Uncertainty-Aware Out-of-Field Dose Prediction in External Beam Radiotherapy

Mohammed El Aichi[1,2,4], Nathan Benzazon[1,2], Meïssane M'Hamdi[1,2], Rodrigue Allodji[3], Florent de Vathaire[3], Eric Deutsch[1,2], Ibrahima Diallo[1,2], Maria Vakalopoulou[4], and Charlotte Robert[1,2⋆]

[1] U1355, Radiothérapie moléculaire et innovations thérapeutiques, Université Paris-Saclay, Gustave Roussy, INSERM, Villejuif, France
[2] Département de Radiothérapie, Gustave Roussy, Villejuif, France
[3] U1018, Épidémiologie des radiations, INSERM, Gustave Roussy, Université Paris-Saclay, Villejuif, France
[4] U1361, Cancer Data Science, Université Paris-Saclay, Gustave Roussy, CentraleSupélec, INSERM, Villejuif, France.

Abstract. Low and very low radiation doses delivered outside the treatment field in radiotherapy can affect the immune system and contribute to radiation-induced lymphopenia, which is associated with poorer clinical outcomes. This is particularly relevant in radio-immunotherapy settings, where preservation of immune function may directly impact therapeutic efficacy. In this work, we propose an energy-conditioned deep learning framework for voxel-wise out-of-field (OOF) dose prediction from the planned in-field dose and whole-body mask. Beam energy is incorporated through additive conditioning, and the deterministic model is further extended to a heteroscedastic mean variance formulation for voxel-wise uncertainty estimation. Our framework is developed and evaluated on 1,519 reconstructed whole-body dose distributions from retrospective photon-based LINAC treatment records. The best-performing model, an energy-conditioned U-Net, achieved an MAE of 9.76±9.28 cGy and an RMSD of 15.61 ± 12.50 cGy. Energy conditioning significantly improved the U-Net baseline, while mean variance modeling provided uncertainty estimates positively correlated with prediction errors. These results demonstrate the potential of energy-conditioned deep learning for rapid and uncertainty-aware whole-body OOF dose estimation. Code is available at https://github.com/maichi98/DoseFieldNet.

Keywords: Out-of-field dose prediction · Radiotherapy · Energy conditioning · Uncertainty estimation

1 Introduction

Since its first clinical application more than a century ago, external beam radiation therapy (EBRT) has become a cornerstone of cancer treatment, with

⋆ Corresponding author: CH.ROBERT@gustaveroussy.fr

roughly 50% of patients undergoing it at some point over the trajectory of their disease [1]. Although modern delivery techniques achieve highly conformal dose distributions, photon transport prevents strict dose confinement to the target volume. In addition to the intended in-field dose, scattering processes within the treatment head and patient, along with leakage from beam shaping components, produce a low-dose distribution throughout the body, known as the *out-of-field* (OOF) dose. These lower doses–defined in AAPM TG-158 [8] as < 3 Gy or $< 5\%$ of the prescribed dose–are generally ignored in treatment planning. However, such unintended exposure has been linked to long-term health effects, including secondary malignancies, cardiac toxicity, and cataracts [8]. The EPI-CT cohort study reported an increased risk of brain cancer after cumulative CT doses of about 50 mGy [5]. These results suggest that radiation levels comparable to OOF doses during photon EBRT may be clinically relevant, underscoring the need to explicitly consider them when optimizing treatment plans and modeling radiation-induced risks. OOF dose remains poorly estimated by commercial treatment planning systems (TPS). Most TPS dose calculation algorithms are specifically optimized for high-dose regions near the target volume and generally neglect distant low-dose deposition. As demonstrated by [6], commercial TPS tend to substantially underestimate OOF dose, with prediction errors increasing as the distance from the treatment field grows.

In this work, we propose a deep learning framework for rapid 3D OOF dose estimation from the planned in-field dose and whole-body binary mask. Our contributions are threefold: *(i)* we propose beam-energy conditioning to account for energy-dependent OOF dose deposition and demonstrate its benefit; *(ii)* we investigate alternative transformations of the target dose to perform learning in a transformed prediction space better suited for low-dose estimation; and *(iii)* we extend the framework to a heteroscedastic mean variance formulation for voxel-wise uncertainty estimation.

