Acıbadem Univ. Sağlık Bilim. Derg. 17 (October, November, December) 2026
https://doi.org/10.31067/acusaglik.1957061

ORIGINAL ARTICLE / ARAŞTIRMA YAZISI

Radiology & Organ Imaging / Radyoloji ve Organ Görüntüleme

Zero-Shot Cross-Modality Dose Prediction in Particle Therapy via Physics-Informed Ensembles

Muhammed Emin Bedir[1,2]

[1] Vocational School of Technical Sciences, Department of Electricity and Energy, Karamanoğlu Mehmetbey University, Karaman, Türkiye

[2] School of Medicine and Public Health, Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA

Muhammed Emin Bedir
0000-0002-4428-1241

Correspondence: Muhammed Emin Bedir
Vocational School of Technical Sciences, Department of Electricity and Energy, Karamanoğlu Mehmetbey University, Karaman, Türkiye
Phone: +90 542 614 37 44
E-mail: meminbedir@kmu.edu.tr

Received: 22.05.2026
Accepted: 25.08.2026

ABSTRACT

Purpose: Deep learning can accelerate radiotherapy dose prediction, but particle-therapy datasets remain scarce. We evaluated whether a physics-informed ensemble trained exclusively on photon plans could predict proton and carbon-ion dose distributions without particle-dose training.

Methods: Using matRad, 90 reference plans (photons, protons and carbon ions) were generated for five anatomies and six beam configurations. A 13-channel tensor combining CT, structure masks and modality-specific ray-tracing/depth-dose channels was input to SOTA-UNet3D, Res-UNet3D and Trans-UNet3D; their voxel-wise mean formed the Voltron ensemble. Leave-one-anatomy-out evaluation used four test anatomies with photon-only training. Performance was assessed with PyMedPhys gamma analysis (3%/3 mm and 2%/2 mm; 10% low-dose threshold) and mean absolute error (MAE).

Results: Across four held-out anatomies, mean 3%/3 mm gamma pass rates were 23.5% for photons, 23.3% for protons and 21.5% for carbon ions; 2%/2 mm rates were 15.8%, 15.5% and 14.5%. Mean MAE was approximately 5.2–5.3 Gy. Gamma values reported in an earlier version of this work (73.2/70.8/69.1%) were computed with a non-standard lenient criterion; all values reported here use the standard gamma index.

Conclusion: Photon-only training with physics-informed inputs provides a reproducible feasibility baseline for zero-shot particle-dose prediction. Performance is not clinical-grade, but the small cross-modality degradation suggests that anatomical and geometric features learned from photon data can transfer to particle therapy.

Keywords: Radiotherapy; Deep Learning; Proton Therapy; Heavy Ion Radiotherapy; Radiotherapy Planning, Computer-Assisted; Machine Learning

ÖZET

Amaç: Derin öğrenme radyoterapi doz tahminini hızlandırabilir; ancak parçacık terapisi veri kümeleri sınırlıdır. Yalnızca foton planlarıyla eğitilmiş fizik tabanlı bir topluluk modelinin, parçacık doz verisiyle eğitilmeden proton ve karbon iyon doz dağılımlarını tahmin edip edemeyeceği değerlendirildi.

Yöntemler: matRad kullanılarak beş anatomi ve altı ışın yapılandırması için 90 referans plan (foton, proton ve karbon iyon) üretildi. BT görüntüsü, yapı maskeleri ve modaliteye özgü ışın izleme/derinlik-doz kanallarını birleştiren 13 kanallı girdi SOTA-UNet3D, Res-UNet3D ve Trans-UNet3D ağlarına verildi; voksel düzeyindeki ortalamaları Voltron topluluğunu oluşturdu. Dört test anatomisinde leave-one-anatomy-out değerlendirme ve yalnızca foton eğitimi kullanıldı. Performans PyMedPhys gama analizi (%3/3 mm ve %2/2 mm; %10 düşük doz eşiği) ve ortalama mutlak hata (MAE) ile değerlendirildi.

Bulgular: Dört dışarıda bırakılan anatomi genelinde ortalama %3/3 mm gama geçiş oranları foton, proton ve karbon iyon için sırasıyla %23,5, %23,3 ve %21,5; %2/2 mm oranları ise %15,8, %15,5 ve %14,5 idi. Ortalama MAE yaklaşık 5,2-5,3 Gy idi. Çalışmanın önceki bir sürümünde bildirilen %73,2/%70,8/%69,1 değerleri standart dışı gevşek bir ölçütle hesaplanmıştı; burada sunulan tüm değerler standart gama indeksine dayanmaktadır.

Sonuç: Fizik tabanlı girdilerle yalnızca foton eğitimi, sıfır-atışlı parçacık doz tahmini için tekrarlanabilir bir fizibilite temeli sağlamaktadır. Performans klinik düzeyde değildir; ancak modaliteler arası küçük performans kaybı, foton verisinden öğrenilen anatomik ve geometrik özelliklerin parçacık terapisine aktarılabildiğini düşündürmektedir.

Anahtar Kelimeler: Radyoterapi; Derin öğrenme; Proton terapi; Ağır iyon radyoterapisi; Radyoterapi planlama, Bilgisayar destekli; Makine öğrenmesi

Copyright © 2026 the Author(s). Published by Acibadem University. This is an open access article licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND 4.0) International License, which is downloadable, re-usable and distributable in any medium or format in unadapted form and for noncommercial purposes only where credit is given to the creator and publishing journal is cited properly. The work cannot be used commercially without permission from the journal.

Radiotherapy remains a cornerstone of modern oncology and is delivered to a large fraction of patients with solid tumours (1). Conventional photon therapy is widely available, but particle therapies, proton beams and, in selected centres, carbon ions, offer more conformal physical dose distributions by virtue of the Bragg peak, in which energy deposition is concentrated at the end of the particle range with a steep distal fall-off (2,3). Carbon ions add a higher, depth-increasing relative biological effectiveness that has been exploited clinically for radioresistant histologies at the Heavy Ion Medical Accelerator in Chiba and at the Heidelberg Ion Therapy Centre (2–4). Despite this physical and biological promise, only a small number of dedicated particle therapy facilities operate globally, and the resulting paucity of openly shareable treatment-planning data has constrained data-driven research in this field (4,5).

