in Figure 3 (Patient a (a), Patient b (b) and Patient c (c)). The doses were normalized to the maximum value obtained from the ground truth out-of-field dose maps (analytical model).

Finally, Figure 5 establishes a link between the doses predicted by the network and the ground truth doses for each of the 3 patients, the objective being a linear curve passing through 0 with a slope of 1. The visualization confirms the observations made earlier from the profiles. In particular, for Patient a, the points corresponding to lower dose values show greater clustering around the linear trend than for the other patients. Figure 5.b clearly illustrates the overestimation of doses far from the field. Figure 5.c shows a more scattered distribution of the points, in line with previous observations.

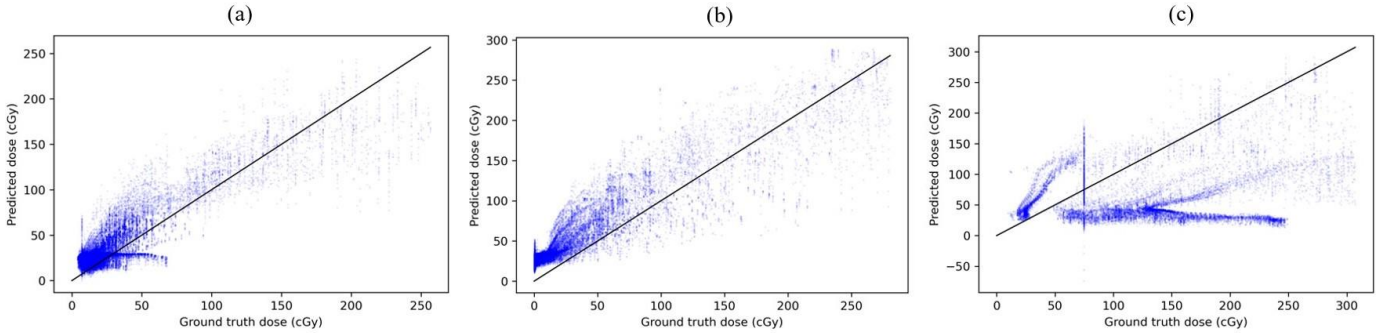

Fig 5. Doses predicted by the 3D U-Net as a function of the ground truth doses. Each point corresponds to a voxel wise dose evaluation. Scatter plots were drawn for Patient a (a), Patient b (b) and Patient c (c).

In additional deep learning experiments, we sought to evaluate the benefits of adding betatron data to the training set, to test its influence on the model's performance on other betatron data, and to test the development of a machine-specific method. For the first deep learning experiment, a mean RMSD of 0.26 ± 0.11 $cGy.Gy^{-1}$ was obtained on the 46 betatron test patients as a benchmark value. The model retrained on the original patient training set plus the 46 betatron patients kept aside, resulted in a mean RMSD of 0.64 ± 0.41 $cGy.Gy^{-1}$. For the second deep learning 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 Alcyon cobalt irradiator training set

patients, and configuration considering the original model fine-tuned with the 100 Alcyon cobalt irradiator training set patients, respectively.

Discussion

Our aim in this proof of concept was to demonstrate the feasibility of out-of-field dose prediction for high-energy photon irradiations using a deep learning neural network, while demonstrating that this approach is an appropriate response to the limitations in terms of computing time, difficulties in accessing experimental measurements and the lack of versatility inherent in analytical and MC methods, which ultimately limit access to out-of-field dose maps for routine clinical use.

Firstly, the results were analyzed in terms of RMSD, which is the most used metric for assessing out-of-field doses[31]. Conventionally, RMSD values were calculated in the training, validation and test sets, in order to assess the ability of the algorithm to generalize to the anatomy of new patients, new tumor locations, and new irradiation geometries, for new irradiation devices. 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 $cGy.Gy^{-1}$) (Table 3), suggesting that the neural network has acquired a strong degree of robustness and generalization. However, a closer look at the results for the different test sets shows that significantly lower results were observed for test set 5, which corresponds to the betatron accelerator. The same observation applies by examining the differences in median absolute doses (Table 4) per patient, where the betatron test set was associated with the highest median absolute error (0.33 Gy). In addition, most of the poorest results in the validation set concerned patients treated with the Sagittaire accelerator operating at 25 MV. As a reminder, the highest voltage of the linacs considered in the training set was equal to 20 MV (Supplemental Table 1). These results logically highlight the fact that the generalization capabilities of the neural network cannot be extended to non-conventional or highly atypical linacs, as long as they have not been considered during the training phase, i.e. with different shielding properties and internal geometries. Similarly, the ability to generalize is limited at very high voltages, as this implies in particular an increase in the pair production cross-section. For example, the betatron device,

which more closely resembles cyclotron systems than conventional devices, and the unique very high-voltage operation of the Sagittaire facility in our dataset contribute to this limited generalization capability.

