Radiation Measurements 193 (2026) 107662

Contents lists available at ScienceDirect

Radiation Measurements

journal homepage: www.elsevier.com/locate/radmeas

Capability of deep learning to predict recoil protons for neutron dosimetry with Fluorescent Nuclear Track Detectors☆,☆☆

Long-Yang Jan Thai [a,b,c,*], Stefan Schmidt [d,a,c,e,f], Alexandra Walter [a,c,g], Richard V. Häcker [a,b,c], Kristina Giske [a,b,c], José Vedelago [d,a,c]

[a] Department of Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[b] Department for Physics and Astronomy, University of Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany
[c] Heidelberg Institute for Radiation Oncology (HIRO), National Center for Radiation Research in Oncology (NCRO), 69120 Heidelberg, Germany
[d] Department of Radiation Oncology, Heidelberg University Hospital (UKHD), Im Neuenheimer Feld 400, 69120 Heidelberg, Germany
[e] Medical Faculty Heidelberg, Heidelberg University, Im Neuenheimer Feld 672, 69120 Heidelberg, Germany
[f] Heidelberg Ion Beam Therapy Center (HIT), Heidelberg University Hospital, Im Neuenheimer Feld 450, 69120 Heidelberg, Germany
[g] Scientific Computing Center, Karlsruhe Institute of Technology (KIT), Zirkel 2, 76131 Karlsruhe, Germany

ARTICLE INFO

Keywords:
Neutron dosimetry
FNTDs
Machine learning
Deep learning network
nnU-net

ABSTRACT

Fluorescent Nuclear Track Detectors (FNTDs) provide high spatial resolution, wide linear energy transfer coverage, and reusability, making them well-suited for high-energy neutron dosimetry. When neutrons traverse a polyethylene converter, recoil protons are generated, and their tracks are stored inside the FNTDs and visualised through optical readout. Traditional analysis of FNTD images relies on deterministic algorithms or machine learning methods with explicit feature definition, limiting their general extension. In contrast, deep learning networks can extract image features enabling generalisation across different neutron energy spectra and dose values. In this study, a deep learning network was trained on images of FNTDs irradiated at six mono-energetic neutron energies and tested on images of FNTDs exposed to a broad-spectrum ^{241}Am-Be neutron source. Using raw images of irradiated FNTDs as input, the network predicted the proton tracks which were later counted. For the ^{241}Am-Be test dataset, a dose-response curve of identified tracks over ambient dose equivalent was fitted, and the sensitivity in terms of $H^*(10)$ was extracted from the slope. When the fit was applied on the whole $H^*(10)$ range, from 0 mSv up to 100 mSv, the predicted sensitivity for ^{241}Am-Be was $S_{pred} = (2280 \pm 20)$ tracks mSv^{-1} cm^{-2}. The relative deviation of this predicted sensitivity from the reference sensitivity was 5.8 %. When the fit considered only the $H^*(10)$ range of the training dataset, namely from 5 mSv to 15 mSv, the predicted sensitivity for ^{241}Am-Be was $S_{pred} = (2500 \pm 60)$ tracks mSv^{-1} cm^{-2}. This led to a relative deviation from the reference sensitivity of only 1.2 %. Despite being trained solely on mono-energetic data, the model successfully generalised to the ^{241}Am-Be energy spectrum.

1. Introduction

Ion beam radiotherapy, an advanced cancer treatment modality known for its optimal tissue-sparing capabilities, comes along with the production of secondary neutrons up to several hundred MeV, which can travel far beyond the primary beam path (Howell and Burgett, 2014; Vedelago et al., 2024). Accurate secondary neutron dose measurement is essential for improving late-effect probability estimations. Particularly, the broad energy distribution and biological effectiveness present a major challenge for accurate neutron dosimetry (ICRP 103, 2007; Baiocco et al., 2016).

A variety of neutron detectors had been used in the past (Gómez-Ros et al., 2023; Vedelago et al., 2024). Passive solid-state nuclear track detectors, such as Plastic Nuclear Track Detectors (PNTDs) and Fluorescent Nuclear Track Detectors (FNTDs), attracted considerable interest

☆ This article is part of a Special issue entitled: 'SSD21' published in Radiation Measurements.

☆☆ Given his role as Co-Guest Editor, Dr. José Vedelago had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to another journal editor.

* Corresponding author at: Department of Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
E-mail address: long-yang.thai@dkfz.de (L.-Y.J. Thai).
URL: https://www.dkfz.de/en/translational-research-for-ion-beam-therapy (J. Vedelago).

https://doi.org/10.1016/j.radmeas.2026.107662
Received 29 October 2025; Received in revised form 3 February 2026; Accepted 26 February 2026
Available online 26 February 2026

due to their capability to visualise charged particle tracks caused by secondary charged particles induced by neutron reactions (Akselrod and Kouwenberg, 2018; Bolzonella et al., 2022). Unlike PNTDs that require chemical etching, FNTDs offer direct optical readout, high spatial resolution and usage for various radiation types, energy and dose ranges (Akselrod et al., 2020; Dahham et al., 2025; Schmidt et al., 2026). Moreover, their linear energy transfer threshold of 0.4 keV/μm in water allows proton measurements up to approximately 230 MeV (Akselrod et al., 2006).

