



A novel method for EPID transmission dose generation using Monte Carlo simulation and deep learning

Tao Qiu¹ · Ning Gao² · Yan-Kui Chang² · Xi Pei² · Huan-Li Luo¹ · Fu Jin¹

Received: 28 May 2025 / Revised: 8 August 2025 / Accepted: 3 September 2025 / Published online: 29 January 2026

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2026

Abstract

This study aimed to integrate Monte Carlo (MC) simulation with deep learning (DL)-based denoising techniques to achieve fast and accurate prediction of high-quality electronic portal imaging device (EPID) transmission dose (TD) for patient-specific quality assurance (PSQA). A total of 100 lung cases were used to obtain the noisy EPID TD by the ARCHER MC code under four kinds of particle numbers (1×10^6 , 1×10^7 , 1×10^8 and 1×10^9), and the original EPID TD was denoised by the SUNet neural network. The denoised EPID TD was assessed both qualitatively and quantitatively using the structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and gamma passing rate (GPR) with respect to 1×10^9 as a reference. The computation times for both the MC simulation and DL-based denoising were recorded. As the number of particles increased, both the quality of the noisy EPID TD and computation time increased significantly (1×10^6 : 1.12 s, 1×10^7 : 1.72 s, 1×10^8 : 8.62 s, and 1×10^9 : 73.89 s). In contrast, the DL-based denoising time remained at 0.13–0.16 s. The denoised EPID TD shows a smoother visual appearance and profile curves, but differences between 1×10^6 and 1×10^9 still remain. SSIM improves from 0.61 to 0.95 for 1×10^6 , 0.70 to 0.96 for 1×10^7 , and 0.90 to 0.97 for 1×10^8 . PSNR increases by > 20% for 1×10^6 and 1×10^7 , and > 10% for 1×10^8 . GPR improves from 48.47% to 89.10% for 1×10^6 , 61.04% to 94.35% for 1×10^7 , and 91.88% to 99.55% for 1×10^8 . The method that combines MC simulation with DL-based denoising for EPID TD generation can accelerate TD prediction and maintain high accuracy, offering a promising solution for efficient PSQA.

Keywords PSQA · EPID · Monte Carlo · Deep learning

1 Introduction

With the growing adoption of advanced techniques such as intensity-modulated radiation therapy (IMRT), volumetric-modulated arc therapy (VMAT), and adaptive radiation

therapy (ART), patient-specific quality assurance (PSQA) is routinely implemented to ensure the accuracy of both treatment planning and dose delivery [1–4]. Numerous studies have explored the role of PSQA and have reported promising developments [5–8]. However, most existing studies have primarily focused on pretreatment PSQA and are limited in their ability to capture inter- and intra-fractional variations. In recent years, the emphasis has shifted toward real-time in vivo dose verification. Among the various available tools, the electronic portal imaging device (EPID) has emerged as a practical and efficient solution over traditional tools, such as films and ion chambers, owing to its superior operational efficiency and ease of use, as highlighted in the AAPM TG-58 report, which recognizes the ability of EPIDs to provide quantitative data and real-time feedback [9–13]. However, the EPID-based transmission dose (TD) is often affected by statistical noise when generated by Monte Carlo (MC) simulation under limited particle settings to reduce the computation time. Noise can significantly degrade the image

Tao Qiu and Ning Gao have contributed equally to this work.

This work was supported by National Key R&D Program of China (No. 2022YFC2404604), Chongqing Research Institution Performance Incentive Guidance Special Project (No. CSTB2023JXJL-YFX0080), and Chongqing Medical Scientific Research Project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (No. 2022DBXM005).

✉ Fu Jin
jfazj@126.com

¹ Departments of Radiation Oncology, Chongqing University Cancer Hospital, Chongqing 400030, China

² School of Nuclear Science and Technology, University of Science and Technology of China, Hefei 230026, China

quality, making it difficult to extract reliable dose information for PSQA. With the development of deep learning (DL) techniques in radiotherapy, these methods provide a promising solution for denoising MC-based dose calculations. Therefore, combining MC simulation with DL-based denoising methods presents an effective approach for enabling accurate dose verification and supporting efficient online ART workflows.

The forward-projection approach is a representative EPID-based PSQA approach that compares measured 2D images or TD with those predicted at the EPID level [14]. It is widely used in clinical practice because it effectively identifies errors related to data transfer, treatment delivery, and anatomical variations, thereby significantly enhancing radiotherapy accuracy. The forward-projection approach relies on both measured and predicted EPID images and the TD. The measured data, representing the actual beam delivery, were acquired as 2D images or dose distributions in calibration units (CU) and could be converted into Gy through appropriate calibration procedures. Predicted data can be generated using dose calculation methods. Dose calculation methods encompass both analytical algorithms and MC simulation. Analytical algorithms have limited capability in accurately modeling tissue heterogeneities, which may lead to dose calculation inaccuracies in complex anatomical areas. MC, widely recognized as the “gold standard” for dose calculation, serves to simulate particle transport processes to closely approximate actual physical interactions, which can provide high accuracy for EPID TD prediction. Currently, two types of MC methods are used to predict the EPID TD. The first is the kernel convolution method, which simulates the dose kernel on the EPID plane and convolves it with the fluence map to obtain dose deposition [15, 16]. The second is the full MC method, which simulates the entire process of particle transport through the accelerator and dose deposition in the EPID [17]. Both methods rely on precise modeling of the accelerator and EPID; however, the EPID model is typically proprietary, and its construction requires specialized expertise. Therefore, finding an appropriate MC simulation tool is crucial. The details of the MC simulation tool used in this study are presented in the next section.

Although MC simulation can yield accurate results, achieving clinically acceptable statistical precision requires simulating a large number of particles. This process consumes substantial time and computational resources, making it challenging to apply in clinical practice, particularly in the context of online ART, which relies on specialized software and hardware platforms with streamlined and automated workflows. It utilizes updated anatomical images and contours to reconstruct the delivered dose, based on the original treatment plan. After evaluating the reconstructed dose on the new anatomy, a decision is made regarding whether adaptation is required. If so, a new treatment plan is

generated based on this new anatomy. The entire online ART process includes image generation (obtaining new anatomy), contour generation (OAR and targets), dose reconstruction, and new treatment planning on new anatomy. Before the adapted plan is delivered, comprehensive QA must be performed. The total online ART workflow will take several to tens of minutes, and the time allocated for QA should be minimized. Excessive delays during this phase can lead to patient discomfort or movement, potentially causing intra-fractional anatomical variations (e.g., bladder filling or respiratory motion), which may compromise the dosimetric benefits of ART and reduce clinical efficiency [18–20]. For QA in online ART, the EPID serves as a critical clinical tool owing to its ability to perform rapid, noninvasive verification of delivered dose distributions, which enables efficient and time-sensitive PSQA. Therefore, the rapid and accurate generation of EPID-based TD is essential to support the implementation of online ART.