The first additional deep learning experiment shows that adding betatron patients to the training set resulted in an unexpected decline in performance for the betatron device (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 small number of betatron dose maps in the training set was insufficient to allow the neural network to converge towards a solution tailored to these accelerators, and that these accelerators may be too unusual and confuse the network. The results obtained from the second additional deep learning experiment demonstrated that the model developed from 100 Alcyon cobalt irradiator patients showed poorer averaged performance on the 28 Alcyon cobalt irradiator test patients than the original model, trained without any Alcyon cobalt irradiator 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 irradiation device, but with a smaller number of patients cases 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.

Despite achieving better $RMSD$ results in areas distant from the field compared to those near the field (Table 3), the neural network does not seem to correctly recover very low dose values as depicted in Figures 3, 4, and 5. This consistent pattern observed in all dose maps generated by the neural network can be explained by two factors. Firstly, the chosen loss function (MSE) tends to minimize the absolute differences in dose between the predicted values and the ground truth. While some dose differences close to the field can be tolerated, because they are moderate in terms of relative differences,

these same differences become more problematic when it comes to predicting very low doses far from the field. Secondly, it is reasonable to assume that near-field doses, mainly influenced by the patient scatter component, are comparatively easier to predict for the network than the other components. In fact, this component depends mainly on the dose within the irradiation field and the irradiated volume, which is information readily available in the in-field dose map. On the other hand, very low doses in remote regions are mainly influenced by the leakage component, which is certainly only very partially present in the input data, apart from the fact that it is a signature of the irradiation device. This observation also explains the association between larger irradiation fields and improved values (Table 3), thanks to more usable information available.

Some publications using analytical models for out-of-field dose estimation reported $RMSD$ results similar to the performance achieved by our neural network (such as 0.91 $cGy.Gy^{-1}$ and 1.67 $cGy.Gy^{-1}$ [58], 0.75 $cGy.Gy^{-1}$ [55], 4.1 $cGy.Gy^{-1}$, 5.6 $cGy.Gy^{-1}$, 4.6 $cGy.Gy^{-1}$ and 6.5 $cGy.Gy^{-1}$ [54], 1.04 $cGy.Gy^{-1}$ [56], 3.7 $mGy.Gy^{-1}$ [59], 0.094 $cGy.MU^{-1}$, 0.279 $cGy.MU^{-1}$, and 0.410 $cGy.MU^{-1}$ [57]). We note that our average test result (0.32 ± 0.15 $cGy.Gy^{-1}$) is in the middle of the performance range. However making a direct comparison may be complex. Indeed, apart from the differences in the normalization process employed, which is a general problem in the context of out-of-field dose evaluation [31], our analysis involves comparing the predictions of a deep learning neural network to out-of-field dose maps obtain from analytical computation, while the studies previously mentioned compared the predictions of analytical models with experimental measurements or MC simulation. The next stage of our work will therefore be to carry out experimental measurements in order to compare them with the predictions of the network.

Admittedly, in this work, the analysis of the aleatoric and epistemic uncertainties of the neural network has not been undertaken [60]. Indeed, given that our learning database is based on analytical computations derived from experimental measurements, it is reasonable to assume that the uncertainties associated with these experimental measurements will have the most impact on the uncertainty associated with dose predictions. According to Vu Bezin et al, the minimum overall uncertainty in out-of-field dose measurements, whatever the detector used, is of the order of 15% [61]. A future study will

involve combining the uncertainties associated with the ground-truth databases with the uncertainties of the network itself, in order to be able to associate final uncertainties with the presented results.

