TURKISH JOURNAL of ONCOLOGY

Turk J Oncol 2026;41(3):198–206
doi: 10.5505/tjo.2026.4983

ORIGINAL ARTICLE

Cross-Geometry Generalisation of a Physics-Informed 3D Dose-Prediction Ensemble

Muhammed Emin BEDİR,[1,2] Fuzuli TUĞRUL[3]

[1]Department of Electric and Energy, Karamanoğlu Mehmetbey University, Vocational School of Technical Sciences, Karaman- *Türkiye*
[2]Department of Medical Physics, University of Wisconsin, Madison-*USA*
[3]Department of Radiation Oncology, Acıbadem Eskişehir Hospital, Eskişehir-*Türkiye*

OBJECTIVE

Deep-learning dose-prediction models are often tied to their training beam configurations, limiting reusability. We tested whether an ensemble of three complementary 3D networks trained on one photon beam configuration can generalise to five unseen configurations on a held-out anatomy.

METHODS

Photon dose plans for five anatomies and six beam configurations were generated with matRad. A 13-channel input tensor (CT image, PTV/OAR masks, physics-informed per-beam ray-tracing maps) was built for every plan. Three networks (SOTA-UNet3D, Res-UNet3D, Trans-UNet3D) were trained on the training configuration (C0) of four anatomies and combined by voxel-wise mean as the Voltron ensemble. The held-out anatomy (TG-119) was evaluated under C0 and C1–C5 using the 3D gamma index (3%/3 mm) and mean absolute error (MAE).

RESULTS

Voltron achieved 51.42% gamma pass rate on C0 (MAE 8.91 Gy). Across the unseen configurations the mean was 52.49% (range 47.38%–56.94%; MAE 8.94–10.44 Gy). The largest decrease relative to C0 was 4.04 points (C1); three configurations exceeded it by 1.75–5.52 points. Performance on unseen geometries was comparable to the in-distribution baseline.

CONCLUSION

A physics-informed 3D dose-prediction ensemble trained on one beam configuration may retain its dosimetric quality on unseen co-planar geometries, suggesting an avenue for reusable models in data-scarce settings.

Keywords: Deep learning; dose distribution; generalization; neural networks; radiation dosimetry; radiotherapy.
Copyright © 2026, Turkish Society for Radiation Oncology

INTRODUCTION

Radiotherapy is a cornerstone of modern oncology, and three-dimensional dose prediction has emerged as a powerful tool for accelerating treatment planning and automating quality-assurance workflows.[1,2] Convolutional architectures based on the U-Net[3] and its three-dimensional extension[4] have become a de facto standard for voxel-wise dose prediction.[5,6] More recent work has extended these ideas with cascaded architectures,[7] structure-aware losses[8] and end-to-end planning pipelines.[2]

Received: June 29, 2026
Accepted: August 05, 2026
Online: September 10, 2026

Accessible online at:
www.onkder.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Dr. Muhammed Emin BEDİR
Karamanoğlu Mehmetbey Üniversitesi,
Elektrik ve Enerji Bölümü,
Teknik Bilimler Meslek Yüksekokulu,
Karaman-Türkiye
E-mail: meminbedir@kmu.edu.tr

Even when anatomy and prescription are fixed, the predicted dose distribution depends strongly on the beam geometry chosen by the planner. Beam-angle selection is itself a challenging combinatorial problem that has received considerable attention, including recent deep-reinforcement-learning approaches.[9] When dose-prediction networks are trained, the input representation must therefore convey the beam geometry; Barragán-Montero et al.[10] showed that encoding the beam configuration as an additional input channel allowed a single model to generalise across heterogeneous beam configurations in lung intensity-modulated radiotherapy. Beam-wise decomposition strategies further exploit the additivity of dose with respect to individual beams,[11,12] and physics-informed inputs that incorporate per-beam ray-tracing have been proposed to improve interpretability and accuracy.[12]

Despite these advances, most published dose-prediction models are trained and tested on plans that share a single beam geometry, or on data drawn from a single clinical site.[5,6] A practical question that follows is whether such models can be reused when the beam geometry departs from the one they were trained on. If the answer were yes, dose-prediction models could be more easily shared between centres that use different beam-angle conventions, which would lower the data burden on smaller centres.[13]

From a machine-learning perspective, this is a question of out-of-distribution generalisation. Recent reviews and benchmark studies in medical imaging have emphasised that deep models often degrade when the test-time distribution drifts from the training distribution.[14,15] Most published work on this problem in medical imaging concerns segmentation across scanners or hospitals, where the input distribution shifts but the task output is largely invariant. Dose prediction differs in that the output itself depends explicitly on the geometric input; the question becomes whether a network can correctly re-weight its predictions when the geometry changes.

The open-source toolkit matRad (matRad Research Group, German Cancer Research Center [DKFZ], Heidelberg, Germany) implements analytical pencil-beam algorithms for photon, scanned-proton and scanned-carbon-ion plans within a single MATLAB framework.[16] Because matRad can compute reference doses for arbitrary beam configurations under consistent algorithms, it provides a controlled testbed for studying robustness to beam-geometry variation, free of the inter-centre confounders that complicate clinical multi-institutional studies.

On the model side, residual connections[17] and attention mechanisms[18] have inspired numerous variants of the three-dimensional U-Net. Modern training recipes that combine the decoupled-weight-decay AdamW optimiser with cosine annealing and warm restarts[19] further stabilise training of large three-dimensional models. Ensembling complementary backbones is a well-known route to reduce variance and improve calibration, but to our knowledge no published study has explicitly tested an ensemble of complementary three-dimensional backbones in a controlled cross-geometry setting for dose prediction.

Building on these elements, in this study we asked a specific and previously unexplored question: can an ensemble of three complementary three-dimensional networks trained on a single beam configuration predict photon dose distributions for the same anatomy under several unseen beam configurations? We trained a Voltron ensemble, composed of SOTA-UNet3D, Res-UNet3D and Trans-UNet3D, on a single five-beam co-planar configuration (C0) and evaluated it on five unseen configurations (C1–C5) covering different beam counts and angular distributions. Reference doses were generated with matRad[16] and predictions were compared with the three-dimensional gamma index (3% / 3 mm).[20,21] To our knowledge, this is the first dosimetric study to evaluate a single three-dimensional ensemble across multiple unseen co-planar beam configurations under matched conditions. Our aim is not to deliver a clinical-grade dose calculator but to provide a transparent baseline that future cross-geometry robustness studies can compare against.

MATERIALS AND METHODS

Ethical approval

This study involved only in-silico simulations on publicly available phantoms and demonstration datasets distributed with the matRad toolkit.[16] No patient-level data, identifiers or human subjects were involved at any stage, and no ethics-committee approval or informed consent were therefore required.