Simulating fewer particles can shorten the computation time, but statistical uncertainty may be introduced, causing noise in the EPID TD. If the noise can be effectively mitigated while preserving the advantage of reduced computation time under a low particle number, the MC-based forward-projection approach could become clinically viable. Traditional denoising methods are typically algorithm-driven and are often referred to as prior- or model-based approaches, such as self-similarity, sparse coding, and total variations. Although these methods are effective in addressing ill-posed problems, they often involve substantial computational costs and may struggle in complex or highly noisy scenarios [21–23]. With the advancement of DL, its applications in radiotherapy have expanded to include automatic delineation, automated planning, and image registration. Zhou et al. developed and tested a 3D DL model for predicting 3D voxel-wise dose distributions for IMRT [24]. Xing et al. attempted to resolve the dilemma that fast algorithms were generally less accurate, whereas accurate dose engines were often time-consuming, by exploring DL for dose calculation [25]. Zhang et al. developed a slice classification model-facilitated 3D encoder–decoder network for segmenting organs-at-risk in head and neck cancer [26]. Zhen et al. proposed a deep convolutional neural network with transfer learning for rectal toxicity prediction in cervical cancer radiotherapy [27]. However, few studies have been conducted on denoising EPID TD using DL. To address the noise under low particle numbers in MC simulation, DL can be employed to denoise by learning the mapping between low-particle-number EPID TD and high-particle-number EPID TD. Its goal is to establish a function F that corrects a noisy low-quality TD_l to a high-quality TD_h [28].

$$TD_h = F(TD_l)$$

U-Net is one of the most widely used DL architectures. To meet various demands, numerous U-Net variants have been

developed, including Dense-UNet, U-Net++, U-Net3+, and 3D-UNet [29–31]. Swin-UNet was proposed as a Transformer-based image processing network that replaces traditional convolutional layers in the classic U-Net with Swin Transformer models. Recently, modifications based on Swin-UNet have been implemented to tailor it for denoising tasks, resulting in the development of the SUNet, which incorporates several denoising-specific enhancements, including a dual upsampling module designed to mitigate checkerboard artifacts and strengthen spatial detail reconstruction [32–34]. These improvements enabled SUNet to effectively suppress noise in the EPID TD while preserving structural fidelity. Although SUNet was originally developed for natural image denoising, its direct application to EPID TD denoising demonstrates meaningful potential for extension into the medical imaging domain. More details and implementation strategies of the SUNet are elaborated in the next section to provide a comprehensive understanding of its design and application.

This study aims to apply SUNet to denoise the EPID TD generated by the MC simulation with a low particle number, enhancing the low-quality TD to a high-quality one while reducing the computation time without compromising the accuracy. First, the EPID TD was generated for all fields at different particle numbers: 1×10^6 , 1×10^7 , 1×10^8 , and 1×10^9 . Existing research indicates that while particle numbers exceeding approximately 1×10^9 continue to enhance the simulation accuracy, the improvement becomes clinically insignificant when considering the substantial increase in computational requirements [28]. Therefore, 1×10^9 was selected as a reasonable upper limit for the MC simulation and was used as the ground truth. Then, three separate SUNet models were trained for 1×10^6 , 1×10^7 , and 1×10^8 . After training, the denoising results were assessed qualitatively and quantitatively to identify the optimal trade-off between computational efficiency and dosimetric accuracy across different particle numbers, thus providing a suitable particle number for future research and evidence-based guidance for practical clinical applications.

2 Material and methods

We collected data from 100 patients with lung cancer who underwent IMRT with a five-field beam arrangement, yielding 500 fields in the dataset. Treatment plans were completed using the Eclipse 15.6 treatment-planning system (Varian Medical Systems, Palo Alto, CA, USA). The corresponding RT files, including RT-Plan, RT-Structure, and RT-Dose, along with CT images, were exported as input data for MC simulation. The employed accelerator model was the Varian TrueBeam operating in a 6 MV flattened photon beam mode.

2.1 Monte Carlo simulation for EPID TD

ARCHER, a GPU-accelerated fast MC code, was employed to generate the EPID TD for training and testing the neural network model. ARCHER has been validated in various radiotherapy dose calculation studies [35–38]. Recently, it has been expanded to EPID dosimetry, providing an efficient and accurate platform for simulating radiation transport through the treatment head, patient phantom, and EPID model [39]. The framework of ARCHER for TrueBeam accelerator is illustrated in Fig. 1.

ARCHER developed a treatment head model of the Varian TrueBeam linear accelerator to facilitate radiation transport simulations. The model incorporates a High-Definition Multi-Leaf Collimator (HDMLC) with 60 leaf pairs, consisting of 32 central pairs with a width of 0.25 mm and 14 pairs on each side with a width of 0.5 mm, providing a total treatment field width of 22 cm. The treatment head describes the process by which electrons are emitted from the accelerating waveguide and accelerated to a high energy. These electrons then strike a metal target, generating secondary photons and electrons. Photons are used for irradiation, whereas secondary electrons contribute to the buildup dose near the surface. The generated photons are collimated and shaped to form the desired treatment beam, which is delivered to the patient for radiotherapy purposes. Because the upper accelerator components are independent of the patient and solely related to the accelerator's mechanical structure, the same dataset can be used for all simulations. To save computation time, the electron-target interaction process was replaced by a phase space file for the radiation source, which was generated using the MC code BEAMnrc. Various modeling components, including SLAB, CONS3R, FLATFILT, CHAMBER, and MIRROR, were used to define the target, primary collimator, ion chamber, flattening filter, and mirror, respectively [35, 36, 39]. Subsequently, ARCHER simulates radiation transport through the treatment head and secondary collimation system, including the jaws and MLC, using an explicit approximation transport. The transported particles that pass through the MLC are then used for subsequent radiation transport simulations in the patient phantom and EPID models.

Three phase space planes are defined in ARCHER: (1) Phase space 1 is located downstream of the treatment head but before the secondary collimators. (2) Phase space 2 is positioned after the secondary collimators but before reaching the patient's body. (3) Phase space 3 is located after the patient. The spatial locations of these phase spaces are illustrated in Fig. 1. In this study, the recorded particle number corresponds to those in phase space 2. The dose calculation on the EPID is performed in two main steps. The first step simulates particles transport within the patient, following the same procedure as

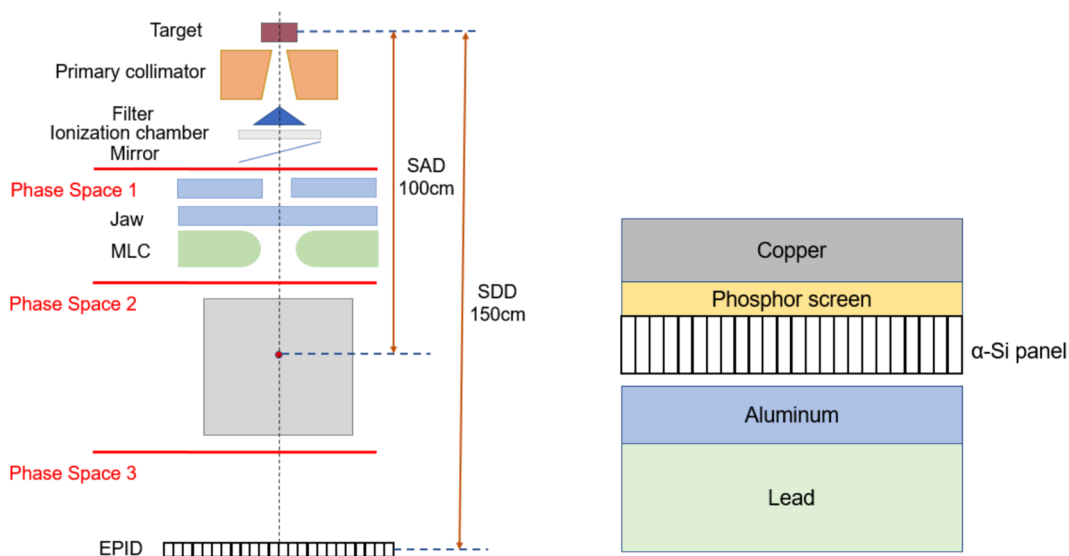


Fig. 1 (Color online) The framework of ARCHER for TrueBeam accelerator. It includes particle transport in the treatment head, patient, and EPID

conventional dose calculations. Through inverse translation and rotation, the transmitted particles are mapped back to the gantry angle of 0° , where they are collected in phase space 3. Finally, the dose deposited by the exiting particles on the EPID was calculated to obtain the TD for each field with a resolution of 160×160 . This resolution was chosen to improve the computational efficiency while maintaining sufficient accuracy in the calculated dose distributions, as supported by a previous study [39].