To further enhance the capabilities of the neural network, it would be advantageous to have a database that encompasses even more diverse data, including information from recent radiotherapy treatments involving recent linacs and intensity modulation irradiation techniques. Indeed, a first limitation of this work is that the database used only contains patients treated with older techniques and irradiation devices, not representative of current patient care. Also, the analytical method used to estimate the whole-body dose maps used as ground truth could also be refined, given that analytical methods have evolved since then, especially recently [31]. The use of MC simulation with our modern hardware capabilities also now seems to be a good candidate for the generation of a learning database. Finally, even if simple estimation of out-of-field dose can be sufficient when studying integral dose trend, a local and more precise estimation can be needed. Indeed, the significance of a high resolution required by an out-of-field dose estimation tool is contingent upon the specific medical question at hand, as for instance when studying apparition and severity of radio-induced lymphopenia and dose to lymph nodes closest to the treatment field. For instance, advancements in neural network strategies beyond our current architecture, such as adversarial auto-encoder, successfully used for extension of anatomopathological whole slide images [62], hold potential for further refining our approach.

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 simulation 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. Almberg, S. S., Frengen, J. & Lindmo, T. Monte Carlo study of in-field and out-of-field dose distributions from a linear accelerator operating with and without a flattening-filter. *Med. Phys.* **39**, 5194–5203 (2012).
2. Howell, R. M., Scarboro, S. B., Kry, S. F. & Yaldo, D. Z. Accuracy of out-of-field dose calculations by a commercial treatment planning system. *Phys. Med. Biol.* **55**, 6999–7008 (2010).
3. Howell, R. M. *et al.* Methodology for determining doses to in-field, out-of-field and partially in-field organs for late effects studies in photon radiotherapy. *Phys. Med. Biol.* **55**, 7009–7023 (2010).
4. Sánchez-Nieto, B., López-Martínez, I. N., Rodríguez-Mongua, J. L. & Espinoza, I. A simple analytical model for a fast 3D assessment of peripheral photon dose during coplanar isocentric photon radiotherapy. *Front. Oncol.* **12**, (2022).
5. Schneider, C. W., Newhauser, W. D., Wilson, L. J. & Kapsch, R.-P. A physics-based analytical model of absorbed dose from primary, leakage, and scattered photons from megavoltage radiotherapy with MLCs. *Phys. Med. Biol.* **64**, 185017 (2019).
6. Wang, L. & Ding, G. X. The accuracy of the out-of-field dose calculations using a model based algorithm in a commercial treatment planning system. *Phys. Med. Biol.* **59**, N113-128 (2014).
7. Goy, E. *et al.* The out-of-field dose in radiation therapy induces delayed tumorigenesis by senescence evasion. *eLife* **11**, e67190 (2022).
8. Lumniczky, K. *et al.* Low dose ionizing radiation effects on the immune system. *Environ. Int.* **149**, 106212 (2021).
9. Paganetti, H. A review on lymphocyte radiosensitivity and its impact on radiotherapy. *Front. Oncol.* **13**, (2023).
10. Little, M. P. *et al.* Ionising radiation and cardiovascular disease: systematic review and meta-analysis. *BMJ* **380**, e072924 (2023).
11. Narasimhamurthy, R. K., Mumbrekar, K. D. & Satish Rao, B. S. Effects of low dose ionizing radiation on the brain- a functional, cellular, and molecular perspective. *Toxicology* **465**, 153030 (2022).
12. Hauptmann, M. *et al.* Brain cancer after radiation exposure from CT examinations of children and young adults: results from the EPI-CT cohort study. *Lancet Oncol.* **24**, 45–53 (2023).
13. 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) doi:10.1002/acm2.13182.
14. Mutic, S. & Low, D. A. Whole-body dose from tomotherapy delivery. *Int. J. Radiat. Oncol. Biol. Phys.* **42**, 229–232 (1998).
15. Tao, Y. *et al.* Comparison of dose contribution to normal pelvic tissues among conventional, conformal and intensity-modulated radiotherapy techniques in prostate cancer. *Acta Oncol. Stockh. Swed.* **47**, 442–450 (2008).
16. 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.* **62**, 9039–9053 (2017).
17. De Bruin, M. L. *et al.* Breast Cancer Risk in Female Survivors of Hodgkin's Lymphoma: Lower Risk After Smaller Radiation Volumes. *J. Clin. Oncol.* **27**, 4239–4246 (2009).
18. Hall, E. J. Intensity-modulated radiation therapy, protons, and the risk of second cancers. *Int. J. Radiat. Oncol. Biol. Phys.* **65**, 1–7 (2006).
19. Hall, E. J. & Wuu, C.-S. Radiation-induced second cancers: the impact of 3D-CRT and IMRT. *Int. J. Radiat. Oncol. Biol. Phys.* **56**, 83–88 (2003).
20. Jahreiß, M.-C. *et al.* The Risk of Second Primary Cancers in Prostate Cancer Survivors Treated in the Modern Radiotherapy Era. *Front. Oncol.* **10**, (2020).
21. Xiang, M., Chang, D. T. & Pollom, E. L. Second cancer risk after primary cancer treatment with three-dimensional conformal, intensity-modulated, or proton beam radiation therapy. *Cancer* **126**, 3560–3568 (2020).