2 Related Work

OOF dose estimation has traditionally relied on Monte Carlo (MC) simulations and analytical models. MC methods explicitly simulate photon transport and secondary electron production within both the linear accelerator head and the patient. However, implementing MC simulations requires vendor-specific information about the accelerator head geometry and material composition, together with careful validation against experimental measurements. OOF dose calculation is particularly challenging with MC simulations because only a small fraction of simulated photon histories contribute to dose deposition outside the treatment field, requiring a large number of histories to get an acceptable statistical uncertainty. Recent variance-reduction methods, such as Last Vertex Splitting combined with a hybrid Track Length Estimator, have reduced computation times by approximately one order of magnitude in patient simulations [7]. However, patient-specific simulations still require several hours on high-performance computers, precluding their use for treatment plan optimization. Analytical ap-

proaches estimate OOF dose using empirical formulations that either represent the total dose through a single-compartment model or describe its main physical contributions separately through a multi-compartment model [2]. Periphocal 3D [10] is an example of a fast analytical model for photon peripheral dose estimation, designed to generate 3D whole-body dose distributions and organ dose estimates outside the 5% isodose for coplanar isocentric treatments. However, analytical models often require experimental measurements to adjust intrinsic machine-specific parameters, which can limit their use in large retrospective studies involving legacy LINACs that are no longer available for measurements. Deep learning offers a promising alternative for rapid OOF dose estimation. Once trained, a neural network can directly infer the whole-body OOF dose distribution from routinely available treatment information, enabling fast predictions suitable for retrospective studies and potential clinical integration. To the best of our knowledge, the only previous work predicting OOF dose from the in-field dose using deep learning is the study of Benzazon et al. [3], which demonstrated feasibility using downsampled dose distributions ($11.6 \times 7.6 \times 8.8$ mm^3), with a mean absolute error (MAE) of 22.14 cGy and a root mean square deviation (RMSD) of 27.20 cGy. Our study goes beyond this by predicting OOF dose at a finer output resolution ($4 \times 4 \times 4$ mm^3), incorporating beam-energy conditioning to account for energy-dependent dose deposition, and providing voxel-wise uncertainty estimation through a heteroscedastic mean variance framework.

3 Proposed Method

3.1 Problem Formulation

For each treatment record, we consider a 3D whole-body dose distribution D_{WB}, a binary whole-body mask M_{WB}, the prescribed dose D_{pres}, and the beam energy E. Following prior work [11], body voxels receiving at least 5% of the prescribed dose are defined as in-field voxels, and the remaining body voxels are defined as OOF voxels. We denote these regions by Ω_{in} and Ω_{out}, respectively.

The base network input consists of the normalized in-field dose $\tilde{D}_{\mathrm{in}}$ and the whole-body mask M_{WB}. The normalized in-field dose is obtained by keeping only the in-field portion of D_{WB} and dividing it by D_{pres}. For energy-conditioned models, the beam energy E is provided as an additional input. The prediction target is the OOF dose normalized by $0.05 D_{\mathrm{pres}}$, denoted by $\tilde{D}_{\mathrm{out}}$.

3.2 Deterministic OOF Dose Prediction

We formulate OOF dose estimation as a 3D image-to-image regression problem. Given the normalized in-field dose, the whole-body mask, and, for energy-conditioned models, the beam energy, a neural network f_θ predicts the normalized OOF dose:

$$\hat{\tilde{D}}_{\mathrm{out}} = f_\theta\left(\tilde{D}_{\mathrm{in}}, M_{\mathrm{WB}}; E\right). \tag{1}$$