In parallel, deep learning has been adopted for automated photon dose prediction. Following the U-Net architecture (6) and its three-dimensional extension (7), three-dimensional convolutional networks have become a de facto standard for voxel-wise dose prediction in radiotherapy (8–10). Nguyen and colleagues trained a hierarchically densely connected U-Net on head-and-neck plans (10); Barragán-Montero and colleagues showed that incorporating beam geometry as an additional input channel allowed a single network to generalise across heterogeneous beam configurations in lung intensity-modulated radiotherapy (9). Babier and colleagues released OpenKBP, standardising benchmarking of photon dose prediction (8). Subsequent work has explored residual-network-based three-dimensional dose prediction for helical tomotherapy (11) and the coupling of dose prediction with automatic plan generation (12). Most of these systems share two design choices: a three-dimensional U-Net-like encoder–decoder backbone, and an input tensor consisting of the computed-tomography image and binary contours of the planning target volume (PTV) and organs at risk (OARs).

Particle therapy has received less attention. Pastor-Serrano and Perkó introduced DoTA, a transformer-based network that approximates Monte Carlo proton pencil-beam dose calculation in milliseconds (13). Recent work has extended deep learning to heavy-ion dose prediction (14) and to direct cone-beam computed-tomography-based dose calculation for protons (15). To our knowledge, every published particle-therapy prediction model has been trained directly on particle dose data and therefore requires datasets that very few centres can produce. The clinical importance of these data is amplified by the well-documented sensitivity of proton dose calculations to range and stopping-power uncertainties (16).

Cross-modality learning has been demonstrated for medical image segmentation, where computed tomography and magnetic resonance imaging share anatomy but differ in appearance. Kieselmann and colleagues trained a model on annotated computed-tomography images and used it to contour magnetic-resonance images of head-and-neck cancer patients without target-modality annotations (17). Domain-adaptation frameworks (18) and challenges such as crossMoDA (19) have matured these ideas. Dose prediction is, however, a different task: unlike segmentation, where the target output is a largely modality-invariant categorical mask, dose distributions depend on the underlying transport physics, with photons depositing dose via exponential attenuation and charged particles via a Bragg peak whose depth depends on initial energy and tissue stopping power (16,20). It is therefore not obvious that a network trained only on photon plans should produce sensible particle dose maps, even with a modality-aware input.

The open-source toolkit matRad (matRad Research Group, German Cancer Research Center [DKFZ], Heidelberg, Germany) provides a reproducible framework for computing photon, proton and carbon-ion reference doses under consistent algorithms (21); its multi-modality character is almost unique for controlled studies of cross-modality questions. On the model side, residual learning (22) and attention mechanisms (23) have inspired numerous U-Net variants for three-dimensional medical imaging (24). To our knowledge, however, no published ensemble has been explicitly tested in a zero-shot cross-modality setting for radiotherapy dose prediction.

In this study we asked a specific and previously unexplored question: can an ensemble of complementary three-dimensional networks, trained exclusively on photon plans, predict proton and carbon-ion dose distributions in a zero-shot setting if the input representation encodes modality-specific ray tracing? Our contributions are threefold. First, we construct a 13-channel input tensor in which the only modality-dependent information is a small set of physics-informed depth-dose channels, keeping the anatomical and geometric inputs identical across modalities. Second, we train a Voltron ensemble of three complementary three-dimensional backbones (SOTA-UNet3D, Res-UNet3D and Trans-UNet3D) on photon plans only and use the same weights to predict proton and

carbon-ion doses. Third, we evaluate this strategy across three modalities and six beam configurations using the three-dimensional gamma index (25). To our knowledge, this is the first dosimetric study to explicitly report zero-shot photon-to-particle transfer with a single physics-informed ensemble.

Material and Methods

Dataset and Reference Dose Generation

All reference dose distributions were generated with the open-source treatment planning toolkit matRad (matRad Research Group, German Cancer Research Center [DKFZ], Heidelberg, Germany) (21), which implements analytical pencil-beam algorithms for photon, scanned-proton and scanned-carbon-ion plans within a single MATLAB framework (MathWorks, Natick, MA, USA). matRad and its example phantoms were downloaded from the official public repository (https://github.com/e0404/matRad). 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. Computed-tomography cubes, structure masks, voxel resolutions and isocentre coordinates were used without further processing of the underlying anatomy.

For each anatomy, dose distributions were computed for three modalities under six beam configurations (configurations 0–5), yielding $5 \times 3 \times 6 = 90$ reference dose volumes. The six configurations comprised five equispaced co-planar beams (configuration 0), six equispaced beams (configuration 1), seven equispaced beams (configuration 2), nine equispaced beams (configuration 3), six beams offset by 30° (configuration 4) and four oblique beams (configuration 5). Exact gantry angles per configuration are reported in Supplementary Figure S1 and Supplementary Table S1. matRad inverse optimisation was run with default settings for each modality. Dose cubes and structure information were exported as compressed NumPy archives (NumPy 1.26, NumPy Developers).