Dataset and Reference Dose Generation

The overall workflow of this study is illustrated in Figure 1. All reference dose distributions were generated using the open-source treatment-planning toolkit matRad (matRad Research Group, German Cancer Research Center [DKFZ], Heidelberg, Germany), which implements analytical pencil-beam algorithms for photon plans within a single MATLAB framework

Turk J Oncol 2026;41(3):198–206
doi: 10.5505/tjo.2026.4983

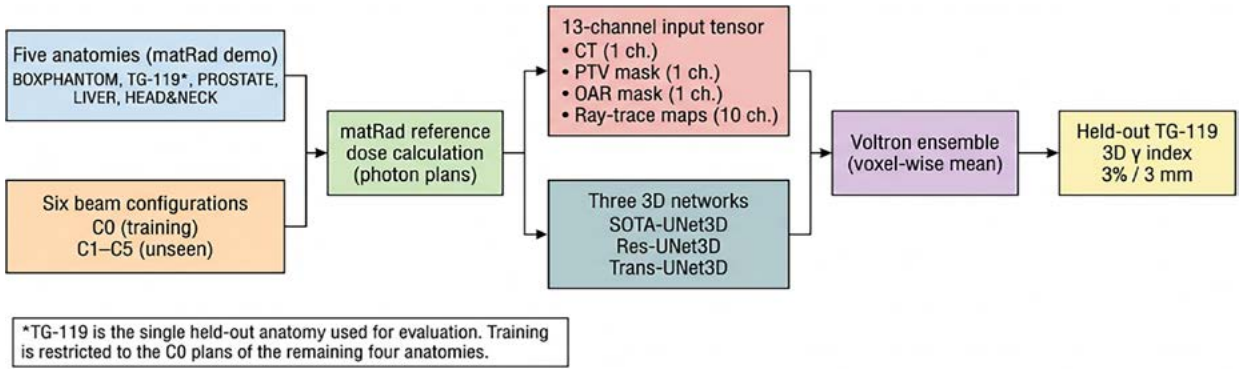


Fig. 1. Schematic overview of the study workflow. The matRad demonstration dataset (five anatomies) is combined with six beam configurations to generate reference photon dose distributions. A 13-channel input tensor (CT, PTV mask, OAR mask and per-beam ray-tracing maps) is provided to three complementary three-dimensional networks, whose voxel-wise mean defines the Voltron ensemble. The held-out anatomy (TG-119) is evaluated under the training configuration (C0) and the five unseen configurations (C1–C5) using the three-dimensional gamma index (3% / 3 mm).

CT: Computed tomography; PTV: Planning target volume; OAR: Organ at risk; MAE: Mean absolute error.

(MathWorks, Natick, MA, USA).[16] Five anatomies were used: A homogeneous water-equivalent box phantom (BOXPHANTOM), the TG-119 phantom (a synthetic concave-target benchmark), and three anthropomorphic cases (PROSTATE, LIVER and HEAD_AND_NECK) distributed with matRad as part of its demonstration dataset. The computed-tomography cubes, structure masks, voxel resolutions and isocentre coordinates supplied with each case were used without further processing.

For each anatomy, photon dose distributions were generated under six beam configurations indexed C0–C5. Configuration C0 used five equispaced co-planar beams (0°, 72°, 144°, 216°, 288°) and served as the single training configuration. The remaining configurations were: six equispaced co-planar beams (C1; 0°, 60°, 120°, 180°, 240°, 300°), seven equispaced co-planar beams (C2), nine equispaced co-planar beams (C3), six co-planar beams offset by 30° (C4; 30°, 90°, 150°, 210°, 270°, 330°) and four oblique co-planar beams (C5; 45°, 135°, 225°, 315°). All six geometries are illustrated in Figure 2. matRad inverse optimisation was run with default settings. The resulting dose cubes and structure information were exported as compressed NumPy archives (NumPy 1.26, NumPy Developers, Austin, TX, USA).

Input Tensor Construction

The input to each network was a 13-channel three-dimensional tensor (Supplementary Fig. S1). Channel 1 contained the computed-tomography image normalised to [0, 1] using the linear transformation (HU+1000)/4000, clipped to that interval. Channel 2 was the binary planning target volume (PTV) mask, defined as the union of all target structures available in matRad. Channel 3 was the binary organ-at-risk (OAR) mask, defined as the union of non-target structures excluding non-clinical labels such as BODY, SKIN, ENTRANCE, DOSE-FALLOFF, ISOCENTER and clip markers; the full exclusion list is given in Supplementary Table S1. Channels 4–13 encoded the beam geometry and physics: channel 4 held the voxel-wise sum of all per-beam ray-tracing maps, and channels 5–13 held up to nine individual per-beam maps.

For each gantry angle, a three-dimensional ray-tracing map was produced by combining a lateral Gaussian attenuation around the beam central axis (lateral standard deviation σ-lat=8.0 voxels) with a one-dimensional percent-depth-dose profile along the beam direction. For photons, the percent-depth-dose used an exponential attenuation with a soft build-up region; the functional form and parameter values are reported in Supplementary Table S2. Unused beam slots in configurations with fewer than nine beams were zero-padded. The computed-tomography image and structure masks are identical across all configurations of the same anatomy, so the only inputs that differ between configurations are channels 4–13.

Network Architectures and The Voltron Ensemble

Three complementary three-dimensional networks were used (Supplementary Fig. S2). All three were implemented in PyTorch 2.x (PyTorch Foundation, San Francisco, CA, USA) and accepted the same