22. Hall, E. J. & Brenner, D. J. In reply to Drs. Macklis Gottschalk, Paganetti, et al. *Int. J. Radiat. Oncol. Biol. Phys.* **66**, 1595 (2006).
23. de Kermenguy, F. *et al.* Radio-induced lymphopenia in the era of anti-cancer immunotherapy. in *International Review of Cell and Molecular Biology* (Academic Press, 2023). doi:10.1016/bs.ircmb.2023.03.002.
24. Lambin, P. *et al.* Lymphocyte-Sparing Radiotherapy: The Rationale for Protecting Lymphocyte-rich Organs When Combining Radiotherapy With Immunotherapy. *Semin. Radiat. Oncol.* **30**, 187–193 (2020).
25. Cai, S. *et al.* Impact of Radiation Dose to Circulating Immune Cells on Tumor Control and Survival in Esophageal Cancer. *Cancer Biother. Radiopharm.* (2021) doi:10.1089/cbr.2021.0250.
26. Heylmann, D., Ponath, V., Kindler, T. & Kaina, B. Comparison of DNA repair and radiosensitivity of different blood cell populations. *Sci. Rep.* **11**, 2478 (2021).
27. Nakamura, N., Kusunoki, Y. & Akiyama, M. Radiosensitivity of CD4 or CD8 positive human T-lymphocytes by an in vitro colony formation assay. *Radiat. Res.* **123**, 224–227 (1990).
28. Colnot, J. *et al.* A new Monte Carlo model of a Cyberknife® system for the precise determination of out-of-field doses. *Phys. Med. Biol.* **64**, 195008 (2019).
29. Sánchez-Nieto, B., Medina-Ascanio, K. N., Rodríguez-Mongua, J. L., 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) doi:10.1002/mp.14356.
30. Schneider, U. *et al.* Accuracy of out-of-field dose calculation of tomotherapy and cyberknife treatment planning systems: a dosimetric study. *Z. Med. Phys.* **24**, 211–215 (2014).
31. Benzazon, N. *et al.* Analytical models for external photon beam radiotherapy out-of-field dose calculation: a scoping review. *Front. Oncol.* **13**, (2023).
32. Shi, F. *et al.* Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nat. Commun.* **13**, 6566 (2022).
33. McIntosh, C. *et al.* Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat. Med.* **27**, 999–1005 (2021).
34. Lerner, M., Medin, J., Jamtheim Gustafsson, C., Alkner, S. & Olsson, L. E. Prospective Clinical Feasibility Study for MRI-Only Brain Radiotherapy. *Front. Oncol.* **11**, 812643 (2021).
35. Huynh, E. *et al.* Artificial intelligence in radiation oncology. *Nat. Rev. Clin. Oncol.* **17**, 771–781 (2020).
36. Kase, K. R., Svensson, G. K., Wolbarst, A. B. & Marks, M. A. Measurements of dose from secondary radiation outside a treatment field. *Int. J. Radiat. Oncol. Biol. Phys.* **9**, 1177–1183 (1983).
37. Benadjaoud, M. A. *et al.* A multi-plane source model for out-of-field head scatter dose calculations in external beam photon therapy. *Phys. Med. Biol.* **57**, 7725–7739 (2012).
38. Kry, S. F. *et al.* AAPM TG 158: Measurement and calculation of doses outside the treated volume from external-beam radiation therapy. *Med. Phys.* **44**, e391–e429 (2017).
39. Veres, C. *et al.* Retrospective reconstructions of active bone marrow dose-volume histograms. *Int. J. Radiat. Oncol. Biol. Phys.* **90**, 1216–1224 (2014).
40. Bewley, D., Bradshaw, A. & Burns, J. Central axis depth dose data for use in radiotherapy. *Br J Radiol* **4**, (1983).
41. Diallo, I. *et al.* Estimation of the radiation dose delivered to any point outside the target volume per patient treated with external beam radiotherapy. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* **38**, 269–271 (1996).
42. Francois, P., Beurtheret, C. & Dutreix, A. Calculation of the dose delivered to organs outside the radiation beams. *Med. Phys.* **15**, 879–883 (1988).
43. Francois, P., Beurtheret, C., Dutreix, A. & De Vathaire, F. A mathematical child phantom for the calculation of dose to the organs at risk. *Med. Phys.* **15**, 328–333 (1988).