2.2 Denoising using SUNet for EPID TD

We used the SUNet [33] to denoise the EPID TD generated by a low particle number using ARCHER. The network architecture is shown in Fig. 2, it can be roughly divided into three modules: 1) Shallow feature extraction module. 2)

U-Net feature extraction module. 3) Reconstruction module. Among these three modules, the U-Net feature extraction module is the most critical, and its architecture is similar to that of Swin-UNet, with a specific model to be detailed later. A noisy EPID TD with a resolution of $H \times W$ was input into the network, where a single 3×3 convolutional layer extracted shallow features such as edges, textures, and basic dose distribution information. Although these shallow features are simple, they are vital for subsequent processing because they provide the network with contextual information. These shallow features are then passed to the U-Net feature extraction module, which aims to hierarchically extract multi-scale features from the image. First, the original input was divided into non-overlapping patches of resolution 4×4 , with each patch treated as an independent unit to reduce the computational load and better capture

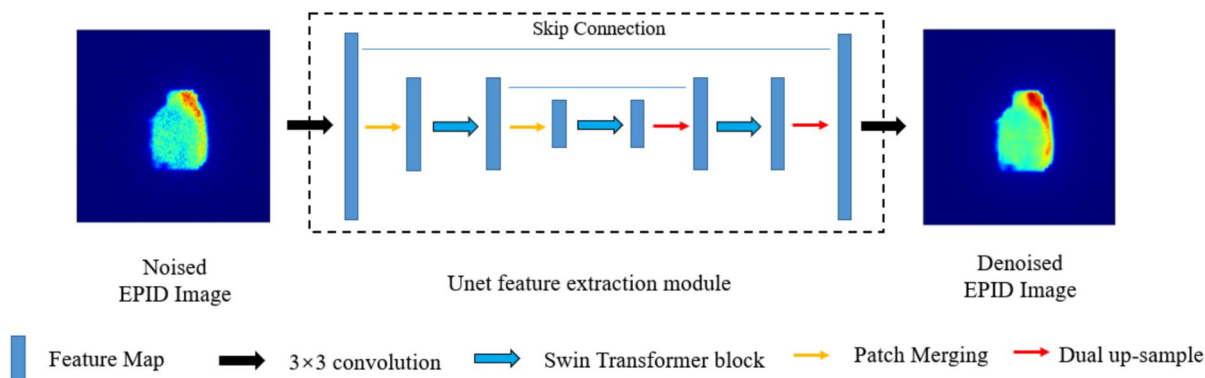


Fig. 2 (Color online) The architecture of SUNet. It includes the following: 1) shallow feature extraction module; 2) U-Net feature extraction module; 3) reconstruction module

local information, resulting in a feature resolution of $H/4 \times W/4$. Subsequently, a patch merging layer merges adjacent 4 patches into a larger patch, reducing the feature resolution by a factor of 2. During the merging process, the features within each patch are concatenated, increasing the new patch’s feature dimension to four times the original. However, to unify the dimensions, the patch merging layer employs a linear layer to transform the concatenated features by adjusting the merged feature dimension to twice the original. This process is repeated multiple times in subsequent network modules, gradually reducing the resolution of the feature maps while increasing their dimensionality, thereby extracting deeper and more multidimensional information.

The Swin Transformer Block (STB) is a key component of the SUNet model, as shown in Fig. 3, replacing the convolutional layers in the original U-Net model. The network uses five STB layers, each containing eight Swin Transformer Layers (STL). Each STL consists of Layer Normalization (LN), Multi-Head Self Attention (MSA), Residual Connection, and a Multi-Layer Perceptron (MLP) with Gaussian Error Linear Unit (GELU) activation. The STB utilizes two attention mechanisms: Window-based Multi-Head Self Attention (W-MSA) and Shifted Window-based Multi-Head Self Attention (SW-MSA). W-MSA reduces computational complexity by performing self-attention within non-overlapping windows, whereas SW-MSA shifts the windows to capture long-range dependencies, enhancing the model’s ability to model global information. The entire process is illustrated as follows:

$$\begin{aligned} \hat{f}^L &= W - MSA(LN(f^{L+1})) + f^{L+1}, \\ f^L &= MLP(LN(\hat{f}^L)) + \hat{f}^L, \\ \hat{f}^{L+1} &= SW - MSA(LN(f^L)) + f^L, \\ f^{L+1} &= MLP(LN(\hat{f}^{L+1})) + \hat{f}^{L+1} \end{aligned}$$

where f^L and \hat{f}^L represent the outputs of the W-MSA module and the MLP module in the L_{th} block, respectively.

To restore the feature map to its original resolution, SUNet introduces a dual upsample method based on two existing upsample methods: Bilinear and PixelShuffle. Compared with the original method, it effectively mitigates checkerboard artifacts. Skip connections fuse high-resolution features from different scales in the encoder with progressively upsampled features in the decoder, helping the

decoder recover fine details and minimizing spatial information loss caused by downsampling.

The resolution of the EPID TD generated by ARCHER was 160×160 , which was resized to 512×512 using opencv library in Python. In this study, 100 cases with 500 fields were collected, 90 cases with 450 fields were used as the training set, and the remaining 10 cases with 50 fields were used as the testing set. Three separate models were trained for the particle number of 1×10^6 , 1×10^7 , and 1×10^8 . For each model, the resampled EPID TD was used as the input, and the EPID TD of 1×10^9 was used as the output. During training, all models were trained for 200 epochs using the L1 loss function and Adam optimizer, with an initial learning rate of 0.0002. A cosine annealing scheduler with a 5-epoch warm-up phase was employed to adjust the learning rate during training. The batch size was set to 4, and both the training and validation patch sizes were fixed at 512×512 . The training loss consistently decreased and stabilized after approximately 150 epochs, indicating a good convergence behavior. The training was carried out on a GTX 3090 GPU (24GB memory), conducted in the Pytorch 1.7.0.

To improve model generalization and emphasize clinically relevant features, several preprocessing and data-augmentation strategies were applied. First, we extracted the central 50% region from both the input and target EPID TD images along each spatial dimension. This cropping focused the network on the clinically relevant high-dose area and helped suppress the low-dose background noise. The cropped images were then resized back to the original resolution of 512×512 using bilinear interpolation to maintain consistent input dimensions across all samples. To further expand the dataset and simulate variations in the patient setup and beam geometry, we applied geometric augmentations: 1) Horizontal flipping (left–right) using NumPy’s `np.fliplr()`. 2) Vertical flipping (top–bottom) using `np.flipud()`. These operations were independently applied to both the input and target TD images to maintain spatial alignment and ensure consistency during supervised learning.

2.3 Quantitative evaluation metrics

Structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are often used to evaluate image performance.

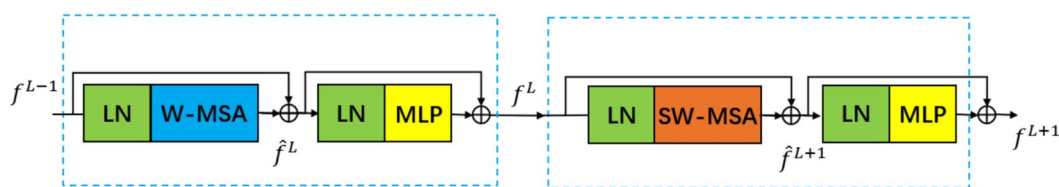


Fig. 3 The architecture of Swin Transformer Block. It includes: Layer Normalization (LN), Multi-Head Self Attention (MSA), Residual Connection, and a Multi-Layer Perceptron (MLP) with Gaussian Error Linear Unit (GELU) activation

The SSIM shows the structural similarity between two TD images on a scale of 0 to 1, with higher values indicating greater similarity. It is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the mean intensities, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance, and C_1 and C_2 are small constants to avoid division by zero.