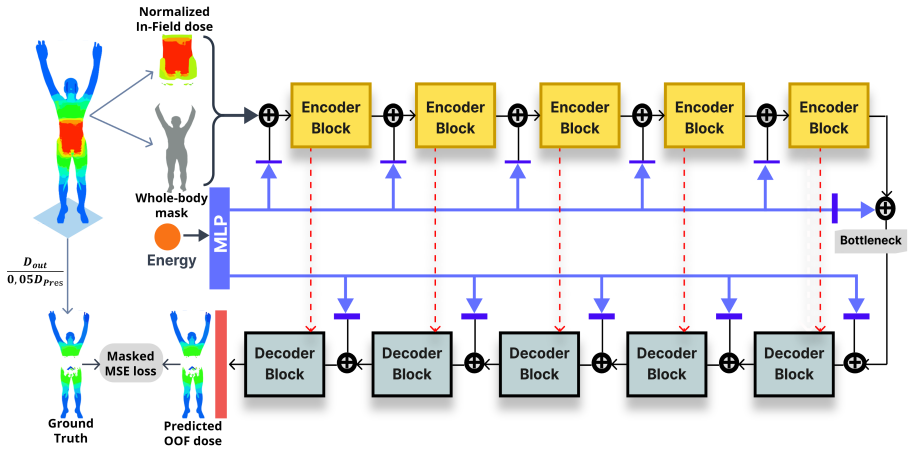

Fig. 1: Overview of the deterministic energy-conditioned OOF dose prediction framework. The model receives the normalized in-field dose and whole-body mask as input channels, while beam energy is embedded with an MLP and injected at multiple encoder–decoder resolution levels. The network predicts the normalized OOF dose and is trained with a masked MSE loss over the OOF region.

We evaluate two encoder–decoder backbones: a 3D U-Net and the transformer-based Swin-UNETR [4]. Both architectures take the normalized in-field dose and whole-body mask as input channels and output a 3D OOF dose prediction. The network is trained using a masked mean squared error (MSE) over the OOF voxels Ω_{out}:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|\Omega_{\text{out}}|} \sum_{v \in \Omega_{\text{out}}} \left(\hat{\tilde{D}}_{\text{out}}^{(v)} - \tilde{D}_{\text{out}}^{(v)} \right)^2 . \tag{2}$$

For the final prediction, we average the outputs of the five models trained during cross-validation. Fig. 1 summarizes the deterministic energy-conditioned framework.

3.3 Beam-Energy Conditioning

To condition the network on the scalar beam energy E, the energy value is first normalized using min–max scaling and embedded using a multilayer perceptron (MLP) $\phi(E)$. The energy embedding is injected into each resolution level of the encoder–decoder network. At level ℓ, the embedding is linearly projected to match the number of feature channels and added to the feature representation as $\mathbf{F}_\ell^{\text{cond}} = \mathbf{F}_\ell + \mathbf{W}_\ell \phi(E)$, where $\mathbf{F}_\ell$ denotes the feature map at level ℓ and $\mathbf{W}_\ell$ is a learned projection. The projected embedding is broadcast over the spatial dimensions before addition.

3.4 Target Transformations

Since the OOF dose spans several orders of magnitude, direct regression can be challenging, especially in very low-dose regions. We therefore investigate two

target transformations: a square-root transformation, $T_{\text{sqrt}}(x) = \sqrt{x}$, and a logarithmic transformation, $T_{\log}(x) = \log(x + \epsilon)$, where ϵ is a small constant added for numerical stability. When a transformation T is used, the network is trained to predict $T(\tilde{D}_{\text{out}})$, and the masked MSE loss is computed in the transformed space over Ω_{out}. Finally, the network outputs are converted back to physical dose values.

3.5 Heteroscedastic Mean Variance Estimation

To estimate voxel-wise predictive uncertainty, we extend the deterministic model to a heteroscedastic mean variance formulation. Starting from a trained deterministic model, we add a second output head that predicts the log-variance in addition to the mean normalized OOF dose. The model therefore predicts a mean $\hat{\mu}$ and log-variance $\hat{s} = \log \hat{\sigma}^2$ at each OOF voxel.

The model defines a Gaussian predictive distribution and is trained using a masked negative log-likelihood loss:

$$\mathcal{L}_{\text{NLL}} = \frac{1}{|\Omega_{\text{out}}|} \sum_{v \in \Omega_{\text{out}}} \left[\frac{1}{2} \exp\left(-\hat{s}^{(v)}\right) \left(\hat{\mu}^{(v)} - \tilde{D}_{\text{out}}^{(v)}\right)^2 + \frac{1}{2} \hat{s}^{(v)} \right]. \tag{3}$$

For the mean variance models, the final OOF dose prediction is obtained by averaging the predicted means from the five cross-validation models. The ensemble uncertainty is computed using the total variance:

$$\left(\hat{\sigma}^2\right)^{\text{ens},(v)} = \frac{1}{5} \sum_{k=1}^{5} \left[\exp\left(\hat{s}_k^{(v)}\right) + \left(\hat{\mu}_k^{(v)}\right)^2 \right] - \left(\hat{\mu}^{\text{ens},(v)}\right)^2. \tag{4}$$