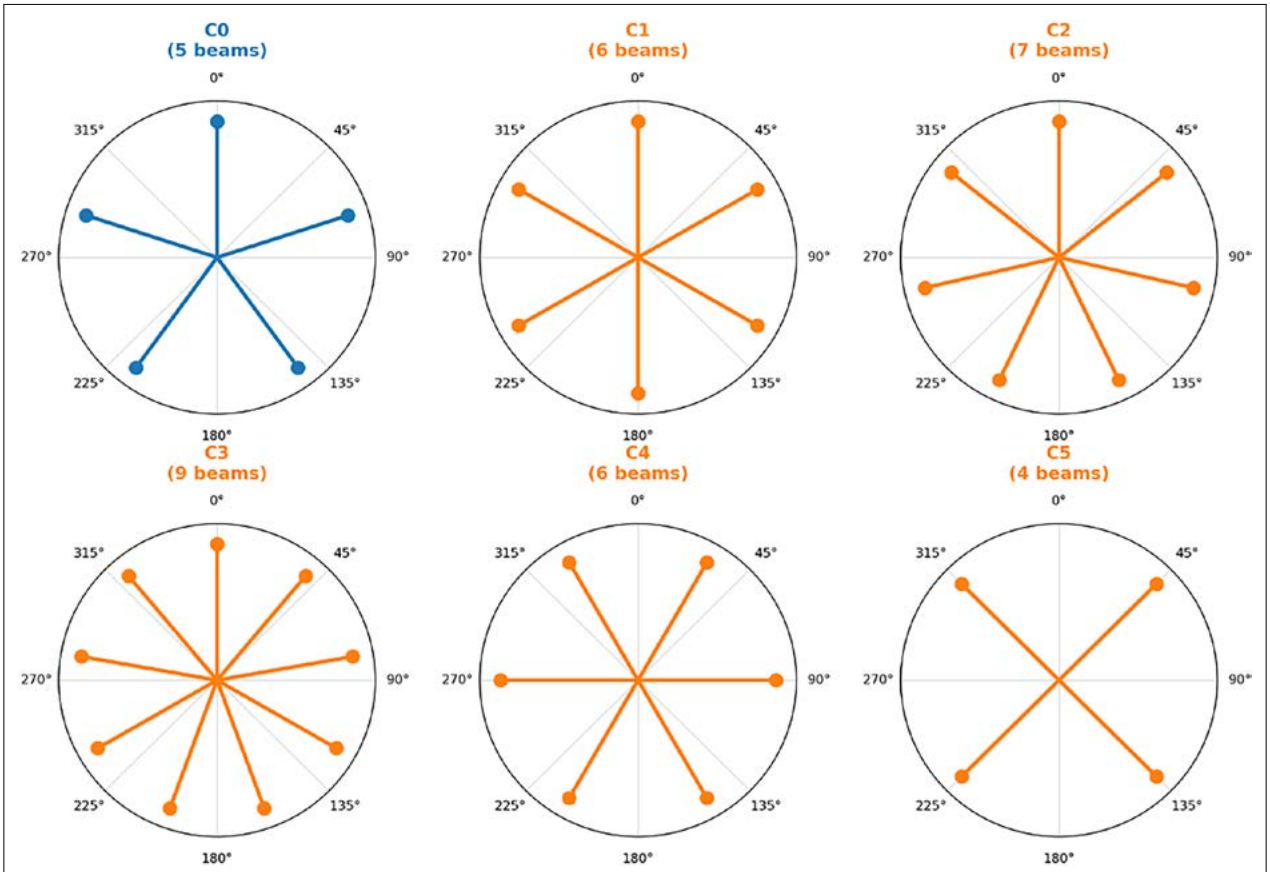


Fig. 2. Polar plots of the six beam configurations used for reference dose generation in matRad. C0 (blue) is the single training configuration; C1–C5 (orange) are used only for evaluation on the held-out anatomy. Each radial line indicates the gantry angle of a single beam. All six configurations are co-planar.

13-channel input tensor of size 128×128×128 voxels. SOTA-UNet3D is a four-level encoder–decoder with a base width of 32 channels in which each stage uses a multi-scale convolutional block (combining dilated 3×3×3 convolutions with dilation rates 1 and 2) and local–global attention-gated skip connections[22] inspired by the attention U-Net family; upsampling is performed by efficient up-convolutional blocks (trilinear upsampling followed by depthwise and pointwise convolutions).Res-UNet3D shares the same encoder–decoder topology but replaces the multi-scale block with a residual convolutional block:[17] two 3×3×3 convolutions with group normalisation and GELU activation, followed by an identity or projection shortcut. Trans-UNet3D inserts a single transformer encoder layer[18] at the bottleneck (16×16×16 spatial extent, 256 channels, four heads, feed-forward dimension 512). All three networks end in a 1×1×1 convolution followed by a sigmoid activation to anchor the output to the [0, 1] dose-fraction interval for numerical stability under mixed-precision training. The Voltron ensemble averages the three voxel-wise predictions at inference time and requires no additional training.

Training Protocol and Data Partitioning

The five anatomies were sorted alphabetically and the second anatomy (TG-119)[23] was held out as the test set; the remaining four (BOXPHANTOM, PROSTATE, LIVER and HEAD_AND_NECK) provided the training set. Crucially, training was restricted to the C0 photon plans of the four training anatomies, so the networks never saw any plan from the C1–C5 geometries at any stage of training. Training used random three-dimensional patches of size 128×128×128 voxels, with a 70% probability of being centred on a voxel inside the PTV (plus a small random jitter) and a 30% probability of being centred on a random foreground voxel; sixteen augmented patches were drawn per plan per epoch. On-the-fly augmentation was

Turk J Oncol 2026;41(3):198–206
doi: 10.5505/tjo.2026.4983

performed with MONAI 1.4 (Project MONAI consortium)[24] and comprised random spatial flips, random affine transformations, random Gaussian noise, random Gaussian smoothing and random intensity shifts. Each network was trained from scratch for 40 epochs with the AdamW optimiser[19] using a cosine-annealing-with-warm-restarts schedule. The loss function combined a mean-squared-error term and a mean-absolute-error term with unit weights. Random seeds were fixed to 42 throughout. Training was performed on NVIDIA A100 GPUs (NVIDIA Corporation, Santa Clara, CA, USA). Complete hyperparameter values are reported in Supplementary Table S3.

Inference and Evaluation Metrics

At inference time, the held-out anatomy was processed for each of the six beam configurations using three-dimensional sliding-window inference (MONAI 1.4) with overlapping 128×128×128 patches at 50% overlap and Gaussian importance weighting (σ-scale=0.125). The Voltron output was rescaled to physical dose units by multiplication with the maximum reference dose of the corresponding matRad plan. Predicted dose distributions were compared with the matRad references using two metrics. The first was the three-dimensional gamma index with criteria 3% / 3 mm and a 10% low-dose cut-off, computed as introduced by Low et al.[20] and following the practical guidance of Hussein et al.;[21] the dose-difference acceptance was set to 3% of the maximum reference dose, and the distance-to-agreement of 3 mm was translated into voxel radii using the actual physical voxel spacing of the anatomy. The second metric was the mean absolute error (MAE) in Gy between predicted and reference dose, computed over all voxels.

Statistical Analysis

Numerical computations and descriptive statistics were performed in Python 3.10 (Python Software Foundation, Wilmington, DE, USA) using NumPy 1.26, SciPy 1.13 and Pandas 2.2 (NumFOCUS, Austin, TX, USA). Figures were generated with Matplotlib 3.9 (Matplotlib Development Team). Because each test condition yielded a single dose volume per beam configuration on the held-out anatomy, summary measures are reported as scalar values per configuration together with the across-configuration mean and range. No formal inferential statistical testing was performed because the limited per-condition sample size (n=6 configurations on a single held-out anatomy) precludes meaningful hypothesis testing; the comparison of greatest interest in this in-silico study is the relative change of the per-configuration gamma pass rate with respect to the C0 baseline.

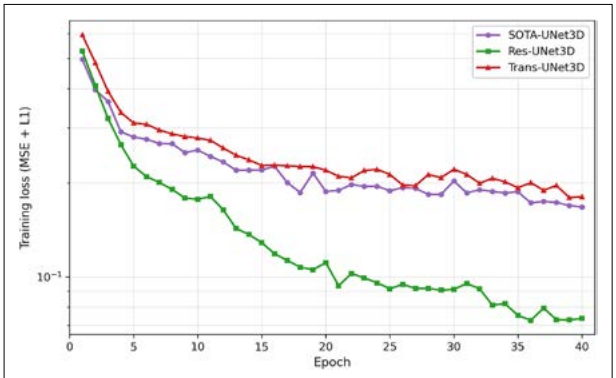


Fig. 3. Training-loss trajectories of the three networks trained from scratch on the single C0 configuration, displayed on a logarithmic vertical axis. The combined mean-squared-error + mean-absolute-error loss is shown for each of the 40 epochs. Per-network curves on a linear scale are provided in Supplementary Figure S3.