Input Tensor Construction

The input to the network was a 13-channel three-dimensional tensor (Figure 1). Channel 1 contained the computed-tomography image normalised to [0, 1] using (HU + 1000)/4000, clipped to the same interval. Channel 2 was the binary PTV mask. Channel 3 was the binary 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 (Supplementary Table S2). Channels 4–13 encoded beam geometry and modality-specific 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 (standard deviation $\sigma_{lat} = 8.0$ voxels) with a one-dimensional depth-dose profile along the beam direction. Depth along each beam was taken as the water-equivalent path length (WEPL), obtained by integrating CT-derived relative stopping power along the ray, so that the depth-dose profile is placed at the correct radiological depth rather than a geometric one. The depth-dose profile was modality-specific: for photons, a WEPL-based build-up plus exponential attenuation; for protons and carbon ions, a Bortfeld-type analytical Bragg model (power-law rise with exponent −0.435, $p = 1.77$, peaking at the WEPL range R_0, with a range-straggling distal fall-off, $\sigma = 0.012 \cdot R_0^{0.935}$), with an additional fragmentation tail beyond the Bragg peak for carbon ions. The full functional forms and per-modality parameters are summarised in Supplementary Table S3. Unused beam slots in configurations with fewer than nine beams were zero-padded. We emphasise that this remains an analytical surrogate rather than a Monte Carlo transport calculation; because it is supplied as an input channel, its role is to encode modality-specific geometry, while the network learns the anatomy-conditional mapping to the reference dose. Modality information enters the network only through this physics-informed depth-dose model in channels 4–13; the computed-tomography image and structure masks are identical between modalities for any given plan, which is the structural property that makes the proposed cross-modality transfer well defined.

Network Architectures and the Voltron Ensemble

Three complementary three-dimensional networks were used (Supplementary Figure S2), all implemented in PyTorch 2.x (PyTorch Foundation, San Francisco, CA, USA) and accepting the same 13-channel input of size $128 \times 128 \times 128$ voxels. SOTA-UNet3D is a four-level encoder–decoder with a base width of 32 channels. Each stage uses a multi-scale convolutional block combining dilated $3 \times 3 \times 3$ convolutions (dilation rates 1 and 2), GELU activation, group normalisation and three-dimensional dropout (probability 0.25). Skip connections are gated

by local–global attention gates inspired by the attention U-Net family. The prediction head is a $1 \times 1 \times 1$ convolution followed by a sigmoid activation.

Res-UNet3D shares the same skeleton but replaces the multi-scale block with a residual convolutional block (two $3 \times 3 \times 3$ convolutions with group normalisation, GELU and identity or projection shortcut). Trans-UNet3D inserts a single transformer encoder layer at the bottleneck (four heads, feed-forward dimension 512, dropout 0.25) with a learnable three-dimensional positional embedding. All normalisation layers in all three networks are group normalisation (8 groups); no batch normalisation is used anywhere. Group normalisation computes its statistics within each individual sample, independently of batch size and of the other samples in the mini-batch, so the micro-batch size of 6 (with eight gradient-accumulation steps, effective batch 48) affects only the optimiser update and not the normalisation statistics, and no batch statistics can cross the train/test boundary. The Voltron ensemble output at every voxel is the arithmetic mean of the three sigmoid-bounded predictions. Ensembling was applied only at inference time and did not require additional training.

Training Protocol and Data Partitioning

Evaluation used leave-one-anatomy-out cross-validation across four anatomies: the synthetic TG-119 benchmark and three anthropomorphic cases (HEAD-AND-NECK, PROSTATE and LIVER). For each held-out anatomy, the networks were trained on photon plans of the remaining anatomies only (the homogeneous BOXPHANTOM was retained in the training pool but not used as a test fold, as it is not anthropomorphic), and then evaluated zero-shot on all three modalities and all six configurations of the held-out anatomy. Crucially, only photon plans were used during training; proton and carbon-ion plans were never seen by any network at any stage. This yields 18 test predictions per held-out anatomy (72 in total).

Training used random three-dimensional patches of size $128 \times 128 \times 128$ voxels, with 70% probability centred on a PTV voxel and 30% on a random foreground voxel. On-the-fly augmentation was performed with MONAI 1.4 (26) and comprised random spatial flips, random affine transformations, random Gaussian noise and smoothing, and random intensity shifts. Each network was warm-started from the weights of a previously trained photon-only baseline model and then trained with the AdamW optimiser (27), initial learning rate 1×10^{-4}, weight decay 5×10^{-4} and a cosine-annealing-with-warm-restarts schedule, for up to 25 epochs with early stopping (patience 6). The mini-batch size was 6 with eight gradient accumulation steps (effective batch size 48). Mixed-precision training was used with PyTorch automatic mixed precision and gradient scaling. The loss was a composite of mean-squared-error, a gradient-weighted term emphasising high-gradient (penumbra) regions, a dose-volume-histogram-quantile term and a boundary term. Random seeds were fixed to 42. Final training losses by backbone and held-out fold are reported in Figure 2a.

Inference and Evaluation Metrics

At inference time, the held-out anatomy was processed with three-dimensional sliding-window inference (MONAI 1.4) using overlapping $128 \times 128 \times 128$ patches at 50% overlap with Gaussian importance weighting. The Voltron output was rescaled to physical dose by multiplication with the maximum reference dose of the corresponding matRad plan. Predicted dose distributions were compared to matRad references using two metrics. The first was the three-dimensional gamma index computed with the open-source PyMedPhys implementation using global dose normalisation, with criteria of 3%/3 mm and a 10% low-dose cut-off (25,28); the dose-difference criterion was set to 3% of the maximum reference dose, and the 3 mm distance-to-agreement criterion was translated into voxel radii using the actual physical voxel spacing. A stricter 2%/2 mm criterion with the same 10% low-dose cut-off was additionally evaluated. Gamma values reported in an earlier version of this work had been computed with a non-standard, lenient voxel-wise criterion and are given in the Results for reference only. The second metric was the mean absolute error in Gy between the predicted and reference dose, computed over all voxels and, regionally, within the PTV and the distal dose fall-off region. Evaluations 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). Plots were generated with Matplotlib 3.9. Per-configuration gamma and MAE values for the 18 modality-configuration combinations, each averaged over the four held-out anatomies, are reported in Supplementary Tables S4–S6.

Results

The complete experimental workflow is summarised in Figure 1. Five anatomies and three modalities under six beam configurations produced 90 reference dose cubes; in each leave-one-anatomy-out split, only the photon plans of the remaining anatomies were used to train the Voltron ensemble. Across the four held-out anatomies, 72 test plans (four anatomies $\times$ three modalities $\times$ six configurations) constituted the evaluation set, the photon plans serving as the in-modality baseline and the proton and carbon-ion plans serving as zero-shot evaluations.