The PSNR quantifies the fidelity of TD images based on the MSE. Higher PSNR values indicate better image quality. Generally, a PSNR above 30 dB suggests high fidelity, whereas values below 20 dB indicate significant distortion.

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right),$$

where MAX is the maximum pixel value. MSE is computed as:

$$MSE = \frac{1}{M \times N} \sum [I(i, j) - K(i, j)]^2,$$

where $I(i, j)$ and $K(i, j)$ are the pixel intensities of the original and denoised images, respectively. $1 \leq i \leq M, 1 \leq j \leq N$.

The gamma passing rate (GPR) quantitatively evaluates dosimetric and spatial agreements between two dose distributions, making it particularly suitable for detecting discrepancies. The GPR evaluates the agreement between the predicted and reference TD based on the gamma index $\Gamma(x, y)$:

$$\Gamma(x, y) = \min \sqrt{\left(\frac{\Delta D}{D_{crit}} \right)^2 + \left(\frac{\Delta d}{d_{crit}} \right)^2},$$

where ΔD and Δd are the dose and spatial differences, respectively, and D_{crit} and d_{crit} are the dose and distance criteria, respectively. The GPR represents the percentage of points satisfying $\Gamma(x, y) \leq 1$ which was evaluated using the 3%/2 mm criterion, with a 10% low-dose threshold based on

AAPM TG-307. For 2D relative mode transit gamma analysis, a passing rate greater than 95% with 3%/2 mm criteria is typically recommended, and for 2D absolute dose mode, the tolerance may be looser owing to the uncertainties of the dose conversion from the EPID image pixel data [14]. The SSIM and PSNR were computed using the skimage and sklearn libraries in Python, and the GPR was calculated using PTW software.

3 Results

To evaluate the impact of different particle numbers on the EPID TD, Fig. 4(a) displays five randomly selected MC-simulated TD images. Each column represents the results generated by the same particle number, whereas each row shows the results for varying particle numbers. From left to right: 1×10^6 , 1×10^7 , 1×10^8 , and 1×10^9 . The TD images were plotted using Python’s Matplotlib library with the “jet” colormap, where blue indicates low-dose regions and yellow to red indicates high-dose areas. It can be observed that TD images corresponding to 1×10^6 exhibit severe graininess, with highly blurred structural details, making it difficult to discern dose boundaries, which is noise caused by insufficient particle number. As the number of particles increased, the noise progressively diminished, revealing clearer internal structures. Notably, the TD images at 1×10^8 are already comparable to those at 1×10^9 . These results highlight a direct correlation between the particle number and the quality

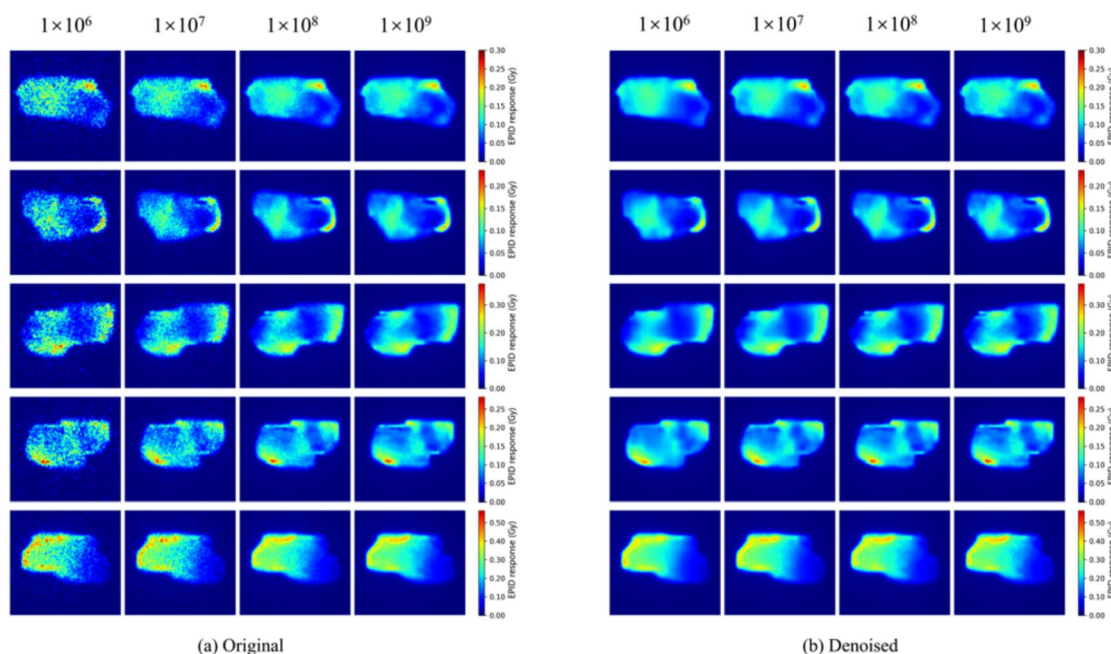


Fig. 4 (Color online) Comparison of EPID TD with different particle number and DL-based denoising. (a) original, (b) denoised

of the TD images, where a higher particle number leads to an improved signal-to-noise ratio (SNR), producing clearer and smoother TD images. Figure 4(b) displays the denoised TD images corresponding to the same five fields shown in Fig. 4(a). The TD images at 1×10^9 , regarded as the ground truth, remain unchanged as the reference. The layout of Fig. 4(b) is consistent with Fig. 4(a). Compared to the original TD images, those with 1×10^6 , 1×10^7 , and 1×10^8 exhibit significant improvements, with noticeable noise reduction and a smoother visual appearance. However, at the lowest particle number, 1×10^6 , some degree of distortion was observed. For instance, in comparison to 1×10^9 , high-dose regions appear overly smoothed, suggesting that under extremely low SNR conditions, DL-based denoising may compromise certain details of the TD images. Differences are also present at 1×10^7 , but they are relatively smaller and less significant.

Figure 5 presents the central horizontal profile curves of the EPID TD, illustrating the effect of DL-based denoising. The selected cases correspond to those shown in Fig. 4, with the first row showing the original profile curves and the second row showing the profile curves after denoising. The blue, orange, green, and red solid lines represent the particle

number of 1×10^6 , 1×10^7 , 1×10^8 , and 1×10^9 , respectively. The profile curve of 1×10^9 was not denoised in either row and is presented as a reference for comparison. Under a low particle number, the profile curves exhibited large fluctuations and a noticeable jagged pattern. After DL-based denoising, the profile curves became smoother, and the noise fluctuation amplitude decreased significantly. However, differences remained between low and high particle numbers. The profile agreement between 1×10^6 and 1×10^9 was relatively limited, with noticeable deviations across the field. In contrast, 1×10^7 showed some discrepancies, which were primarily localized to high-gradient regions, such as dose peaks and valleys. This aligns with the slight distortion observed in the denoised low-particle-number TD images, as shown in Fig. 4. To quantitatively evaluate the denoising performance, the SSIM and PSNR were calculated. Figure 6 illustrates the improvement in the EPID TD across different particle numbers after applying the DL-based denoising model. Table 1 and 2 present the SSIM, PSNR, and relative improvement ratio of the DL-based denoising model across 10 cases in the test set. The relative improvement ratio was calculated as follows:

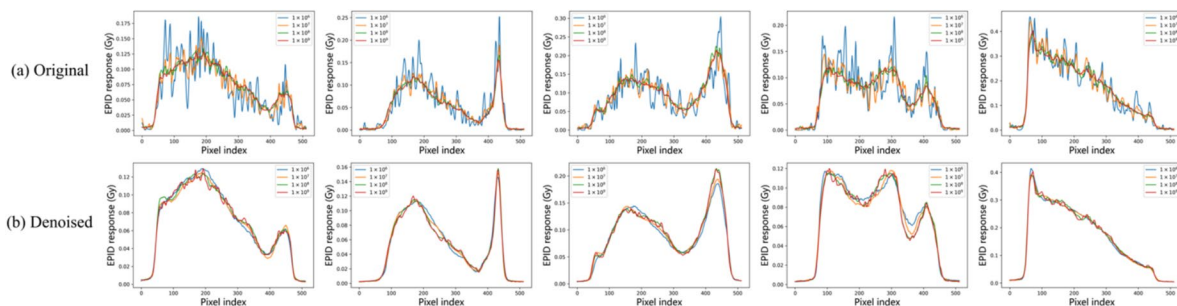


Fig. 5 (Color online) Comparison of profile curves with different particle number and DL-based denoising. (a) original, (b) denoised

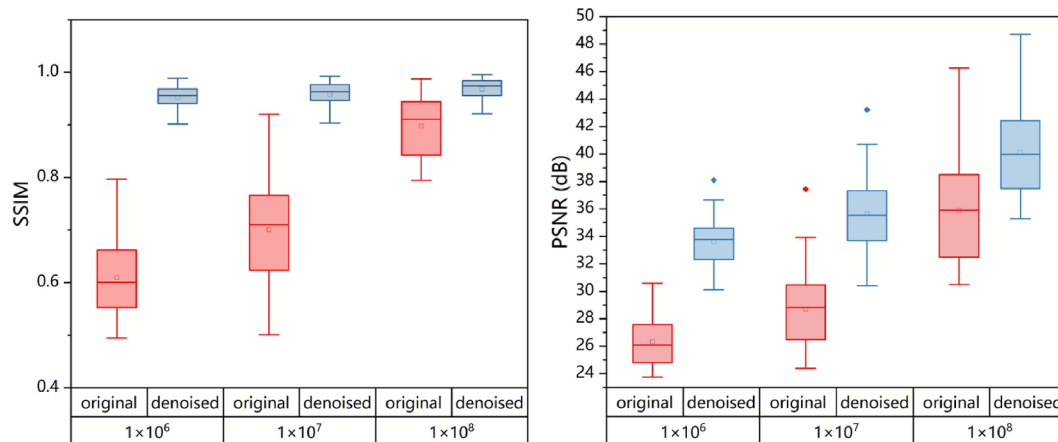


Fig. 6 The comparison of SSIM and PSNR before and after denoising. (a) SSIM, (b) PSNR

$$R = \frac{I_{\text{denoised}} - I_{\text{original}}}{I_{\text{original}}} \times 100\%$$

where R denotes the relative improvement ratio, and I_{original} and I_{denoised} represent the metric values before and after denoising, respectively.

In Fig. 6, the X-axis categorizes EPID TD by particle number: 1×10^6 , 1×10^7 , and 1×10^8 with each group divided into “original” (before denoising) and “denoised” (after denoising) subgroups. The Y-axis represents the SSIM and PSNR values compared to the reference 1×10^9 . For the original EPID TD, the 1×10^6 cases had the lowest SSIM of 0.61 and PSNR of 26.33, indicating the highest noise level and poorest accuracy. As the number of particles increased, the quality of the EPID TD improved, with 1×10^7 reaching an SSIM of 0.70 and PSNR of 28.70. However, this was still below 1×10^8 , which exhibited the highest SSIM of 0.90 and PSNR of 35.89. The boxplot for 1×10^7 shows the greatest variation, indicating a higher inconsistency and suggesting room for further improvement. After DL-based denoising, the SSIM and PSNR values increased significantly for all cases. The 1×10^6 cases showed the most substantial

improvement, with the SSIM rising to 0.95, even surpassing the original 1×10^8 cases. However, the PSNR remained slightly lower, indicating residual noise. The 1×10^7 and 1×10^8 cases also exhibit marked enhancement, with 1×10^8 achieving the highest quality, SSIM of 0.97, PSNR 40.11. Because the original 1×10^8 cases already have high quality, the relative improvement is less pronounced.

For SSIM, under low particle numbers, 1×10^6 and 1×10^7 , the improvement in SSIM is most notable, with the average improvement reaching up to 73.88% and 64.05% for individual cases, respectively. Although the original cases at 1×10^8 already have a very high SSIM, denoising still results in an average improvement of approximately 8.11%. After denoising, the overall average SSIM across all cases in the test set exceeded 0.95. The PSNR followed a trend similar to that of the SSIM. The DL-based denoising method effectively improved the PSNR under all conditions. For 1×10^6 and 1×10^7 , the PSNR improvement ratio was mostly above 20%. For 1×10^8 , the mean improvement ratio is still above 10%. However, it is noteworthy that while the PSNR significantly improves for 1×10^6 and 1×10^7 , their average values

Table 1 Mean SSIM for 10 cases in the test set

ID	1×10^6			1×10^7			1×10^8		
	Original	Denoised	Ratio	Original	Denoised	Ratio	Original	Denoised	Ratio
#1	0.65	0.95	45.71%	0.75	0.96	28.75%	0.93	0.98	5.37%
#2	0.74	0.98	33.26%	0.87	0.99	13.60%	0.98	0.99	1.55%
#3	0.57	0.95	68.12%	0.65	0.96	46.57%	0.88	0.97	9.42%
#4	0.56	0.94	68.72%	0.66	0.95	44.26%	0.85	0.95	11.77%
#5	0.61	0.96	57.51%	0.70	0.97	37.95%	0.91	0.97	6.94%
#6	0.67	0.97	44.79%	0.79	0.97	24.07%	0.94	0.98	4.17%
#7	0.58	0.92	60.60%	0.62	0.93	50.16%	0.84	0.94	11.75%
#8	0.56	0.95	69.97%	0.65	0.95	46.89%	0.86	0.96	11.89%
#9	0.63	0.97	52.81%	0.76	0.97	27.84%	0.95	0.98	3.97%
#10	0.54	0.93	73.88%	0.57	0.94	64.05%	0.83	0.95	14.23%
Mean	0.61	0.95	57.54%	0.70	0.96	38.41%	0.90	0.97	8.11%

Table 2 Mean PSNR for 10 cases in test set

ID	1×10^6			1×10^7			1×10^8		
	Original	Denoised	Ratio	Original	Denoised	Ratio	Original	Denoised	Ratio
#1	27.46	33.50	21.99%	29.82	35.27	18.29%	36.87	40.57	10.01%
#2	29.05	36.30	24.94%	34.08	40.82	19.76%	43.03	45.96	6.80%
#3	25.56	33.07	29.36%	27.38	34.96	27.70%	34.72	39.41	13.51%
#4	24.93	32.07	28.66%	26.94	33.62	24.80%	32.93	37.53	13.97%
#5	26.45	34.20	29.31%	28.32	36.09	27.47%	35.66	40.31	13.04%
#6	27.31	35.00	28.16%	30.81	37.64	22.19%	38.63	42.58	10.21%
#7	25.93	33.19	28.01%	27.04	33.89	25.35%	33.12	37.10	12.02%
#8	25.20	32.73	29.86%	27.25	34.42	26.29%	33.45	38.38	14.73%
#9	26.65	33.97	27.45%	29.99	37.01	23.38%	38.53	42.27	9.72%
#10	24.79	31.94	28.84%	25.34	32.70	29.03%	31.98	36.99	15.67%
Mean	26.33	33.60	27.66%	28.70	35.64	24.43%	35.89	40.11	11.97%

remain lower than those for 1×10^8 . These results suggest that the inherent information loss caused by lower simulated particle numbers may still limit the maximum achievable quality. Nevertheless, the improvements in the SSIM and PSNR strongly demonstrate that the DL-based denoising method can substantially enhance the quality of the EPID TD with a low particle number.