RESULTS

All three networks converged within 40 epochs of training on the single C0 configuration (Fig. 3). Res-UNet3D reached the lowest training loss (composite MSE+L1 loss=0.074 at epoch 40), while SOTA-UNet3D (0.167) and Trans-UNet3D (0.181) converged to similar values. Because the same training set, augmentation pipeline and hyperparameters were used for all three backbones, this difference is most plausibly attributed to the lower regularisation pressure of the plain residual block compared with the multi-scale and transformer-equipped blocks. Per-network training curves are shown in Supplementary Figure S3, and per-epoch loss values for all three networks are tabulated in Supplementary Table S4.

On the held-out TG-119 anatomy, the Voltron ensemble achieved a three-dimensional gamma pass rate (3% / 3 mm) of 51.42% on the training configuration C0, with a mean absolute error of 8.91 Gy (Table 1). This value is the in-distribution baseline against which the unseen-geometry predictions are compared. The full set of per-configuration values is visualised in Figure 4. A complementary scatter view of gamma versus MAE, and a distribution comparison of training versus unseen geometries, are provided in Supplementary Figure S4.

Table 1 Per-configuration three-dimensional gamma pass rate (3% / 3 mm) and mean absolute error of the Voltron ensemble on the held-out TG-119 anatomy

Configuration	Beams (gantry angles)	γ 3%/3 mm (%)	MAE (Gy)
C0 (training)	5 (0°, 72°, 144°, 216°, 288°)	51.42	8.91
C1	6 (0°, 60°, 120°, 180°, 240°, 300°)	47.38	9.17
C2	7 (equispaced)	51.34	9.08
C3	9 (equispaced)	53.17	9.11
C4	6 (offset by 30°)	53.59	8.94
C5	4 oblique (45°, 135°, 225°, 315°)	56.94	10.44
Unseen mean (C1–C5)	—	52.49	9.35

Configuration C0 was used to train the networks; C1–C5 were unseen during training. The row labelled 'Unseen mean (C1–C5)' reports the across-configuration mean for the unseen geometries. γ: Three-dimensional gamma index pass rate with 3% dose-difference criterion (relative to maximum reference dose) and 3 mm distance-to-agreement, computed under a 10% low-dose cut-off; MAE: Mean absolute error. All values were computed on the held-out anatomy (TG-119) against matRad reference plans.

Across the five unseen beam configurations, the Voltron ensemble achieved a mean gamma pass rate of 52.49% (range 47.38%–56.94%, Table 1). This value is slightly higher than the in-distribution C0 baseline (51.42%), with three of the five unseen configurations (C3, C4 and C5) producing higher gamma values than C0, while C1 and C2 were lower. The largest single deviation from baseline was a 4.0 percentage-point decrease on C1 (Fig. 4c). Configuration C5, which uses only four oblique beams and is geometrically the most different from the equispaced five-beam training geometry, in fact produced the highest gamma pass rate across all six configurations (56.94%). The heatmap in Figure 5a summarises gamma and MAE side by side,

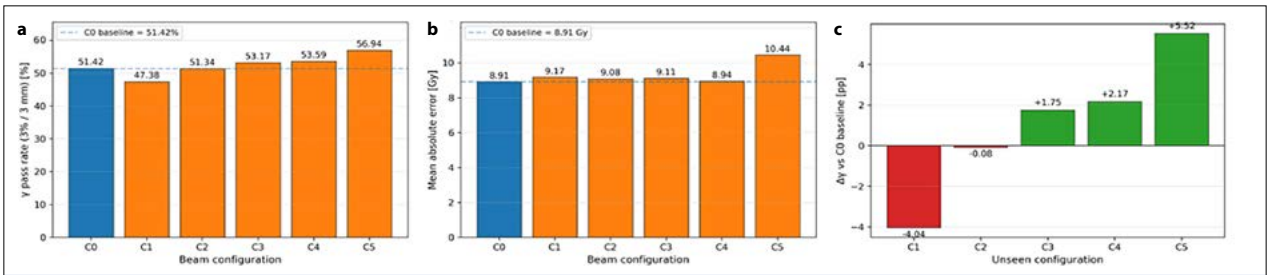


Fig. 4. Per-configuration evaluation of the Voltron ensemble on the held-out TG-119 anatomy. (a) Three-dimensional gamma pass rate (3% / 3 mm) for the six configurations; C0 (training) is shown in blue and the unseen geometries C1–C5 in orange. (b) Corresponding mean absolute error (MAE) in Gy. (c) Change in gamma pass rate of each unseen configuration relative to the C0 baseline (in percentage points; pp); positive values indicate higher gamma on the unseen geometry than on the training geometry.

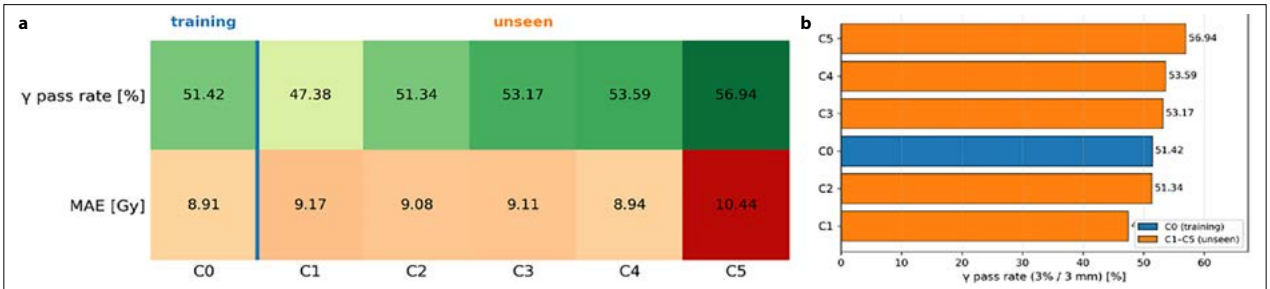


Fig. 5. Robustness summary across the six beam configurations. (a) Heatmap of the Voltron gamma pass rate (top row, green colour scale) and mean absolute error (bottom row, orange–red scale) for each configuration; the vertical blue line separates the training geometry (C0) from the five unseen geometries (C1–C5). (b) Configurations ranked by gamma pass rate, with C0 (blue) interleaved with the unseen geometries (orange).

Turk J Oncol 2026;41(3):198–206
doi: 10.5505/tjo.2026.4983

and the ranked view in Figure 5b orders the six configurations by gamma. Overall, the in-distribution and out-of-distribution geometries are interleaved in the ranking rather than separated.

Taken together, the per-configuration analysis indicates that the Voltron ensemble retained its dosimetric quality across the five unseen co-planar geometries: the mean gamma pass rate across C1–C5 exceeded the C0 baseline, and the largest individual decrease relative to baseline was on the order of four percentage points. MAE values stayed within approximately 8.9 to 10.4 Gy for all configurations, with no configuration producing a markedly larger error than the training geometry.

DISCUSSION

We trained a single ensemble of three complementary three-dimensional networks on one beam configuration (C0) and showed that, on a held-out anatomy, the ensemble produced dose distributions for five unseen co-planar geometries whose gamma pass rate at 3% / 3 mm was comparable to, and on average slightly higher than, the in-distribution baseline. To our knowledge, this is the first explicit dosimetric demonstration of single-configuration to multi-configuration transfer for physics-informed three-dimensional dose-prediction ensembles. The result supports the view that, when beam geometry is encoded through a small set of physically meaningful input channels rather than baked into the training distribution, much of the information a network needs to predict the dose can be shared across configurations.