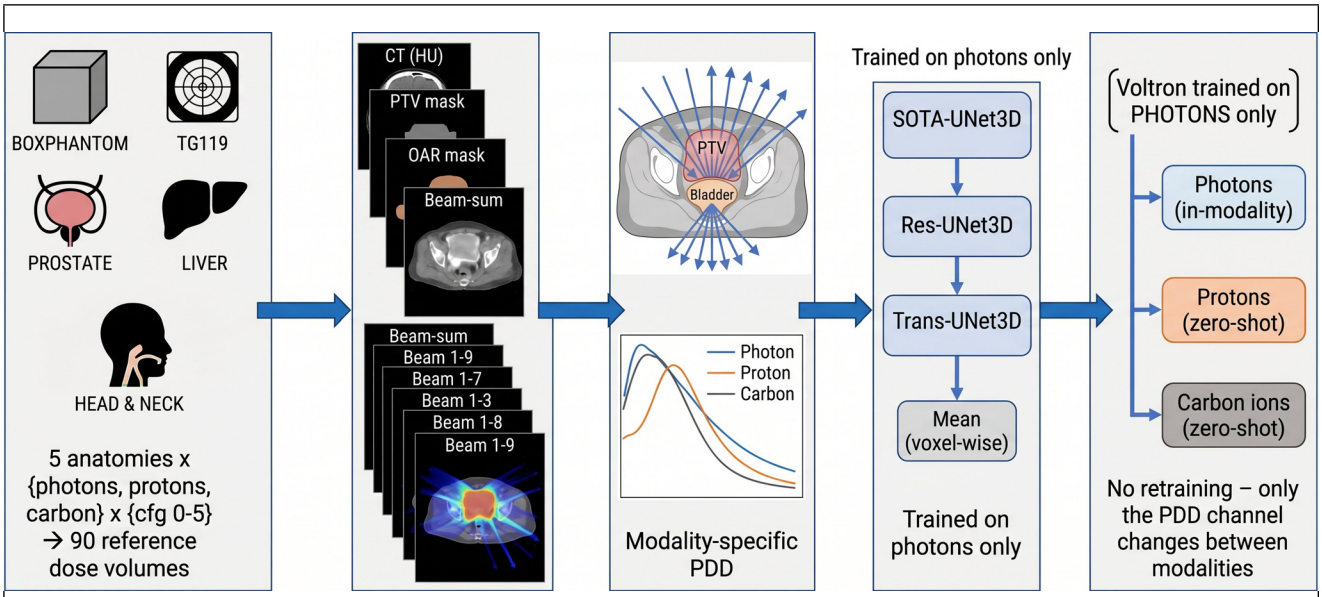

Figure 1. Workflow of the Voltron ensemble for multi-modality dose prediction. Five anatomies (BOXPHANTOM, TG-119, PROSTATE, LIVER and HEAD_AND_NECK) and six beam configurations were combined with three modalities (photons, protons and carbon ions) to produce 90 reference dose volumes using matRad. The 13-channel input tensor (CT image, PTV and OAR masks, beam-sum and per-beam ray-tracing channels with modality-specific depth-dose profile) was fed to three complementary three-dimensional networks (SOTA-UNet3D, Res-UNet3D and Trans-UNet3D) whose outputs were averaged voxel-wise to form the Voltron prediction. The ensemble was trained only on photon plans; modality changes between predictions only through the per-beam depth-dose channel.

All three networks converged under early stopping (maximum 25 epochs, warm-started from the photon-only baseline weights; Figure 2a). Res-UNet3D reached the lowest final training loss (composite ≈ 0.003), while SOTA-UNet3D (≈ 0.028) and Trans-UNet3D (≈ 0.023) converged to values roughly an order of magnitude higher. 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 to multi-scale or transformer-equipped blocks. Averaged over the four held-out anatomies, the Voltron ensemble achieved standard gamma (3%/3 mm) pass rates of 23.5% for photons (in-modality baseline), 23.3% for protons (zero-shot) and 21.5% for carbon ions (zero-shot) (Figure 2b, 2c). For reference, the non-standard lenient criterion used in an earlier version of this work gives 57.3%, 53.1% and 51.7% for the same predictions, and the values reported in that earlier version were 73.2%, 70.8% and 69.1%; the large gap is attributable mainly to the corrected metric (see Methods). The stricter 2%/2 mm standard gamma gives 15.8%, 15.5% and 14.5%.