Table 3 presents the original, denoised GPR and relative improvement ratio across 10 cases in test set. The GPR strongly depends on the particle number, and a higher particle number typically reduces the statistical noise, which is evident in the original GPR values. As the number of particles increased from 1×10^6 to 1×10^7 , the mean GPR increased from 48.47% to 61.04%, and further to 91.88% at 1×10^8 . This confirms that low-particle-number simulations suffer from noise-induced degradation, which affects the dose verification reliability. After applying DL-based denoising, the GPR for all particle numbers improved dramatically. For 1×10^6 , the mean GPR increased from 48.47% to 89.10%, representing an 83.83% improvement. For 1×10^7 , the mean GPR increased from 61.04% to 94.35%, reflecting a 54.57% enhancement. However, for 1×10^8 , the increase is more modest, from 91.88% to 99.55%, with 8.34% gain. Notably, the denoised results for 1×10^7 exceed the original GPR for 1×10^8 , highlighting the model’s capability to recover accuracy even in simulations that are computationally less expensive. The ability of the DL-based denoising method to enhance the accuracy of low particle simulations beyond that of high particle simulations underscores its potential to reduce computational costs while maintaining high accuracy, offering a practical and efficient solution for PSQA.

To assess the computation efficiency, Table 4 summarizes the time of the MC simulation, DL-based denoising, and total computation across different particle numbers for one field. As expected, the time of the MC simulation increased

with the particle number, whereas the time of DL-based denoising remained relatively stable at approximately 0.1 s, with minor fluctuations depending on the system performance. Notably, the denoising time was significantly lower than that of the MC simulation and remained largely unaffected by the particle number. Overall, the total computation time was primarily driven by the MC simulation, increasing with the number of particles. However, incorporating DL-based denoising dramatically reduces the computation time while preserving the quality of the EPID TD. For 1×10^6 and 1×10^7 , the total computation time is reduced to 1.24 s and 1.88 s, respectively, representing a 40-fold to 60-fold increase in efficiency compared to the 73.89 s required for 1×10^9 . While improvements are also observed at 1×10^8 particles, the efficiency gain is comparatively modest, with an approximately 9-fold increase compared to 1×10^9 . It is also worth noting that the computation time for 1×10^6 and 1×10^7 remains under 2 s and shows little difference. However, as discussed earlier, the denoised TD images and profile curves for 1×10^6 exhibit distortions, and their GPR is also relatively low. Considering both quality and computational efficiency, 1×10^7 offers a reasonable trade-off, ensuring a high-quality EPID TD; if higher accuracy is required, the particle number can be increased to 1×10^8 .

Table 3 Mean GPR for 10 cases in test set

ID	1×10^6			1×10^7			1×10^8		
	Original	Denoised	Ratio	Original	Denoised	Ratio	Original	Denoised	Ratio
#1	51.14%	87.44%	70.98%	63.50%	91.60%	44.25%	94.00%	99.62%	5.98%
#2	50.82%	89.42%	75.95%	77.76%	98.32%	26.44%	98.96%	99.98%	1.03%
#3	46.08%	89.56%	94.36%	57.78%	94.88%	64.21%	90.78%	99.48%	9.58%
#4	46.18%	86.68%	87.70%	50.48%	92.40%	83.04%	86.70%	99.08%	14.28%
#5	45.20%	89.22%	97.39%	58.74%	94.34%	60.61%	90.90%	99.48%	9.44%
#6	49.54%	89.44%	80.54%	66.90%	96.16%	43.74%	96.26%	99.90%	3.78%
#7	56.46%	95.04%	68.33%	62.44%	96.20%	54.07%	91.60%	99.62%	8.76%
#8	44.34%	88.86%	100.41%	55.36%	93.36%	68.64%	87.40%	99.48%	13.82%
#9	47.80%	87.14%	82.30%	65.88%	95.88%	45.54%	96.66%	99.84%	3.29%
#10	47.12%	88.18%	87.14%	51.60%	90.40%	75.19%	85.54%	98.98%	15.71%
Mean	48.47%	89.10%	83.83%	61.04%	94.35%	54.57%	91.88%	99.55%	8.34%

Table 4 Computation time for different particle number and methods

Time	1×10^6	1×10^7	1×10^8	1×10^9
MC	1.12 s	1.72 s	8.62 s	73.89 s
Denoised	0.13 s	0.16 s	0.15 s	–
Total	1.24 s	1.88 s	8.76 s	73.89 s

4 Discussion

PSQA plays a crucial role in radiotherapy, ensuring treatment accuracy and safety. The forward-projection approach compares the measured and predicted EPID TD, whose clinical applicability depends largely on the suitability of the predicted data. MC, recognized as the “gold standard” for dose calculation, can be used to predict the EPID TD with high accuracy. The precise modeling of linear accelerators and EPID remains a substantial challenge. Zhang et al. attempted to use the PRIMO MC code to compute the EPID TD and opted to use a homogeneous water phantom as a substitute for the complex EPID [40]. Li et al. performed EPID dosimetry studies using DL models with water-equivalent materials [41]. Approximating the EPID model may introduce inaccuracies; therefore, finding an appropriate MC simulation tool is crucial. Although MC-based EPID TD prediction can provide accurate results, it relies on statistical sampling, requiring a sufficient number of particles to achieve an accurate simulation. Martins et al. spent 14 h using the Geant4 MC code to simulate accurate dose distributions and respective EPID signals for only one subfield of the treatment plan, which may not be a good choice for clinical applications [42]. Lazaro et al. utilized an MC code to develop an EPID model. When simulating a 1024×1024 image on a Linux cluster with 100 processors (2.26 GHz), the computation time was 30 min for a simulation involving over 100 million photons from the PSF. Conversely, increasing the photon number to 500 million extended the computation time to 2 h and 30 min [43]. Such computation times are impractical for clinical use, particularly in time-critical scenarios. This study employed ARCHER, a GPU-accelerated MC code, to predict the EPID TD, which incorporates detailed models of the linear accelerator and EPID, enabling accurate results. Although ARCHER offers faster simulations, it cannot meet all clinical conditions, such as online ART, where rapid calculations are required.

Based on lung IMRT cases, this study showed the impact of different particle numbers on the quality of the EPID TD in ARCHER. As the number of particles increased, the quality of the TD images improved progressively, from noticeable graininess under 1×10^6 to clear structural details at 1×10^9 . However, a high-particle-number demands extensive computational costs, reducing its feasibility for clinical applications. Applying a low particle number often results in significant noise, making the EPID TD clinically unacceptable. Therefore, it is crucial to find an effective denoising approach for EPID TD prediction. Traditional image processing methods, such as Gaussian filtering and non-local means filtering, can suppress noise to some extent but often sacrifice details and may even distort the dose distribution. This study proposes the use of DL to denoise the EPID TD

generated by MC simulation at low particle numbers, aiming to mitigate noise while ensuring quality and reducing computation costs. By introducing the SUNet neural network model, the model learns the characteristics of low-particle-number and high-particle-number TD images, enabling noise removal. The results indicate that the DL-based denoising method significantly enhanced the SSIM and PSNR. However, for 1×10^6 , the extremely low original SNR may still lead to detail loss and distortion after the denoising. A computational efficiency analysis showed that DL-based denoising remained stable at approximately 0.1 s, significantly reducing the total computation time compared with the MC simulation. At 1×10^7 , the total computation time is only 1.88 s, achieving a 40-fold speedup, and at 1×10^8 , the total computation time is only 8.76 s, achieving a 9-fold speedup compared to 1×10^9 . The timing results should be interpreted in the context of hardware used. These results were obtained using a single RTX-3090 GPU, which reflects the hardware setup available in many clinical environments. Although the use of multi-GPU or cloud-based platforms may further accelerate denoising, such resources are not universally available. Therefore, our study provides a practical and accessible solution that balances performance and broad applicability.