4 Experimental Setup

4.1 Dataset and Preprocessing

Our dataset consists of retrospective radiotherapy treatment records from the French Childhood Cancer Survivor Study (FCCSS), a national cohort established to investigate the long-term effects of cancer treatments in children and adolescents. The cohort includes patients diagnosed with solid tumors or lymphomas before 21 years of age, with 7,205 radiotherapy treatment records collected between 1945 and 2000. Each record provides patient-body voxel coordinates and corresponding dose values at $2 \times 2 \times 2$ mm^3 resolution, defining the whole-body mask M_{WB} and aligned dose distribution D_{WB}. For our experiments, we retained photon treatments delivered using high-voltage linear accelerators (> 1 MV), resulting in $N = 1{,}519$ dose records. Data were split into training/validation (85%) and held-out test (15%) sets, with five-fold cross-validation used for model training and selection. Splits were stratified by treatment machine, beam energy, and patient body volume. Finally, dose volumes were resampled to $4 \times 4 \times 4$ mm^3.

4.2 Implementation Details

Backbone architectures: For the 3D U-Net, we used five resolution levels with feature widths $(32, 64, 128, 256, 320)$, instance normalization, and a dropout rate of 0.2. For Swin-UNETR, we used a patch size of $2 \times 2 \times 2$, an embedding dimension of 36, hierarchical depths $(2, 2, 2, 2)$, a window size of $7 \times 7 \times 7$, multi-head self-attention with $(3, 6, 12, 24)$ heads, an MLP ratio of 4.0, dropout of 0.2, and a drop-path rate of 0.1.
Evaluated configurations: We evaluated both backbones with and without energy conditioning. Energy-conditioned models are denoted with the prefix c, resulting in cU-Net and cSwin-UNETR. For cU-Net, we additionally evaluated square-root and logarithmic target transformations, denoted cU-Net + sqrt and cU-Net + log. Finally, we evaluated mean variance variants, denoted cU-Net + MV and cSwin-UNETR + MV. For energy-conditioned models, the beam energy embedding dimension was set to 16, and energy conditioning was randomly dropped during training with probability 0.5. For mean variance models, the predicted log-variance was constrained to the interval $[-8, -2]$ to stabilize training with the Gaussian negative log-likelihood loss and prevent the model from assigning unrealistically small or large variances.
Data augmentation: Training samples were augmented using random flipping, translations, random affine rotations around the head-to-toe axis with a maximum angle of $\pm 10^{\circ}$, and partial body cropping while preserving the in-field region. We also used dose-gradient-guided elastic deformations, where local displacement magnitudes were constrained according to the dose gradient so that deformations remained consistent with a local 3%/3 mm gamma criterion. This prevents the augmentation from introducing physically implausible dose changes in high-gradient regions.
Training: Gradient accumulation over 8 mini-batches was used, with unreduced voxel-wise losses normalized by the total number of OOF voxels across accumulation steps. Deterministic models were trained for 800 epochs using automatic mixed precision. Mean variance models were initialized from the corresponding deterministic models and fine-tuned for 200 additional epochs using the masked Gaussian negative log-likelihood loss. Optimization was performed using AdamW [9] with an initial learning rate of 1×10^{-4} and weight decay of 1×10^{-4}. A cosine annealing schedule with 10 warmup epochs reduced the learning rate to 1×10^{-6} over training. Swin-UNETR based models were trained on NVIDIA H100 GPUs, and U-Net based models on NVIDIA A100 GPUs.

4.3 Evaluation Metrics

All evaluation metrics were computed in physical dose units after converting model outputs back from any normalization or target transformation. Metrics were evaluated within the OOF region Ω_{out} and computed separately for each test dose distribution. Results are reported as mean (SD) across test cases. Model performance was assessed using the MAE, RMSD, peak signal-to-noise ratio

(PSNR), and normalized MAE (NMAE), defined as MAE divided by the prescribed dose. To assess performance across dose levels, we additionally report the relative absolute error (REL) within three OOF dose ranges: 0.01–0.1%, 0.1–1%, and 1–5% of the prescribed dose. For the mean variance models, we report the Gaussian negative log-likelihood (NLL) computed over OOF voxels in physical dose units. We also report the Spearman rank correlation $\rho_s(|e|, \sigma)$ between the voxel-wise absolute prediction error $|e|$ and the predicted standard deviation σ. Organ-level performance was evaluated for selected organs involved in immune function or lymphocyte circulation. For each organ, we computed the relative error in mean organ dose, defined as $(D^{\text{pred}}_{\text{organ}} - D^{\text{GT}}_{\text{organ}})/D^{\text{GT}}_{\text{organ}}$.