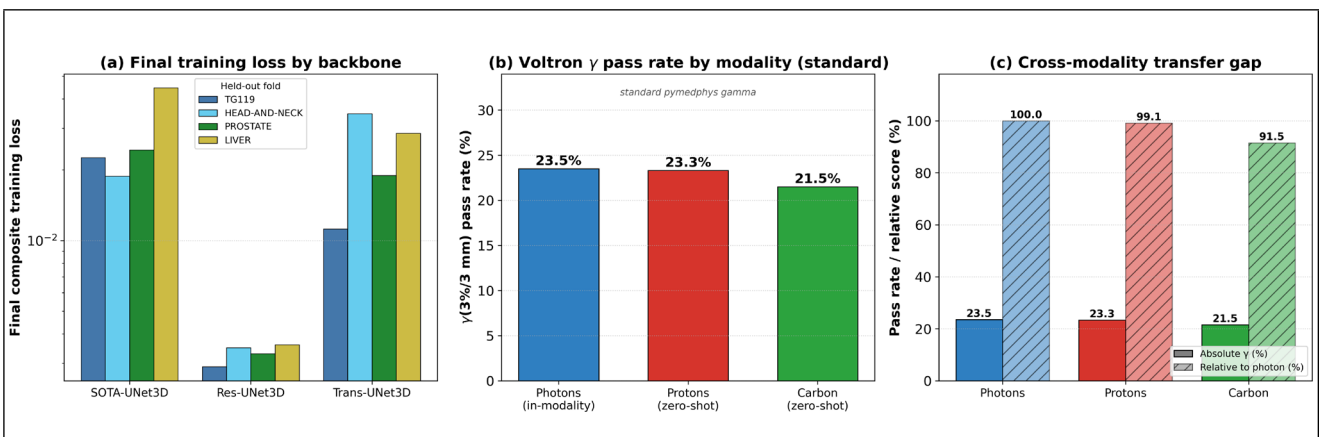

Figure 2. Convergence and overall performance of the Voltron ensemble. (a) Final composite training loss of the three backbones (SOTA-UNet3D, Res-UNet3D, Trans-UNet3D) for each held-out fold, log scale. (b) Modality-averaged three-dimensional gamma (3%/3 mm) pass rate (mean over the six beam configurations and the four held-out anatomies) for photons (in-modality baseline), protons (zero-shot) and carbon ions (zero-shot); values are the standard pymedphys gamma. (c) Cross-modality transfer gap: solid bars show absolute gamma pass rates; hatched bars show the relative score with respect to the photon baseline (=100%).

Examining the standard-gamma results in detail, the Voltron ensemble (arithmetic mean) achieved photon (in-modality) pass rates of 23.5% (3%/3 mm) and 15.8% (2%/2 mm), with per-anatomy values ranging from 15.2% (LIVER) to 29.7% (TG-119); per-configuration values are shown in Figure 3 and Supplementary Table S4. For zero-shot protons the pass rate was 23.3% (3%/3 mm) and for zero-shot carbon ions 21.5% (Supplementary Tables S5-S6), i.e. within a few percentage points of the in-modality photon baseline on the same standard metric. A per-model breakdown shows that the best single backbone is Trans-UNet3D (24.1% averaged over modalities and anatomies); the mean Voltron ensemble (22.8%) performs comparably while reducing inter-model variance, and a median-fusion variant (23.8%) recovers essentially all of the best single-model performance, so ensembling here improves robustness rather than raising the peak score. The mean absolute error was lowest in the PTV (about 4.9–5.2 Gy) and largest in the distal fall-off region (about 8.5–10.4 Gy), the distal error increasing from photons to protons to carbon ions, which localises the transfer gap to the steep-gradient region where the sharper particle Bragg peaks are hardest to reproduce. Dose-volume-histogram analysis shows the predicted PTV mean dose tracking the reference within about 2 Gy while the high-dose coverage tails (D95/D98) are under-predicted, consistent with a plausible but over-smoothed feasibility-stage prediction. Per-configuration values are provided in Supplementary Tables S4–S6 and Supplementary Figures S3–S4.

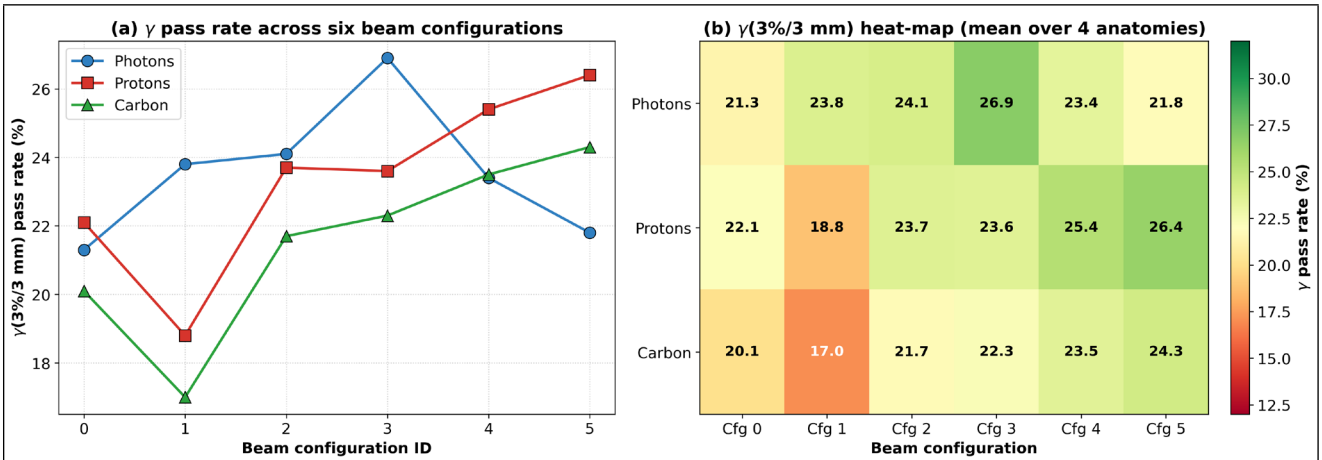

Figure 3. Per-beam-configuration performance of the Voltron ensemble, averaged over the four held-out anatomies. (a) Standard three-dimensional gamma (3%/3 mm) pass rate of each modality across the six beam configurations. (b) Heat-map of the same data (three modalities × six configurations, mean over the four held-out anatomies); numbers in each cell are pass rates in percent. Configuration indices match Supplementary Table S1.

The mean absolute error distributions overlapped substantially across the three modalities (Figure 4a). Mean MAE was approximately 5.2–5.3 Gy across modalities, while individual plans spanned approximately 3.1-7.1 Gy, indicating that the absolute magnitude of voxel-wise error did not grow markedly when the model was applied to particle dose data it had never seen during training. MAE and gamma pass rate were not monotonically related (Figure 4b): points formed distinct MAE clusters, within which gamma pass rates varied widely; there was no systematic separation between in-modality and zero-shot points.

Figure 3b summarises the 18 modality-configuration combinations (each averaged over the four held-out anatomies) in a single heat-map, and Figure 5 shows the distribution of gamma pass rates across all 72 test plans stratified by modality. Two patterns are visible. First, no row is uniformly red, indicating that the ensemble does not catastrophically fail on any single modality. Second, configuration dependence differed between modalities: for photons, configuration 3 (nine equispaced beams) produced the highest pass rate (26.9%) and configuration 0 (five equispaced beams) the lowest (21.3%), whereas for protons and carbon ions configuration 1 (six equispaced beams) produced the lowest pass rates (18.8% and 17.0%) and configuration 5 (four oblique beams; Supplementary Figure S1) the highest (26.4% and 24.3%).