Our design deliberately disentangles anatomy, geometry and the underlying photon physics. The computed-tomography image and structure masks are identical between configurations of any given plan; the only configuration-dependent inputs are the per-beam ray-tracing maps in channels 4–13. This is a much weaker form of geometric encoding than would be required if the network had to memorise each beam set from data, and is conceptually related to the beam-wise decomposition approaches proposed by other groups.[11,25] Our findings suggest that, once the geometry is supplied as an explicit input, the anatomy-conditional dose-shaping function the network has to learn is, to first approximation, invariant with respect to the number and angular distribution of co-planar beams, at least in the regimes implemented by matRad.[16]

It is interesting that several unseen configurations produced gamma pass rates above the in-distribution baseline. One plausible explanation is that, for a given anatomy, configurations with more beams or with beams that better cover the planning target volume produce smoother dose distributions that are simpler to reproduce voxel-wise, even though they were never seen during training. The lower gamma value on C1 (six equispaced beams with a different angular sampling from C0) and the higher gamma values on C3 (nine beams) and C5 (four oblique beams) hint that the local difficulty of dose reconstruction may depend more on the geometric coverage of the target than on similarity to the training geometry. Similar observations have been reported in other deep-learning dose-prediction studies that varied the input configuration.[10]

The Voltron ensemble combines a residual U-Net that converges to a very low training loss with two higher-bias estimators (a multi-scale plus attention-gate U-Net and a transformer-bottleneck U-Net). Ensembling complementary backbones is a well-established route to reduce variance and improve calibration of dose predictions.[1,7] We did not pursue weighted voting or stacking, in order to keep the pipeline reproducible and inexpensive. That this simple unweighted average already retains its dosimetric quality across multiple unseen geometries suggests that further gains from more sophisticated ensemble combinations are likely possible.

Beyond the controlled in-silico setting reported here, the broader idea that distributional differences across centres or scanners can be encapsulated in a physics-aware input layer rather than in centre-specific weights connects naturally to ongoing work on domain generalisation in medical imaging[14,15] and to the recent line of work on physics-informed dose-prediction models.[12] A particularly informative experiment would be to add a small number of plans from an additional clinical centre, with different beam-angle conventions, and to measure how much centre-specific data are needed to recover full performance; this would quantify the practical value of the cross-geometry starting point for data-scarce centres.

Limitations of the Study

Several limitations should be acknowledged. First, the reference dose distributions come from an analytical pencil-beam toolkit rather than from full Monte Carlo simulation. matRad has been validated against commercial treatment-planning systems[16], but analytical algorithms are known to be less accurate than Monte Carlo in highly heterogeneous geometries.[26,27] Our findings should therefore be read

as a controlled algorithm-to-algorithm comparison rather than an algorithm-to-clinical-truth comparison. Second, the held-out test set consists of a single anatomy, which is sufficient to demonstrate feasibility of cross-geometry transfer but does not characterise inter-patient variability rigorously. Third, the absolute gamma pass rates reported here are well below the threshold typically required for patient-specific quality assurance (commonly above 90%); the aim of this study, however, was not clinical-grade dose prediction but a transparent test of whether a single ensemble could remain robust across unseen beam configurations, and the relative change with respect to the C0 baseline is the primary outcome rather than the absolute pass rate. Fourth, our evaluation was restricted to photon plans and co-planar geometries; performance on non-co-planar geometries or on particle modalities was not assessed here. Fifth, because of the limited per-condition sample size (n=6 configurations on a single anatomy), only descriptive statistics could be reported; multi-centre, multi-patient validation against Monte Carlo references is the necessary next step before clinical translation can be considered.

CONCLUSION

An ensemble of three complementary three-dimensional networks trained on a single co-planar beam configuration produced photon dose distributions for five unseen co-planar configurations on a held-out anatomy with a mean three-dimensional gamma pass rate (3% / 3 mm) of 52.49%, compared with 51.42% for the in-distribution training configuration, and with mean absolute errors between 8.91 and 10.44 Gy across all six configurations. The largest single decrease relative to the training geometry was approximately four percentage points, while three configurations exceeded the baseline by 1.75 to 5.52 percentage points. These findings are far from clinical-grade performance, but they suggest that, when the beam geometry is encoded through a physics-informed input layer, a substantial fraction of the dose-prediction task remains transferable across co-planar beam configurations. We recommend that future cross-geometry robustness studies use this design as a reproducible baseline and explore extensions to non-co-planar geometries, Monte Carlo references and multi-centre datasets.

Informed Consent: Not applicable. This study involved only in-silico simulation on publicly available phantoms and demonstration datasets; therefore, no informed consent was required.

Conflict of Interest Statement: The authors declare that they have no conflict of interest related to this work.

Funding: The authors declared that this study received no financial support.

Use of AI for Writing Assistance: While the scientific focus of this research involves the development and evaluation of artificial intelligence (specifically, deep-learning models for dose prediction), this declaration concerns the use of generative AI tools in the preparation of the manuscript text. During preparation, the authors used Gemini (Google LLC, Mountain View, CA, USA) to improve language readability and to ensure grammatical correctness. After using this tool, the authors carefully reviewed and edited the content as needed. AI-assisted technologies were strictly limited to language editing and were not used in the conceptualisation, execution of the research or data generation. The authors take full responsibility for the final content of the publication.

Author Contributions: Concept – M.E.B., F.T.; Design – M.E.B., F.T.; Supervision – M.E.B., F.T.; Materials – M.E.B.; Data collection and/or processing – M.E.B., F.T.; Data analysis and/or interpretation – M.E.B., F.T.; Literature search – M.E.B., F.T.; Writing – M.E.B.; Critical review – M.E.B., F.T.

Acknowledgments: The authors gratefully acknowledge the matRad development team at the German Cancer Research Center (DKFZ; Heidelberg, Germany) for making their open-source treatment-planning toolkit and example datasets publicly available, and the Project MONAI consortium for the open-source medical-imaging deep-learning library used in this study. No other individuals contributed to the preparation of this manuscript in a manner that did not fulfil the authorship criteria.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Babier A, Zhang B, Mahmood R, Moore KL, Purdie TG, McNiven AL, et al. OpenKBP: The open-access knowledge-based planning grand challenge and dataset. Med Phys 2021;48(9):5549–61.
2. Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys 2019;46(1):370–81.
3. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation.

Turk J Oncol 2026;41(3):198–206
doi: 10.5505/tjo.2026.4983

In Navab N, Hornegger J, Wells WM, Frangi AF, eds. Medical image computing and computer-assisted intervention - MICCAI 2015. Cham: Springer; 2015. p. 234–41.

4. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. Medical image computing and computer-assisted intervention - MICCAI 2016. Cham: Springer; 2016. p. 424–32.
5. Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: A volumetric dose prediction algorithm using 3D fully-convolutional neural networks. Phys Med Biol 2018;63(23):235022.
6. Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-Net deep learning architecture. Phys Med Biol 2019;64(6):065020.
7. Liu S, Zhang J, Li T, Yan H, Liu J. Technical note: A cascade 3D U-Net for dose prediction in radiotherapy. Med Phys 2021;48(9):5574–82.
8. Koike Y, Takegawa H, Anetai Y, Ohira S, Nakamura S, Tanigawa N. Patient-specific three-dimensional dose distribution prediction via deep learning for prostate cancer therapy: Improvement with the structure loss. Phys Med 2023;107:102544.
9. Bao P, Xia W, Yang K, Cai L, Cui Y. Beam-angle optimization for intensity-modulated radiation therapy using deep reinforcement learning. Med Phys 2024;51(5):3441–55.
10. Barragán-Montero AM, Nguyen D, Lu W, Lin MH, Norouzi-Kandalan R, Geets X, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: Robust learning from heterogeneous beam configurations. Med Phys 2019;46(8):3679–91.
11. Teng L, Wang B, Xu X, Zhang J, Mei L, Feng Q, et al. Beam-wise dose composition learning for head and neck cancer dose prediction in radiotherapy. Med Image Anal 2024;92:103045.
12. Wang Y, Piao Z, Gu H, Chen M, Zhang D, Zhu J. Deep learning-based prediction of radiation therapy dose distributions in nasopharyngeal carcinomas: A preliminary study incorporating multiple features including images, structures, and dosimetry. Technol Cancer Res Treat 2024;23:15330338241256594.
13. Pastor-Serrano O, Perkó Z. Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy. Phys Med Biol 2022;67(10):105006.
14. Sanner A, González C, Mukhopadhyay A. How reliable are out-of-distribution generalization methods for medical image segmentation? In Bauer S, Drechsler K, ed. Pattern recognition - DAGM GCPR 2021. Cham: Springer; 2021. p. 604–17.
15. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med Image Anal 2020;63:101693.
16. Wieser HP, Cisternas E, Wahl N, Ulrich S, Stadler A, Mescher H, et al. Development of the open-source dose calculation and optimization toolkit matRad. Med Phys 2017;44(6):2556–68.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit 2016:770–8.
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30:5998–6008.
19. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv 2019;1711:05101
20. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25(5):656–61.
21. Hussein M, Clark CH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy - towards good practice. Phys Med 2017;36:1–11.
22. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal. 2019;53:19–207.
23. Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys 2009;36(11):5359–73.
24. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. MONAI: An open-source framework for deep learning in healthcare. arXiv 2022;2211:02701.
25. Wang B, Teng L, Mei L, Cui Z, Xu X, Feng Q, et al. Beam-wise dose composition for head-and-neck dose prediction. In Wang L, Dou Q, Fletcher PT, Speidel S, Li S, eds. Medical image computing and computer assisted intervention - MICCAI 2022. Cham: Springer; 2022. p. 575–84.
26. Newhauser WD, Zhang R. The physics of proton therapy. Phys Med Biol 2015;60(8):R155–R209.
27. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol 2012;57(11):R99–R117.

Supplementary Materials

Cross-Geometry Generalisation of a Physics-Informed 3D Dose-Prediction Ensemble

This supplementary file accompanies the main manuscript submitted to the Turkish Journal of Oncology. It contains the supplementary figures and tables referred to in the main text, presented in the order in which they are first cited.

Contents

• Supplementary Table S1 — Structure labels excluded from the OAR mask

• Supplementary Table S2 — Functional form and parameter values of the photon depth-dose model

• Supplementary Table S3 — Complete training hyperparameters

• Supplementary Table S4 — Per-epoch training-loss values for the three networks

• Supplementary Figure S1 — Visualisation of the 13-channel input tensor

• Supplementary Figure S2 — Schematic of the three networks in the Voltron ensemble

• Supplementary Figure S3 — Per-network training-loss trajectories (linear scale)

• Supplementary Figure S4 — Complementary views of per-configuration performance

Supplementary Table S1 Structure labels excluded when building the organ-at-risk (OAR) mask (channel 3 of the input tensor). All non-target labels listed below are excluded; the remaining non-target labels are unioned into a single binary OAR mask. Matching is case-insensitive and ignores leading and trailing whitespace.

Label (case-insensitive)	**Description**
BODY	Whole patient body / external contour
SKIN	Patient skin contour
ENTRANCE	Beam entrance contour
DOSE_FALLOFF	Dose fall-off helper contour
ISOCENTER	Isocentre marker
CLIP	Generic clip marker

Couch	Treatment-couch contour
EXTERNAL	Synonym for body external
AIR	Air region inside the body contour
Lungs (combined)	Composite lungs mask when separate L/R lungs are present

OAR: organ at risk. The exclusion list is applied during matRad structure parsing and intentionally removes contours that are not clinically meaningful organs at risk (helper or marker structures).

Supplementary Table S2 Functional form and parameter values of the one-dimensional percent-depth-dose (PDD) model used to construct the per-beam ray-tracing maps in channels 4–13 of the input tensor.

Functional form (photons): $PDD(d) = (1 - exp(-\kappa \cdot d)) \cdot exp(-\mu \cdot d) \cdot (1 + \beta \cdot exp(-\alpha \cdot d))$, where d is the depth along the beam direction in voxels.

Parameter	**Value**	**Meaning**
μ (linear attenuation)	0.025 vox^{-1}	Bulk attenuation along the beam
κ (build-up rate)	0.05 vox^{-1}	Soft build-up region near beam entrance
β (build-up amplitude)	0.30	Strength of the build-up correction
α (build-up decay)	0.10 vox^{-1}	Decay of the build-up correction
σ-lateral	8.0 voxels	Lateral Gaussian width about the central axis
Max beam depth	160 voxels	Truncation depth along the beam

vox: Voxel. Parameter values were selected to give a smooth monotone PDD with a soft build-up region and were held fixed across all anatomies and beam configurations.

Supplementary Table S3 Complete training hyperparameters used for all three networks of the Voltron ensemble. The same hyperparameters were applied to SOTA-UNet3D, Res-UNet3D and Trans-UNet3D.

Hyperparameter	**Value**

Patch size	128 × 128 × 128 voxels
Patches per plan per epoch	16
Centre-on-PTV probability	0.70
Centre-on-foreground probability	0.30
Batch size (per step)	6
Gradient accumulation steps	8
Effective batch size	48
Number of epochs	40
Optimiser	AdamW
Initial learning rate	1×10^{-4}
Weight decay	5×10^{-4}
Learning-rate schedule	Cosine annealing with warm restarts
Warm-restart period (T_o)	10 epochs
Warm-restart multiplier (T_mult)	2
Loss function	MSE + L1 (unit weights)
Mixed-precision	Enabled (FP16 + dynamic loss scaling)
Gradient clipping	L2 norm 1.0
Random seed	42 (NumPy + PyTorch + Python random)
Augmentation library	MONAI 1.4 (Project MONAI consortium)
Augmentations applied	random flips, random affine, random Gaussian noise, random Gaussian smoothing, random intensity shift
Hardware	NVIDIA A100 GPU (NVIDIA Corp., Santa Clara, CA, USA)
Framework	PyTorch 2.x (PyTorch Foundation, San Francisco, CA, USA)