4.4 Comparison with Prior OOF Dose Models

We compared our models with the deep learning model of Benzazon et al. [3] and the analytical Peripheral 3D model [10]. To ensure a fair comparison, all literature-model comparisons were restricted to the 25 dose distributions that were present in both our held-out test set and the test set of Benzazon et al. [3].

For the Benzazon et al. model, we used the trained model provided by the original study and resampled its predicted dose distributions to our $4\times4\times4$ mm^3 grid before evaluation. For Peripheral 3D, machine-specific model parameters were fitted using training dose distributions from each treatment machine. The fitted parameters were then applied to the corresponding test cases according to treatment machine. Both literature models were evaluated on the same 25 shared test cases using the metrics described above.

5 Results and Discussion

Table 1 reports voxel-wise performance across the evaluated architectures. Without energy conditioning, Swin-UNETR performed better than U-Net, achieving a lower MAE (10.50 vs. 11.00 cGy) and RMSD (15.92 vs. 16.64 cGy). However, energy conditioning substantially improved the U-Net baseline: cU-Net reduced the MAE to 9.76 cGy and the RMSD to 15.61 cGy (U-Net vs. cU-Net, paired Wilcoxon signed-rank test, $p < 0.001$). cU-Net also outperformed both Swin-UNETR and cSwin-UNETR in terms of MAE and RMSD.

Table 1: Performance comparison across architectures. Values are reported as mean (SD). Overall metrics are computed over all OOF voxels; REL is reported by dose range. (c: energy conditioning; sqrt/log: OOF dose transformations; MV: mean variance modeling).

	Overall Metrics				REL by Dose Range (%)		
Model	MAE (cGy)	NMAE	RMSD (cGy)	PSNR (dB)	0.01–0.1%	0.1–1%	1–5%
Swin-UNETR	10.50 (9.74)	0.281 (0.214)	15.92 (12.91)	22.94 (5.08)	332.53 (539.19)	57.47 (48.74)	**20.87** (7.98)
U-Net	11.00 (10.01)	0.292 (0.211)	16.64 (13.19)	22.53 (4.98)	330.83 (579.35)	59.55 (60.27)	21.79 (7.72)
cSwin-UNETR	10.39 (9.76)	0.280 (0.224)	16.22 (12.67)	22.74 (5.09)	277.43 (589.65)	52.97 (45.26)	21.94 (7.46)
cU-Net	**9.76** (9.28)	**0.261** (0.205)	**15.61** (12.50)	**23.13** (5.10)	258.77 (570.29)	48.94 (42.10)	21.11 (7.18)
cU-Net + sqrt	9.86 (10.15)	0.264 (0.225)	16.17 (13.29)	22.88 (5.19)	211.72 (568.79)	**42.59** (38.57)	22.00 (7.43)
cU-Net + log	11.32 (12.08)	0.304 (0.275)	18.32 (15.20)	21.82 (5.20)	**184.68** (497.39)	43.37 (40.33)	25.32 (8.71)
cU-Net + MV	9.79 (9.29)	0.262 (0.207)	15.70 (12.51)	23.09 (5.13)	279.96 (586.41)	48.20 (42.85)	21.12 (7.09)
cSwin-UNETR + MV	10.22 (9.71)	0.275 (0.224)	16.13 (12.54)	22.81 (5.11)	270.55 (587.40)	51.16 (44.82)	21.79 (7.16)