Taken together, relative to the in-modality photon baseline, the photon-only Voltron ensemble retained approximately 99.1% of the standard gamma pass rate when applied zero-shot to protons and approximately 91.5% when applied zero-shot to carbon ions across the four held-out anatomies. The transfer gap is small considering that, apart from the modality-specific depth-dose channels, no information about the target modality entered the model during training.

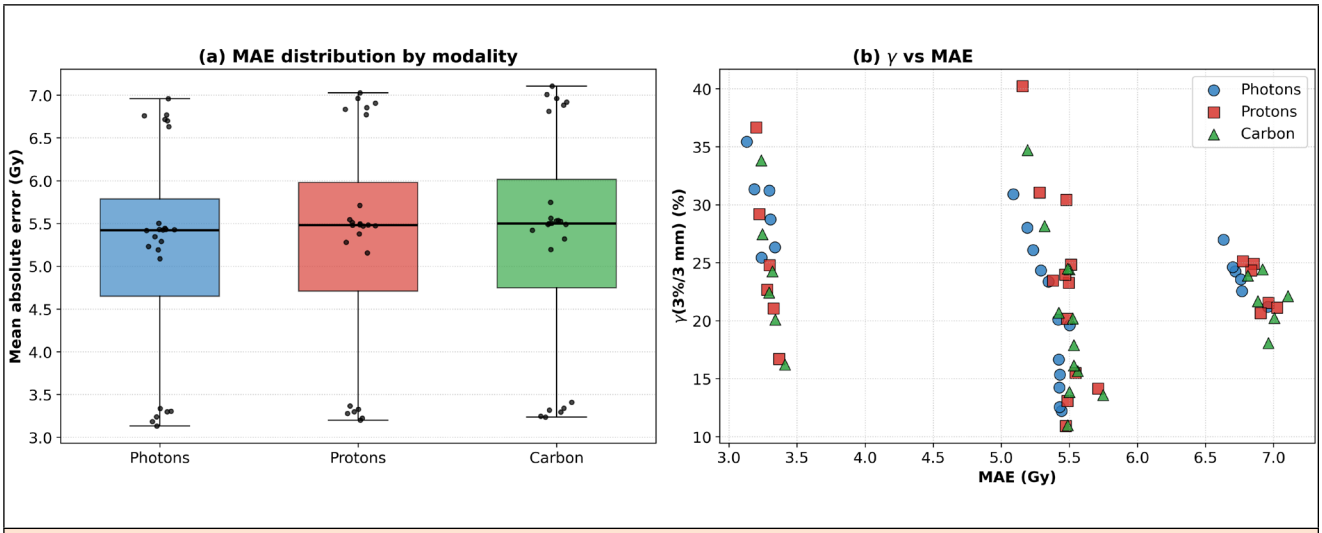

Figure 4. Error analysis of the Voltron ensemble on the four held-out anatomies. (a) Distribution of mean absolute error (MAE, in Gy) across all 24 test plans per modality (four held-out anatomies × six beam configurations); boxes show the interquartile range, the horizontal line is the median, whiskers extend to 1.5 × interquartile range, and individual points are overlaid. (b) Scatter plot of three-dimensional gamma (3%/3 mm) pass rate versus MAE across all 72 test plans (four held-out anatomies); points cluster by MAE level and no monotonic relationship is evident.

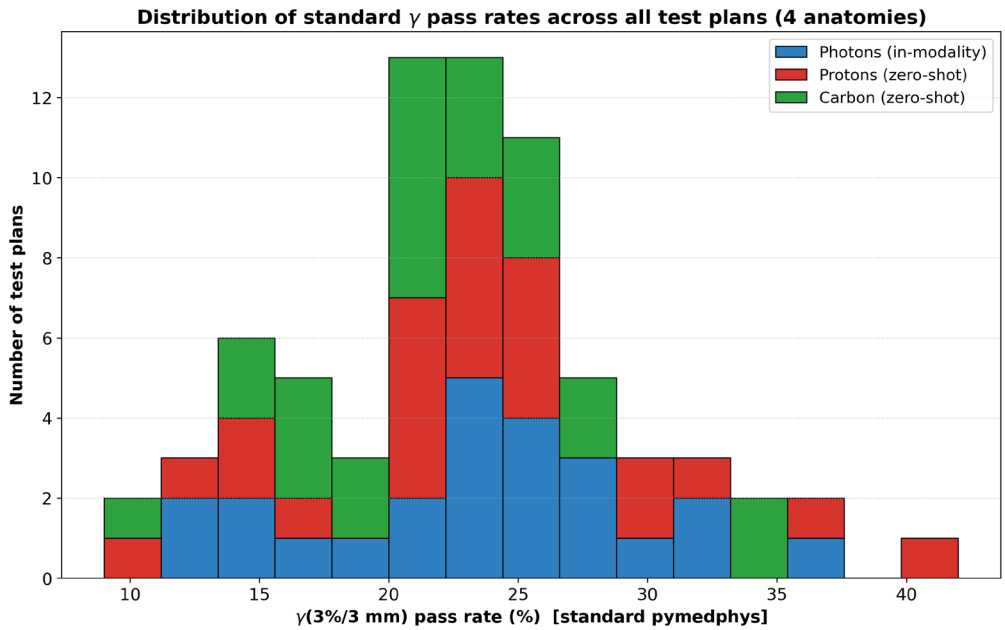

Figure 5. Distribution of standard three-dimensional gamma (3%/3 mm) pass rates across all 72 test plans (four held-out anatomies), stratified by modality (photons in-modality, protons zero-shot and carbon ions zero-shot). The overlapping distributions across the three modalities illustrate that zero-shot particle predictions span a similar performance range to the in-modality photon baseline.