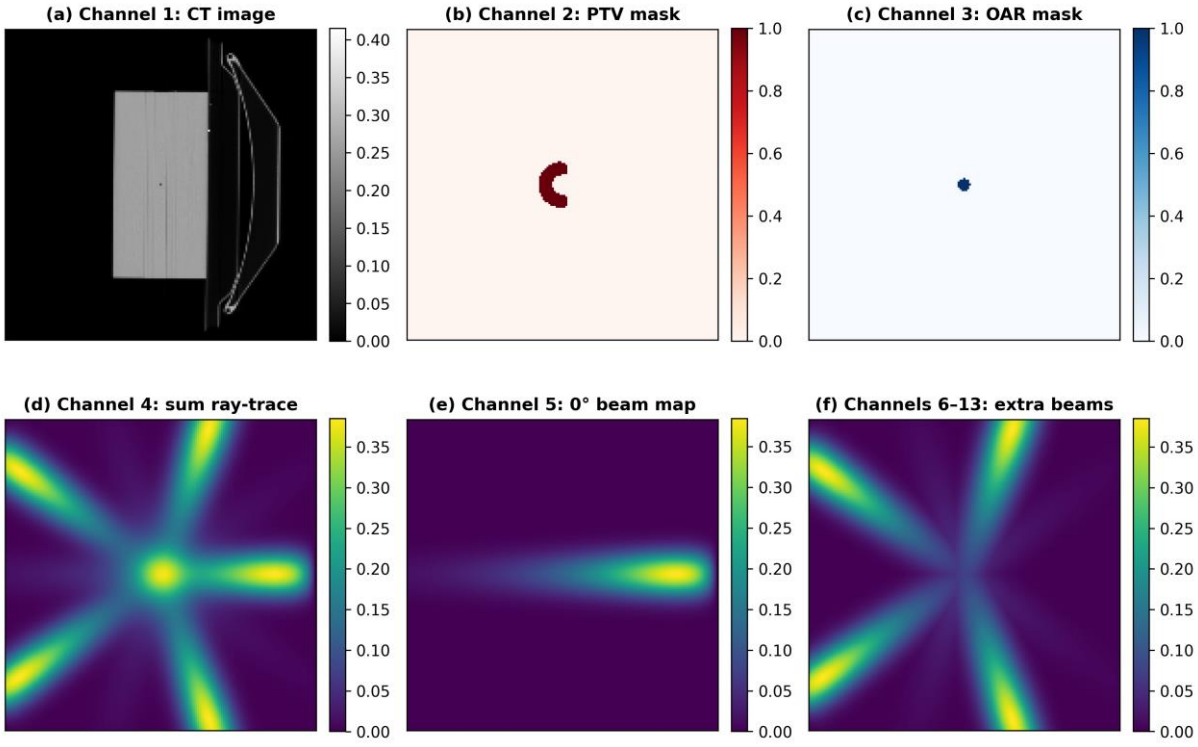
MSE: Mean squared error; L1: Mean absolute error; FP16: Half precision; T_o and T_mult are the standard parameters of the cosine-annealing-with-warm-restarts schedule.

Supplementary Table S4 Per-epoch training loss values (combined MSE + L1, unit weights) for the three networks of the Voltron ensemble. All three networks were trained for 40 epochs from scratch on the single C0 beam configuration.

Epoch	SOTA-UNet3D	Res-UNet3D	Trans-UNet3D
1	0.4967	0.5282	0.5950
2	0.3972	0.4095	0.4849
3	0.3642	0.3216	0.3927
4	0.2912	0.2651	0.3361
5	0.2804	0.2263	0.3108
6	0.2758	0.2095	0.3079
7	0.2673	0.2007	0.2953
8	0.2668	0.1905	0.2868
9	0.2499	0.1789	0.2813
10	0.2545	0.1773	0.2782
11	0.2429	0.1809	0.2735
12	0.2332	0.1639	0.2584
13	0.2195	0.1430	0.2451
14	0.2197	0.1367	0.2368
15	0.2198	0.1287	0.2274
16	0.2260	0.1187	0.2278
17	0.2003	0.1129	0.2269
18	0.1863	0.1075	0.2255
19	0.2148	0.1055	0.2254
20	0.1877	0.1111	0.2195
21	0.1892	0.0937	0.2098
22	0.1975	0.1026	0.2075
23	0.1949	0.0993	0.2190
24	0.1950	0.0958	0.2209
25	0.1887	0.0917	0.2128
26	0.1930	0.0947	0.1967
27	0.1921	0.0919	0.1958

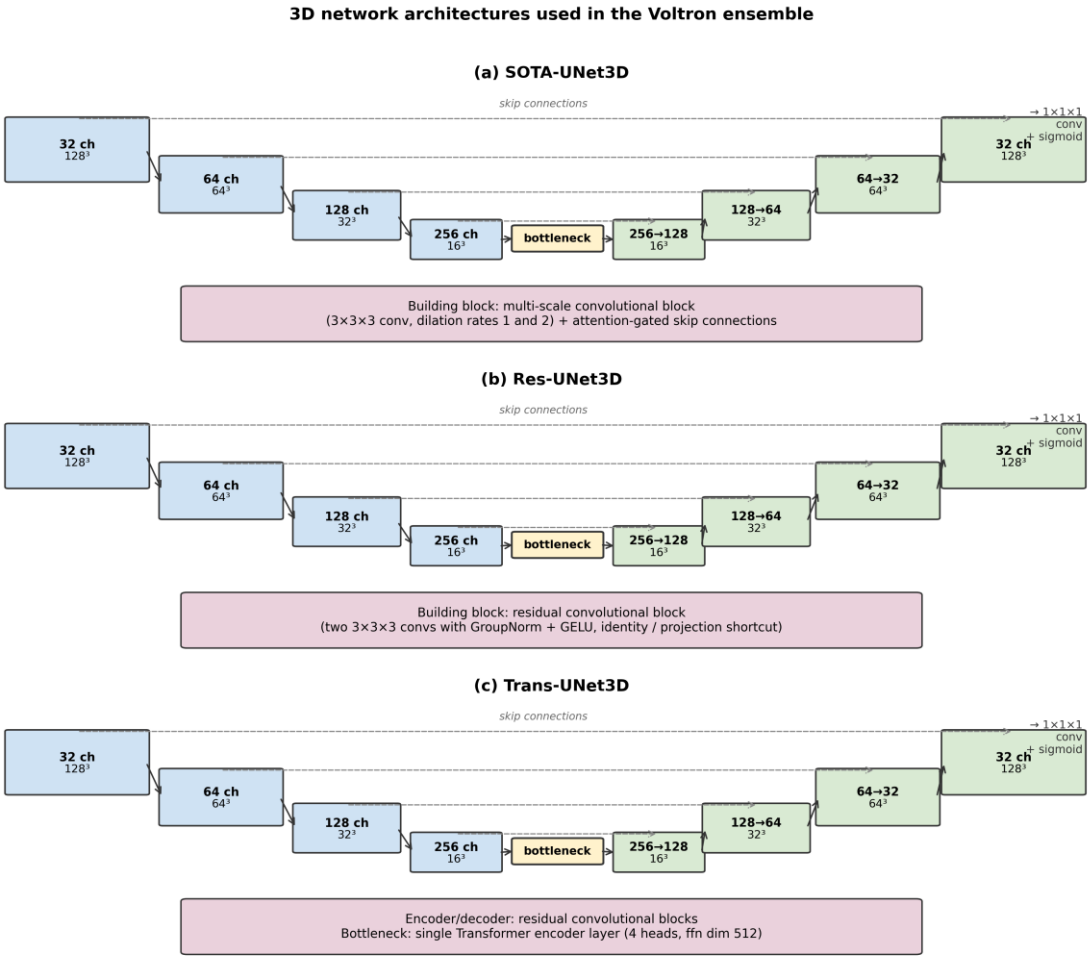
28	0.1837	0.0919	0.2127
29	0.1838	0.0907	0.2078
30	0.2027	0.0913	0.2210
31	0.1858	0.0953	0.2127
32	0.1900	0.0917	0.1994
33	0.1877	0.0813	0.2068
34	0.1856	0.0821	0.2015
35	0.1873	0.0753	0.1929
36	0.1727	0.0727	0.2003
37	0.1745	0.0795	0.1893
38	0.1732	0.0730	0.1962
39	0.1693	0.0729	0.1796
40	0.1674	0.0736	0.1806