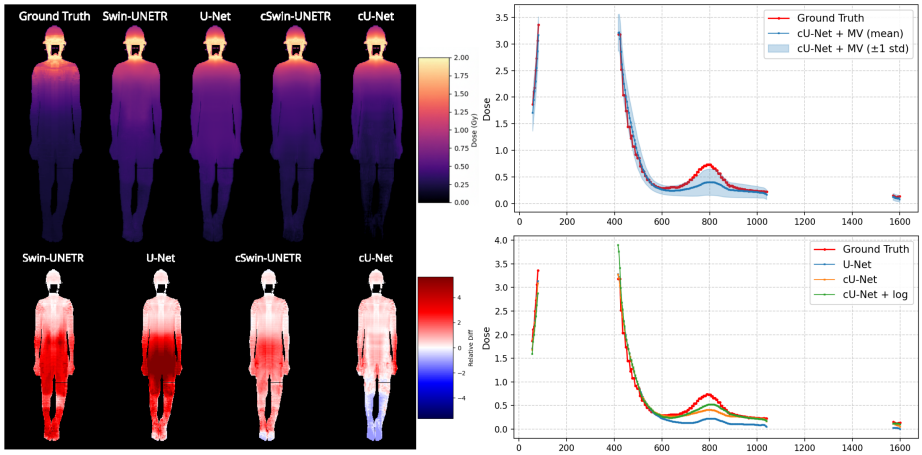

Fig. 2: Example OOF dose maps and dose profiles from two OOF predictions. **Left:** OOF dose prediction and relative error maps for U-Net, Swin-UNETR, cU-Net, and cSwin-UNETR. **Right:** Dose profiles from another OOF prediction with a secondary peak, showing cU-Net + MV with a ±1 SD uncertainty band (top) and the effect of conditioning and logarithmic transformation (bottom).

Figure 2 illustrates the effect of energy conditioning: in the left panel, cU-Net substantially improves predictions compared with U-Net, with a similar but less pronounced improvement observed for Swin-UNETR. The target transformations mainly improved prediction in the lower OOF dose ranges. In the 0.01–0.1% range, REL decreased from 258.77% for cU-Net to 211.72% for cU-Net + sqrt and 184.68% for cU-Net + log. However, these improvements came at the cost of reduced performance in the higher 1–5% range, where REL increased to 22.00% for cU-Net + sqrt and 25.32% for cU-Net + log, compared with 21.11% for cU-Net. This trade-off is visible in the right panel of Fig. 2: cU-Net improves the prediction compared with U-Net, and the logarithmic transformation further improves the lowest-dose regions, but introduces larger errors around higher-dose regions near the in-field boundary.

The mean variance models achieved dose prediction performance close to their deterministic counterparts while additionally providing voxel-wise uncertainty estimates. cU-Net + MV achieved an MAE of 9.79 cGy and an RMSD of 15.70 cGy, close to the deterministic cU-Net values of 9.76 cGy and 15.61 cGy.

Table 2 and Fig. 3 summarize the uncertainty evaluation. cU-Net + MV achieved a lower overall NLL than cSwin-UNETR + MV (−1.1165 vs. −1.0687), and also performed better in the 1–5% dose range (0.3922 vs. 0.4868). The uncertainty-error correlation was strongest in the lowest dose range, reaching $\rho = 0.7017$ for cU-Net + MV and $\rho = 0.7117$ for cSwin-UNETR + MV in the 0.01–0.1% range. However, the calibration curves show conservative uncertainty estimates in the lower dose ranges, with predicted intervals slightly wider than needed. In the 1–5% range, the correlation remains positive but lower, while the calibration curves are closer to the diagonal.

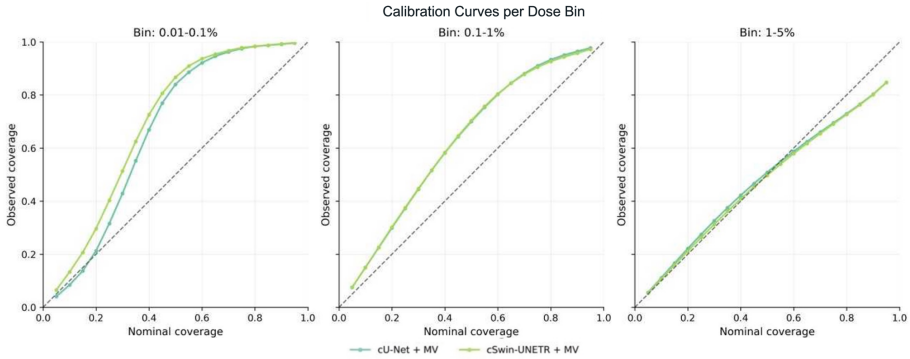

Fig. 3: Uncertainty calibration curves for cU-Net + MV and cSwin-UNETR + MV across three OOF dose intervals. The dashed diagonal line represents perfect calibration.