Discussion

We trained a single ensemble of three complementary three-dimensional networks on photon plans only and showed that, when proton or carbon-ion ray-tracing channels are supplied at inference time, the same ensemble produces zero-shot particle dose distributions whose gamma pass rate is within a few percentage points of the in-modality photon baseline. To our knowledge, this is the first explicit dosimetric demonstration of photon-to-particle zero-shot transfer using a physics-informed input representation. The result supports the view that, when modality differences are expressed through a small set of physically meaningful input channels, much of the information a network needs to predict the dose is anatomical and geometric and can be shared across modalities.

Our design deliberately disentangles anatomy, geometry and modality-specific physics. The computed-tomography image and structure masks are identical between modalities for any given plan; the gantry-angle distribution enters the model through the same per-beam ray-tracing channels for all modalities; and only the one-dimensional depth-dose profile along each beam differs between photons and particles. This is a much weaker form of modality awareness than would be required if the network had to learn the Bragg peak from data. Because the physics model is supplied as part of the input, the network only needs to learn how a given anatomical configuration modulates an idealised depth-dose profile into an actual three-dimensional dose distribution. Our results suggest that this anatomy-conditional dose-shaping function is, to first approximation, modality-invariant for the treatment-planning regimes implemented in matRad (21).

The drop in gamma pass rate from photons (23.5%) to protons (23.3%) and carbon ions (21.5%) is on the order of a few percentage points, consistent with the increasing sharpness of the Bragg peak: photons have a broad depth-dose profile well approximated by the exponential model used in our depth-dose channel, protons have a moderately sharper peak that the ensemble must reshape from training distributions it has only seen on photons, and carbon ions have the sharpest peak and the steepest distal fall-off. The pattern is also consistent with the body of literature on range and dose-calculation uncertainties in proton therapy (16,20). The photon-trained model is also not equally good across photon configurations: configurations 0 and 5 produced the lowest mean photon gamma pass rates (21.3% and 21.8%, respectively), while configuration 3 produced the highest (26.9%), indicating that part of the remaining error is geometric rather than modality-related.

The Voltron ensemble combines a residual U-Net that converges to the lowest training loss with two complementary estimators (a multi-scale attention-gate U-Net and a transformer-bottleneck U-Net). The three backbones were combined by simple unweighted averaging to reduce model-specific variability without introducing an additional trained meta-model. We did not pursue weighted voting or stacking, to keep the pipeline reproducible and inexpensive. That this simple unweighted average already achieves consistent zero-shot performance suggests that further gains from smarter ensembling are likely possible.

Several limitations should be acknowledged. First, the reference doses come from an analytical pencil-beam toolkit rather than from full Monte Carlo simulation. matRad has been validated against commercial treatment planning systems (21), but analytical algorithms are known to be less accurate than Monte Carlo in highly heterogeneous geometries (16); our results should therefore be read as a controlled algorithm-to-algorithm comparison. Second, the evaluation comprises only four held-out anatomies from a public demonstration dataset (one synthetic TG-119 benchmark and three anthropomorphic cases: HEAD-AND-NECK, PROSTATE and LIVER), which is sufficient to demonstrate that zero-shot transfer is feasible but not to characterise inter-patient variability or establish clinical generalisability. Third, the absolute gamma pass rates are below the threshold typically required for clinical patient-specific quality assurance (commonly above 90%); the aim of this study, however, was not clinical-grade particle dose prediction but a transparent test of whether photon-only training combined with a physics-informed input could yield non-trivial zero-shot performance. Fourth, our depth-dose model is a simplified analytical approximation; replacing it with tabulated stopping-power-based depth-dose curves or a learned ray-tracing module could plausibly close part of the remaining gap.

Several future directions follow naturally. The most direct extension is to validate the same ensemble on a larger and more heterogeneous patient cohort and against Monte Carlo reference doses where available. A particularly informative experiment would be to add a small number of particle plans for fine-tuning and measure how much particle-specific data are needed to close the remaining gap, which would quantify the practical value of the zero-shot starting point for data-scarce centres. The same input representation could be combined with more accurate depth-dose models or with a fast learned dose-calculation engine in the spirit of Pastor-Serrano and Perkó (13). Beyond dose prediction, the broader idea that modality differences can be encapsulated in a physics-aware input layer rather than in modality-specific weights may be transferable to other areas of medical imaging where data scarcity in one modality coexists with data abundance in another (19).

Conclusion

An ensemble of three complementary three-dimensional networks trained exclusively on photon plans produced

standard gamma (3%/3 mm) pass rates of 23.3% for protons and 21.5% for carbon ions across four held-out anatomies, compared with 23.5% for the in-modality photon baseline. The relative degradation when moving from photons to particles was small. These pass rates are well below clinical-grade performance, and we therefore present the study explicitly as an algorithmic feasibility study rather than a clinical validation. Within that scope, the result suggests that, with a physics-informed input representation, a substantial fraction of the dose-prediction task is modality-invariant. Photon-only training may therefore offer a useful starting point for particle therapy research in data-scarce settings and provides a transparent, fully-reproducible baseline for future cross-modality methods.

Declarations

Funding

The author declares that this study received no financial support.

Conflicts of Interest/Competing Interests

The author has declared that no conflict of interest exists.

Ethics Approval

Not applicable. This study involved only in-silico simulation on a public phantom and demonstration dataset; therefore, no ethics committee approval was required.

Informed Consent

Not applicable. This study involved only in-silico simulation on a public phantom and demonstration dataset; therefore, no informed consent was required.

Availability of Data and Material

The data that support the findings of this study are available from the corresponding author upon reasonable request. All reference dose volumes used in this study were generated with the publicly available matRad toolkit (https://github.com/e0404/matRad). The Supplementary Materials document also contains data and supporting information used in this study. Supplementary Materials are available with the online version of this article.