This approach, which integrates MC simulation with a DL-based denoising technique, offers substantial potential for clinical applications, particularly in online ART, in which treatment plans must be updated and verified quickly during each treatment session to account for changes in tumor shape and position. This process requires rapid processing and analysis to adjust treatment plans quickly and ensure precision and efficacy. EPID TD serves as a critical tool for dose verification in this process. By applying the trained DL-based model to low-particle-number EPID TD, we can effectively suppress noise and provide reliable support for PSQA, which can help in clinical quick radiotherapy decision-making. Additionally, this work not only serves as an independent PSQA tool based on EPID TD but also lays a foundation for subsequent 3D dose reconstruction on patient CT images. DVH-based dosimetric evaluations of targets and organs-at-risk rely on this. Despite these promising results, this study has several limitations. First, it only includes 100 lung IMRT cases, and the current dataset is limited to a single anatomical site. Models trained solely on lung data may not generalize well to other regions, such as the pelvis, owing to differences in anatomical complexity and beam configurations. Therefore, the applicability of the proposed method to other treatment sites requires further investigation. Future studies will focus on evaluating the model’s performance across different anatomical regions to further assess its generalizability. Second, although the SUNet model demonstrates significant improvements in denoising and effectively removes noise from low-particle-number simulations, such as 1×10^6 particles or fewer, remains a challenge. Further enhancements to the model

or integration of complementary techniques may be necessary to overcome this limitation. Finally, future research should focus on systematically comparing a range of state-of-the-art neural network architectures to identify models with superior denoising capabilities for EPID TD images. This includes exploring hybrid models, ensemble learning approaches, and transfer learning strategies that could potentially improve robustness and generalization. Concurrently, iterative refinements to the existing SUNet architecture will be pursued to optimize its performance and computational efficiency for practical clinical implementation.

5 Conclusion

This study introduces an integrated approach that combines MC simulation with DL-based denoising for EPID TD generation and provides a recommended particle number for the MC simulation. The EPID TD at different particle numbers (1×10^6 , 1×10^7 , 1×10^8 , and 1×10^9) was obtained using the ARCHER MC code for lung IMRT cases, which were used to train the neural network model SUNet. The trained model is applied to denoise TD images at low particle numbers, significantly enhancing the quality while reducing the computational cost. Considering both image quality and computational efficiency, 1×10^7 offers a reasonable trade-off, particularly for time-constrained scenarios in which expedited dose verification is critical. However, when higher dosimetric precision is required, 1×10^8 may be more appropriate because of its higher accuracy. This flexibility allows the method to be adapted to varying clinical demands, providing a practical and scalable solution for PSQA in online ART. Future work will focus on further optimizing the model performance, expanding its clinical applications, and integrating complementary computational techniques to enhance its clinical utility.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection and analysis were performed by Tao Qiu, Ning Gao, and Yan-Kui Chang. Xi Pei and Huan-Li Luo helped in developing the idea of the study. Fu Jin helped in guiding the direction of research and edited the manuscript. The first draft of the manuscript was written by Tao Qiu, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at <https://cstr.cn/31253.11.sciencedb.j00186.00943> and <https://doi.org/10.57760/sciencedb.j00186.00943>.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. G.A. Ezzell, J.M. Galvin, D. Low et al., Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee. *Med. Phys.* **30**, 2089–2115 (2003). <https://doi.org/10.1118/1.1591194>
2. M. Miften, A. Olch, D. Mihailidis et al., Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med. Phys.* **45**, e53–e83 (2018). <https://doi.org/10.1002/mp.12810>
3. S.B. Lim, P. Godoy Sripes, M. Napolitano et al., An investigation of using log-file analysis for automated patient-specific quality assurance in MRgRT. *J. Appl. Clin. Med. Phys.* **22**, 183–188 (2021). <https://doi.org/10.1002/acm2.13361>
4. I. Olaciregui-Ruiz, S. Beddar, P. Greer et al., In vivo dosimetry in external beam photon radiotherapy: Requirements and future directions for research, development, and clinical practice. *Phys. Imaging Radiat. Oncol.* **15**, 108–116 (2020). <https://doi.org/10.1016/j.phro.2020.08.003>
5. S. Li, H.L. Luo, X. Tan et al., The impact of plan complexity on calculation and measurement-based pre-treatment verifications for sliding-window intensity-modulated radiotherapy. *Phys. Imaging Radiat. Oncol.* **31**, 100622 (2024). <https://doi.org/10.1016/j.phro.2024.100622>
6. Z.W. Shen, X. Tan, S. Li et al., Correlation between the gamma passing rates of IMRT plans and the volumes of air cavities and bony structures in head and neck cancer. *Radiat. Oncol.* **16**, 134 (2021). <https://doi.org/10.1186/s13014-021-01861-y>
7. L.Y. Chen, H.L. Luo, S. Li et al., Pretreatment patient-specific quality assurance prediction based on 1D complexity metrics and 3D planning dose: classification, gamma passing rates, and DVH metrics. *Radiat. Oncol.* **18**, 192 (2023). <https://doi.org/10.1186/s13014-023-02376-4>
8. J.Q. Li, L. Wang, X.L. Zhang et al., Machine learning for patient-specific quality assurance of VMAT: Prediction and classification accuracy. *Int. J. Radiat. Oncol. Biol. Phys.* **105**, 893–902 (2019). <https://doi.org/10.1016/j.ijrobp.2019.07.049>
9. L.E. Antonuk, J. Boudry, W. Huang et al., Demonstration of megavoltage and diagnostic x-ray imaging with hydrogenated amorphous silicon arrays. *Med. Phys.* **19**, 1455–1466 (1992). <https://doi.org/10.1118/1.596802>
10. P.B. Greer, Correction of pixel sensitivity variation and off-axis response for amorphous silicon EPID dosimetry. *Med. Phys.* **32**, 3558–3568 (2005). <https://doi.org/10.1118/1.2128498>
11. M. Van Zijtveld, M. Dirx, M. Breuers et al., Portal dose image prediction for in vivo treatment verification completely based on EPID measurements. *Med. Phys.* **36**, 946–952 (2009). <https://doi.org/10.1118/1.3070545>
12. J. Chen, C.F. Chuang, O. Morin et al., Calibration of an amorphous-silicon flat panel portal imager for exit-beam dosimetry. *Med. Phys.* **33**, 584–594 (2006). <https://doi.org/10.1118/1.2168294>
13. M.G. Herman, J.M. Balter, D.A. Jaffray et al., Clinical use of electronic portal imaging: Report of AAPM Radiation Therapy Committee Task Group 58. *Med. Phys.* **28**, 712–737 (2001). <https://doi.org/10.1118/1.1368128>
14. N. Dogan, B.J. Mijneer, K. Padgett et al., AAPM task group report 307: Use of EPIDs for patient-specific IMRT and VMAT QA. *Med. Phys.* **50**, e865–e903 (2023). <https://doi.org/10.1002/mp.16536>
15. K. Chytk-Praznik, E. VanUytven, T.A. vanBeek et al., Model-based prediction of portal dose images during patient treatment. *Med. Phys.* **40**, 031713 (2013). <https://doi.org/10.1118/1.4792203>