MSE: Mean squared error; L1: Mean absolute error. Values rounded to four decimal places.



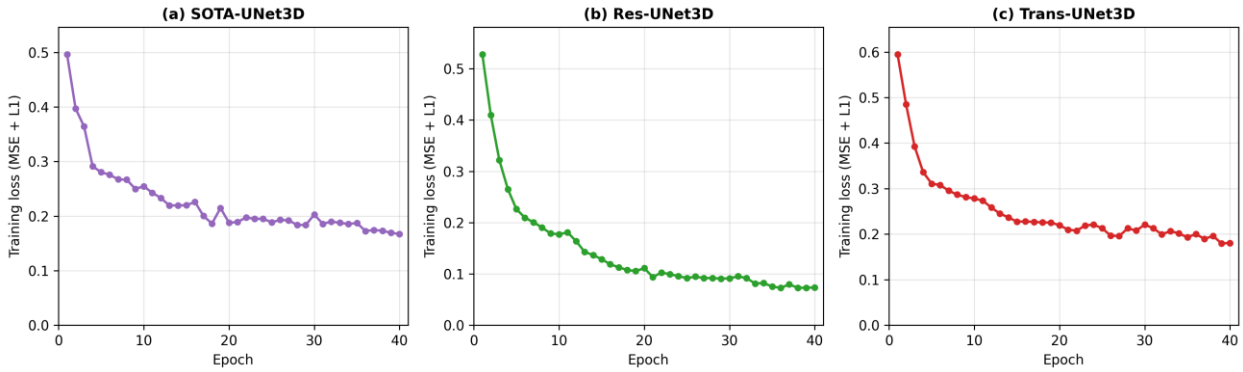
Supplementary Fig. S1. Visualisation of the 13-channel input tensor used by all three networks on a representative axial slice of the held-out TG-119 anatomy. (a) Channel 1: CT

image (Hounsfield-unit values normalised to [0, 1]). (b) Channel 2: Binary planning target volume (PTV) mask. (c) Channel 3: Binary organ-at-risk (OAR) mask. (d) Channel 4: Voxel-wise sum of all per-beam ray-tracing maps shown for the C0 configuration. (e) Channel 5: An individual per-beam ray-tracing map for the 0° beam of C0. (f) Channels 6–13: Maximum projection of the remaining per-beam maps; unused beam slots in configurations with fewer than nine beams are zero-padded. CT, computed tomography; PTV, planning target volume; OAR, organ at risk.

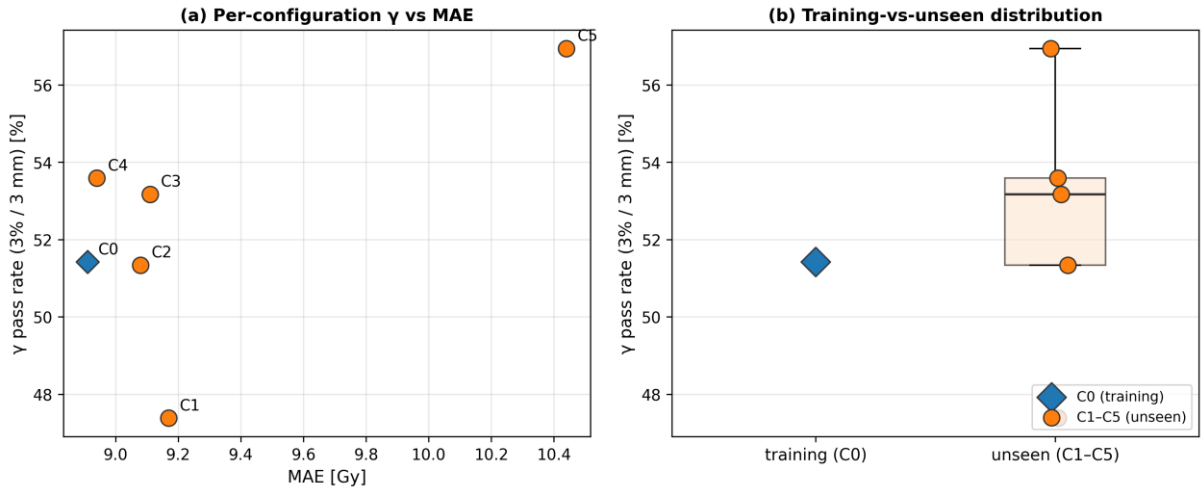


Supplementary Fig. S2. Schematic of the three three-dimensional networks that make up the Voltron ensemble. All three share a four-level encoder–decoder topology with skip connections; they differ in the type of computational block applied at each stage. (a) SOTA-UNet3D uses multi-scale convolutional blocks (dilated 3 × 3 × 3 convolutions with dilation rates 1 and 2) combined with attention-gated skip connections. (b) Res-UNet3D uses simple residual convolutional blocks (two 3 × 3 × 3 convolutions with group normalisation and

GELU, plus an identity or projection shortcut). (c) Trans-UNet3D inserts a single transformer encoder layer at the bottleneck (four heads, feed-forward dimension 512). All three networks end in a 1 × 1 × 1 convolution followed by a sigmoid activation to constrain the output to the [0, 1] dose-fraction interval. The Voltron ensemble averages the three voxel-wise predictions at inference time.



Supplementary Fig. S3. Per-network training-loss trajectories on a linear vertical axis. (a) SOTA-UNet3D. (b) Res-UNet3D. (c) Trans-UNet3D. The combined mean-squared-error + mean-absolute-error loss is shown for each of the 40 training epochs on the single C0 configuration. The combined log-scale view of all three networks is presented as Figure 3 in the main text.



Supplementary Fig. S4. Complementary views of the per-configuration Voltron ensemble performance on the held-out TG-119 anatomy. (a) Scatter plot of the three-dimensional

gamma pass rate (3% / 3 mm) against the mean absolute error (MAE) per beam configuration; C0 (training) is marked by a blue diamond and the unseen configurations C1–C5 are shown in orange. (b) Distribution comparison: the single gamma value of the training configuration (C0; blue diamond) versus the distribution of values across the five unseen configurations (orange dots; box plot drawn for reference).