Authors' Contributions

M.E.B. is the sole author of this study and was responsible for all aspects of the research, including the Concept, Design, Supervision, Materials, Data Collection or Processing, Analysis and/or Interpretation, Literature Review, Writing, and Critical Review.

Declaration of Generative AI and AI-Assisted Technologies

While the core scientific focus of this research involves the development and evaluation of artificial intelligence (specifically, deep learning models for dose prediction), this declaration addresses the use of generative AI tools in the preparation of the manuscript text. During the preparation of this work, the author used Gemini to improve language readability, polish the English phrasing and ensure grammatical correctness. After using this tool, the author closely reviewed and edited the text as needed. The AI-assisted technologies were strictly limited to language editing and were not used in the conceptualisation, execution of the research or data generation. The author takes full responsibility for the final content of the publication and confirms that all ethical guidelines regarding scientific integrity have been strictly observed.

Acknowledgement

The author thanks the matRad development group at the German Cancer Research Center (DKFZ, Heidelberg, Germany) for making the matRad toolkit and its example dataset publicly available, which made this study possible.

References

1. Atun R, Jaffray DA, Barton MB, et al. Expanding global access to radiotherapy. Lancet Oncol. 2015;16:1153–86. DOI:10.1016/S1470-2045(15)00222-3
2. Kamada T, Tsujii H, Blakely EA, et al. Carbon ion radiotherapy in Japan: an assessment of 20 years of clinical experience. Lancet Oncol. 2015;16:e93–100. DOI:10.1016/S1470-2045(14)70412-7
3. Tsujii H and Kamada T. A review of update clinical results of carbon ion radiotherapy. Jpn J Clin Oncol. 2012;42:670–85. DOI:10.1093/jjco/hys104
4. Schulz-Ertner D and Tsujii H. Particle radiation therapy using proton and heavier ion beams. J Clin Oncol. 2007;25:953–64. DOI:10.1200/JCO.2006.09.7816
5. Mohan R and Grosshans D. Proton therapy – present and future. Adv Drug Deliv Rev. 2017;109:26–44. DOI:10.1016/j.addr.2016.11.006

6. Ronneberger O, Fischer P and Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, et al., eds. MICCAI 2015. Cham: Springer; 2015. p. 234–41. DOI:10.1007/978-3-319-24574-4_28
7. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, et al., eds. MICCAI 2016. Cham: Springer; 2016. p. 424–32. DOI:10.1007/978-3-319-46723-8_49
8. Babier A, Zhang B, Mahmood R, et al. OpenKBP: the open-access knowledge-based planning grand challenge and dataset. Med Phys. 2021;48:5549–61. DOI:10.1002/mp.14845
9. Barragán-Montero AM, Nguyen D, Lu W, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. Med Phys. 2019;46:3679–91. DOI:10.1002/mp.13597
10. Nguyen D, Jia X, Sher D, 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;64:065020. DOI:10.1088/1361-6560/ab039b
11. Liu Z, Fan J, Li M, et al. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. Med Phys. 2019;46:1972–83. DOI:10.1002/mp.13490
12. Fan J, Wang J, Chen Z, et al. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys. 2019;46:370–81. DOI:10.1002/mp.13271
13. Pastor-Serrano O and Perkó Z. Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy. Phys Med Biol. 2022;67:105006. DOI:10.1088/1361-6560/ac692e
14. He R, Zhang H, Wang J, et al. Deep learning-based prediction of Monte Carlo dose distribution for heavy ion therapy. Phys Imaging Radiat Oncol. 2025;34:100735. DOI:10.1016/j.phro.2025.100735
15. Li M, Choulilitsa E, Fankhauser L, et al. Neural network-driven direct CBCT-based dose calculation for head-and-neck proton treatment planning. Phys Med Biol. 2025;70:235027. DOI:10.1088/1361-6560/ae222a
16. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol. 2012;57:R99–117. DOI:10.1088/0031-9155/57/11/R99
17. Kieselmann JP, Fuller CD, Gurney-Champion OJ, et al. Cross-modality deep learning: contouring of MRI data from annotated CT data only. Med Phys. 2021;48:1673–84. DOI:10.1002/mp.14619
18. Chen C, Dou Q, Chen H, et al. Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. Proc AAAI Conf Artif Intell. 2019;33:865–72. DOI:10.1609/aaai.v33i01.3301865
19. Dorent R, Kujawa A, Ivory M, et al. CrossMoDA 2021 challenge: benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. Med Image Anal. 2023;83:102628. DOI:10.1016/j.media.2022.102628
20. Newhauser WD and Zhang R. The physics of proton therapy. Phys Med Biol. 2015;60:R155–209. DOI:10.1088/0031–9155/60/8/R155
21. Wieser H-P, Cisternas E, Wahl N, et al. Development of the open-source dose calculation and optimization toolkit matRad. Med Phys. 2017;44:2556–68. DOI:10.1002/mp.12251
22. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. 2016 IEEE Conf Comput Vis Pattern Recognit (CVPR). 2016:770–8. DOI:10.1109/CVPR.2016.90
23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Advances in Neural Information Processing Systems. Vol. 30. Curran Associates; 2017. p. 5998–6008.
24. Hatamizadeh A, Tang Y, Nath V, et al. UNETR: transformers for 3D medical image segmentation. 2022 IEEE/CVF Winter Conf Appl Comput Vis (WACV). 2022:1748–58. DOI:10.1109/WACV51458.2022.00181
25. Low DA, Harms WB, Mutic S, et al. A technique for the quantitative evaluation of dose distributions. Med Phys. 1998;25:656-61. DOI:10.1118/1.598248
26. Cardoso MJ, Li W, Brown R, et al. MONAI: an open-source framework for deep learning in healthcare. arXiv. 2022. DOI:10.48550/arXiv.2211.02701
27. Loshchilov I and Hutter F. Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR); 2019.
28. Hussein M, Clark CH and Nisbet A. Challenges in calculation of the gamma index in radiotherapy — towards good practice. Phys Med. 2017;36:1–11. DOI:10.1016/j.ejmp.2017.03.001