Table 2: Uncertainty evaluation of the mean variance models. NLL is the Gaussian negative log-likelihood computed over OOF voxels; ρ is the Spearman correlation between absolute prediction error and predicted uncertainty, $\mathrm{Corr}(|y - \hat{y}|, \sigma)$.

	All OOF Voxels		0.01–0.1%		0.1–1%		1–5%	
Model	NLL ↓	ρ ↑	NLL ↓	ρ ↑	NLL ↓	ρ ↑	NLL ↓	ρ ↑
cU-Net + MV	−1.1165	0.5939	−2.4269	0.7017	−1.4027	0.6359	0.3922	0.3907
cSwin-UNETR + MV	−1.0687	0.5930	−2.4752	0.7117	−1.3379	0.6567	0.4868	0.3496

We further compared our models with prior OOF dose prediction models using the protocol described in Sec. 4.4. As shown in Table 3, cU-Net + MV achieved the best overall performance, with an MAE of 7.67 cGy and an RMSD of 13.06 cGy. Both cU-Net variants substantially outperformed Benzazon et al. [3] and Periphocal 3D [10] in terms of MAE and RMSD.

Table 3: Comparison with prior OOF dose models on the 25 shared test cases. Values are mean (SD); REL is reported by dose range.

	Overall Metrics				REL by Dose Range (%)		
Model	MAE (cGy)	NMAE	RMSD (cGy)	PSNR (dB)	0.01–0.1%	0.1–1%	1–5%
Benzazon et al. [3]	24.28 (10.94)	0.671 (0.328)	66.38 (78.98)	12.74 (5.99)	1228.95 (1538.91)	207.66 (148.86)	51.90 (18.36)
Periphocal 3D [10]	20.16 (12.34)	0.601 (0.505)	32.01 (12.05)	15.75 (3.99)	430.35 (744.67)	119.06 (151.82)	63.97 (22.66)
cU-Net	7.81 (3.79)	0.206 (0.098)	13.17 (5.81)	23.95 (3.57)	186.19 (226.87)	46.90 (49.81)	**24.16** (5.63)
cU-Net + MV	**7.67** (3.56)	**0.202** (0.096)	**13.06** (5.77)	**24.07** (3.64)	**178.62** (208.51)	**45.73** (39.68)	24.23 (5.58)

Finally, Fig. 4 shows relative errors in mean dose for immune-related organs. For cU-Net and cU-Net + MV, most organ-level errors fall between −15% and 15%.

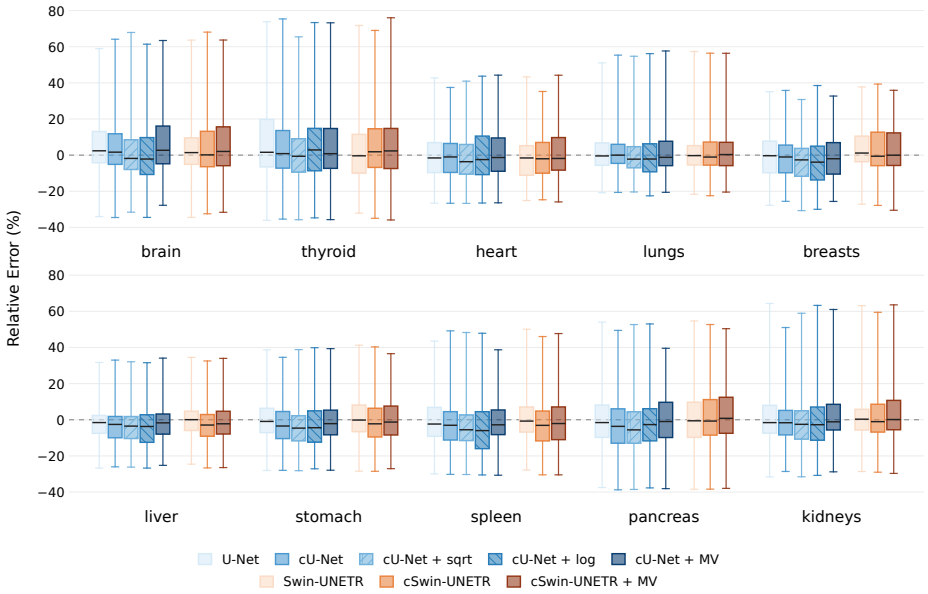

Fig. 4: Relative error distributions for mean dose in immune-related organs. Boxplots compare model-specific errors across the evaluated organs.