44. Allodji, R. S. *et al.* Risk of Subsequent Leukemia After a Solid Tumor in Childhood: Impact of Bone Marrow Radiation Therapy and Chemotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **93**, 658–667 (2015).
45. Diallo, I. *et al.* Frequency distribution of second solid cancer locations in relation to the irradiated volume among 115 patients treated for childhood cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **74**, 876–883 (2009).
46. de Vathaire, F., François, P., Schweisguth, O., Oberlin, O. & Le, M. G. Irradiated neuroblastoma in childhood as potential risk factor for subsequent thyroid tumour. *Lancet Lond. Engl.* **2**, 455 (1988).
47. Tukenova, M. *et al.* Radiation therapy and late mortality from second sarcoma, carcinoma, and hematological malignancies after a solid cancer in childhood. *Int. J. Radiat. Oncol. Biol. Phys.* **80**, 339–346 (2011).
48. Byrne, J. *et al.* The PanCareSurFup consortium: research and guidelines to improve lives for survivors of childhood cancer. *Eur. J. Cancer* **103**, 238–248 (2018).
49. *Radiation Dose in Radiotherapy from Prescription to Delivery*. (INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 1996).
50. Cardoso, M. J. *et al.* MONAI: An open-source framework for deep learning in healthcare. Preprint at https://doi.org/10.48550/arXiv.2211.02701 (2022).
51. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Preprint at https://doi.org/10.48550/arXiv.1606.06650 (2016).
52. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
53. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015).
54. De Saint-Hubert, M. *et al.* Out-of-field doses from pediatric craniospinal irradiations using 3D-CRT, IMRT, helical tomotherapy and electron-based therapy. *Phys. Med. Biol.* **62**, 5293–5311 (2017).
55. Gallagher, K. J. *et al.* Supplemental computational phantoms to estimate out-of-field absorbed dose in photon radiotherapy. *Phys. Med. Biol.* **63**, 025021 (2018).
56. Jagetic, L. J. & Newhauser, W. D. A simple and fast physics-based analytical method to calculate therapeutic and stray doses from external beam, megavoltage x-ray therapy. *Phys. Med. Biol.* **60**, 4753–4775 (2015).
57. Sanchez, B. *et al.* Analytical model for photon peripheral dose estimation in radiotherapy treatments Analytical model for photon peripheral dose estimation in radiotherapy treatments. *Biomed. Phys. Eng. Express* **1**, 045205 (2015).
58. Taddei, P. J. *et al.* Analytical model for out-of-field dose in photon craniospinal irradiation. *Phys. Med. Biol.* **58**, 7463–7479 (2013).
59. Wilson, L. J. *et al.* Method to quickly and accurately calculate absorbed dose from therapeutic and stray photon exposures throughout the entire body in individual patients. *Med. Phys.* **47**, 2254–2266 (2020).
60. Abdar, M. *et al.* A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion* **76**, 243–297 (2021).
61. Vũ Bezin, J. *et al.* A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose-response relationships. *J. Radiol. Prot. Off. J. Soc. Radiol. Prot.* **37**, R1–R18 (2017).
62. Boyd, J. *et al.* Self-Supervised Representation Learning using Visual Field Expansion on Digital Pathology. in 639–647 (IEEE Computer Society, 2021). doi:10.1109/ICCVW54120.2021.00077.