16. K. Chytky, B.M. McCurdy, Comprehensive fluence model for absolute portal dose image prediction. *Med. Phys.* **36**, 1389–1398 (2009). <https://doi.org/10.1118/1.3083583>
17. J.V. Siebers, J.O. Kim, L. Ko et al., Monte Carlo computation of dosimetric amorphous silicon electronic portal images. *Med. Phys.* **31**, 2135–2146 (2004). <https://doi.org/10.1118/1.1764392>
18. O.L. Green, L.E. Henke, G.D. Hugo, Practical clinical workflows for online and offline adaptive radiation therapy. *Semin. Radiat. Oncol.* **29**, 219–227 (2019). <https://doi.org/10.1016/j.semradonc.2019.02.004>
19. D. Winkel, G.H. Bol, P.S. Kroon et al., Adaptive radiotherapy: The Elekta Unity MR-linac concept. *Clin. Transl. Radiat. Oncol.* **18**, 54–59 (2019). <https://doi.org/10.1016/j.ctro.2019.04.001>
20. O.M. Dona Lemus, M. Cao, B. Cai et al., Adaptive radiotherapy: Next-generation radiotherapy. *Cancers* **16**, 1206 (2024). <https://doi.org/10.3390/cancers16061206>
21. A. Buades, B. Coll, J.M. Morel, A non-local algorithm for image denoising. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA, **2**, 60–65 (2005). <https://doi.org/10.1109/CVPR.2005.38>
22. K. Dabov, A. Foi, V. Katkovnik et al., Image denoising with block-matching and 3D filtering. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 6064, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning. 2006-02-17, pp. 606414. <https://doi.org/10.1117/12.643267>
23. W.S. Dong, X. Li, L. Zhang et al., Sparsity-based image denoising via dictionary learning and structural clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA, 2011, pp.457–464. <https://doi.org/10.1109/CVPR.2011.5995478>
24. J.P. Zhou, Z. Peng, Y.C. Song et al., A method of using deep learning to predict three-dimensional dose distributions for intensity-modulated radiotherapy of rectal cancer. *J. Appl. Clin. Med. Phys.* **21**, 26–37 (2020). <https://doi.org/10.1002/acm2.12849>
25. Y.X. Xing, D. Nguyen, W.G. Lu et al., Technical note: A feasibility study on deep learning-based radiotherapy dose calculation. *Med. Phys.* **47**, 753–758 (2020). <https://doi.org/10.1002/mp.13953>
26. S.M. Zhang, H. Wang, S.Q. Tian et al., A slice classification model-facilitated 3D encoder-decoder network for segmenting organs at risk in head and neck cancer. *J. Radiat. Res.* **62**, 94–103 (2021). <https://doi.org/10.1093/jrr/rraa094>
27. X. Zhen, J.W. Chen, Z.C. Zhong et al., Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. *Phys. Med. Biol.* **62**, 8246–8263 (2017). <https://doi.org/10.1088/1361-6560/aa8d09>
28. Z. Peng, H.M. Shan, T.Y. Liu et al., Deep learning for accelerating Monte Carlo radiation transport simulation in intensity-modulated radiation therapy:physics. *arXiv:1910.07735* (2019). <https://doi.org/10.48550/arXiv.1910.07735>
29. S. Guan, A.A. Khan, S. Sikdar et al., Fully dense UNet for 2D sparse photoacoustic tomography artifact removal. *IEEE J. Biomed. Health Inform.* **24**, 568–576 (2020). <https://doi.org/10.1109/JBHI.2019.2912935>
30. H.M. Huang, L.F. Lin, R.F. Tong et al., UNet 3+: A full-scale connected UNet for medical image segmentation. In: 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020, pp. 1055–1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>
31. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh et al., UNet++: A nested U-Net architecture for medical image segmentation. In: Stoyanov D, et al., editors. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer, 2018, pp. 3–11. (Lecture Notes in Computer Science, vol. 11045) https://doi.org/10.1007/978-3-030-00889-5_1
32. H. Cao, Y.Y. Wang, J. Chen et al., Swin-Unet: unet-like pure transformer for medical image segmentation. In: Karlinsky L, Michaeli T, Nishino K, editors. *Computer Vision – ECCV 2022 Workshops*. Cham: Springer, 2023, pp. 205–218. (Lecture Notes in Computer Science, vol. 13803). <https://doi.org/10.1007/978-3-031-25066-8-9>
33. C.M. Fan, T.J. Liu, K.H. Liu, SUNet: Swin transformer UNet for image denoising. In: Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS). Austin, TX, USA: IEEE, 2022, pp. 2333–2337. <https://doi.org/10.1109/ISCAS48785.2022.9937486>
34. Z. Liu, Y.T. Lin, Y. Cao et al., Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021, pp. 9992–10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
35. D.P. Adam, T. Liu, P.F. Caracappa et al., New capabilities of the Monte Carlo dose engine ARCHER-RT: Clinical validation of the Varian TrueBeam machine for VMAT external beam radiotherapy. *Med. Phys.* **47**, 2537–2549 (2020). <https://doi.org/10.1002/mp.14143>
36. B. Cheng, Y. Xu, S.J. Li et al., Development and clinical application of a GPU-based Monte Carlo dose verification module and software for 1.5T MR-LINAC. *Med. Phys.* **50**, 3172–3183 (2023). <https://doi.org/10.1002/mp.16337>
37. S.J. Li, B. Cheng, Y.X. Wang et al., A GPU-based fast Monte Carlo code that supports proton transport in magnetic field for radiation therapy. *J. Appl. Clin. Med. Phys.* **25**, e14208 (2024). <https://doi.org/10.1002/acm2.14208>
38. Z. Peng, Y. Lu, Y. Xu et al., Development of a GPU-accelerated Monte Carlo dose calculation module for nuclear medicine, ARCHER-NM: Demonstration for a PET/CT imaging procedure. *Phys. Med. Biol.* **67**, 065002 (2022). <https://doi.org/10.1088/1361-6560/ac58dd>
39. N. Gao, B. Cheng, Z. Wang et al., Feasibility of reconstructing in-vivo patient 3D dose distributions from 2D EPID image data using convolutional neural networks. *Phys. Med. Biol.* **70**, 015018 (2025). <https://doi.org/10.1088/1361-6560/ada19b>
40. J. Zhang, Z.B. Cheng, Z.T. Fan et al., A feasibility study for in vivo treatment verification of IMRT using Monte Carlo dose calculation and deep learning-based modelling of EPID detector response. *Radiat. Oncol.* **17**, 31 (2022). <https://doi.org/10.1186/s13014-022-01999-3>
41. Y.B. Li, F. Xiao, B.S. Liu et al., Deep learning-based 3D in vivo dose reconstruction with an electronic portal imaging device for magnetic resonance-linear accelerators: A proof of concept study. *Phys. Med. Biol.* **66**, 235011 (2021). <https://doi.org/10.1088/1361-6560/ac3b66>
42. J.C. Martins, J. Maier, C. Gianoli et al., Towards real-time EPID-based 3D in vivo dosimetry for IMRT with Deep Neural Networks: A feasibility study. *Phys. Medica.* **114**, 103148 (2023). <https://doi.org/10.1016/j.ejmp.2023.103148>
43. D. Lazaro, E. Barat, C. Le Loirec et al., Denoising techniques combined to Monte Carlo simulations for the prediction of high-resolution portal images in radiotherapy treatment verification. *Phys. Med. Biol.* **58**, 3433–3459 (2013). <https://doi.org/10.1088/0031-9155/58/10/3433>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.