6 Conclusion

We proposed an energy-conditioned deep learning framework for rapid 3D OOF dose prediction from the planned in-field dose and whole-body mask. Energy conditioning substantially improved the U-Net baseline and allowed cU-Net to achieve the best overall performance among the evaluated models, with an MAE of 9.76 ± 9.28 cGy and an RMSD of 15.61 ± 12.50 cGy. Target transformations improved prediction in the lowest OOF dose ranges, although this came at the cost of reduced performance in higher-dose regions near the in-field boundary. The mean variance extension preserved dose prediction accuracy while providing voxel-wise uncertainty estimates. Finally, comparison with models from the literature showed that the proposed cU-Net variants substantially improved OOF dose prediction on the shared test cases. These results support energy-conditioned deep learning as a promising approach for fast, uncertainty-aware whole-body OOF dose estimation.

Acknowledgments The authors thank the U1355 research team at Gustave Roussy and the Biomathematics team at CentraleSupélec for their support. This research was partially funded by the French National Research Agency (ANR) under the France 2030 investment plan (grant ANR-21-RHUS-0005). This research has also received funding from the European Union's EURATOM Research and Training Programme under Grant Agreement No. 101061037 as part of the PIANOFORTE partnership (project DOSELIA, 2024–2029).

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Atun, R., Jaffray, D.A., Barton, M.B., Bray, F., Baumann, M., Vikram, B., Hanna, T.P., Knaul, F.M., Lievens, Y., Lui, R., et al.: Expanding global access to radiotherapy. The Lancet Oncology **16**(10), 1153–1186 (2015)
2. Benzazon, N., Colnot, J., de Kermenguy, F., et al.: Analytical models for external photon beam radiotherapy out-of-field dose calculation: a scoping review. Frontiers in Oncology **13**, 1197079 (2023). https://doi.org/10.3389/fonc.2023.1197079
3. Benzazon, N., Carré, A., de Kermenguy, F., Niyoteka, S., Maury, P., Colnot, J., M'hamdi, M., Aichi, M.E., Veres, C., Allodji, R., de Vathaire, F., Sarrut, D., Journy, N., Alapetite, C., Grégoire, V., Deutsch, E., Diallo, I., Robert, C.: Deep-learning for rapid estimation of the out-of-field dose in external beam photon radiation therapy - a proof of concept. International Journal of Radiation Oncology Biology Physics **120**(1), 253–264 (Sep 2024). https://doi.org/10.1016/j.ijrobp.2024.03.007
4. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images (2022)
5. Hauptmann, M., Daniels, R.D., Cardis, E., et al.: Radiation dose from paediatric ct scans and risk of brain cancer: a multinational cohort study. The Lancet Oncology **24**(1), 45–53 (2023)
6. 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. Physics in Medicine and Biology **55**(23), 6999–7008 (2010). https://doi.org/10.1088/0031-9155/55/23/015
7. Jacquet, M., Michel Létang, J., Baudier, T., Boissard, P., M'hamdi, M., Diallo, I., Robert, C., Sarrut, D.: Last vertex splitting: a new retroactive monte carlo splitting technique applied to linac out-of-field dose computation. Physics in Medicine Biology **70**(15), 155010 (aug 2025). https://doi.org/10.1088/1361-6560/adf1d2
8. Kry, S.F., Bednarz, B., Howell, R.M., Dauer, L., Followill, D., Klein, E., Paganetti, H., Wang, B., Wuu, C.S., George Xu, X.: Aapm tg 158: Measurement and calculation of doses outside the treated volume from external-beam radiation therapy. Medical Physics **44**(10), e391–e429 (2017). https://doi.org/10.1002/mp.12462
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2019)
10. 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. Frontiers in Oncology **12**, 872752 (2022). https://doi.org/10.3389/fonc.2022.872752
11. 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. Physics in Medicine Biology **59**(13), N113 (jun 2014). https://doi.org/10.1088/0031-9155/59/13/N113