To date, the FNTD image analysis relies mostly on classical image processing techniques or simple machine learning methods. Principal component analysis was applied on Poly-Allyl Diglycol Carbonate (PADC) detectors and FNTDs to filter proton tracks from delta electron tracks (Stabilini et al., 2021). Tracks along with their angular distribution in FNTDs were identified with a support vector classifier, a classical machine learning algorithm (Nikaido et al., 2024). The potential in the application of deep learning networks in order to detect recoil protons within FNTD images remains largely unexplored; only a few publications deal with deep learning networks. In Akselrod et al. (2020), a convolutional neural network was used for track classification of FNTD images, where both training and testing were performed with ^{252}Cf and ^{241}Am-Be. Usually, these classical approaches rely on pre-defined thresholds for post-processing that are case-specific, for example, only useful for an ^{241}Am-Be irradiation, and it is not possible to generalise them to other neutron sources. These limitations can be overcome by more general approaches like deep learning networks.

The field of biomedical imaging has witnessed a paradigm shift, where traditional cell and tissue segmentation algorithms were outperformed by deep learning networks (Morelli et al., 2021; Rayed et al., 2024; Xu et al., 2024). These networks learn features from experimental data, even hidden properties, which were not uncovered by classical deterministic approaches. Afterwards, the trained model can be used to infer on related datasets. In particular, the nnU-Net, with the ability to adapt its architecture and parameters to different datasets, has become a state-of-the-art framework (Isensee et al., 2021). As a self-configuring deep learning framework, it is the first choice for biomedical image segmentation and demonstrated high performance for various datasets (Guo et al., 2024; Isensee et al., 2024; Joshi et al., 2025). Its prediction potential can be used for neutron dosimetry, particularly considering the limited experimental (quasi-)mono-energetic neutron facilities available worldwide, to cover the energy gaps for detector calibration and dosimetry (Nolte et al., 2004; Zakalek et al., 2025).

The goal of this study was to train a deep learning network using images that were obtained by irradiating FNTDs with mono-energetic neutron fields between 1.2 MeV and 19.0 MeV to segment proton track regions on images. The trained network was then used to segment the recoil proton tracks on images that were obtained by irradiating FNTDs with a broad-spectrum ^{241}Am-Be neutron source. The number of proton tracks was then counted. Finally, a dose-response curve, showing the identified tracks over ambient dose equivalent $H^*(10)$, was established for ^{241}Am-Be. From the slope of the dose-response curve, the sensitivity can be extracted and compared with a reference sensitivity.

2. Materials and methods

2.1. Irradiated detectors: training and test datasets

The detectors used for the experiments were FNTDs. These Al_2O_3:C,Mg crystals were manufactured by Landauer Inc-Crystal Growth Division (Stillwater, OK, USA) and have the size of 8 mm × 4 mm × 0.5 mm (Akselrod et al., 2006; Akselrod and Kouwenberg, 2018). A 1 mm-thick polyethylene (PE) converter sheet was placed before the FNTD. Neutrons interacting with this sheet can produce recoil protons through elastic scattering. These recoil protons induced luminescent tracks inside the FNTD that were used to perform neutron dosimetry. A dedicated confocal laser scanning microscope (FXR700RG) was used for the readout of the irradiated FNTDs (Akselrod et al., 2014). For each FNTD, 100 x-y-fields were read out at a fixed depth of z = 2 μm, with each field measuring a size of 100 μm × 100 μm. Each field was automatically cropped by the software to approximately 98.44 μm × 98.44 μm and then stored as a 16-bit greyscale fluorescent image in png format, representing the raw image input for the subsequent image analysis.

For the training, FNTDs irradiated with mono-energetic neutrons at the Physikalisch-Technische Bundesanstalt (PTB, Braunschweig, Germany) were used. At PTB, six different mono-energetic neutron fields ranging from 1.2 MeV to 19.0 MeV were used (Schmidt et al., 2025a), delivering an ambient dose equivalent $H^*(10)$ from 2.72 mSv to 15.3 mSv, depending on the energy (Supplementary Table S1).

The test data originates from FNTDs irradiated with a broad-spectrum ^{241}Am-Be source at the Paul Scherrer Institut (PSI, Villingen, Switzerland), as described in a previous study (Schmidt et al., 2025a). $H^*(10)$ values range from 0 mSv up to 100 mSv with a corresponding expanded uncertainty (k = 2) of approximately 8.3 %. Table 1 and Fig. 1 visualise the details of both datasets. There were six subsets of mono-energetic irradiations with different energies and seven subsets of Am-Be irradiations with different $H^*(10)$ values, including background (0 mSv). For every FNTD 100 readout fields or images were acquired. Every subset included three FNTDs, totalling in 300 images per subset. This led to a total dataset size of 1800 images for training and 2100 images for testing. The training and test datasets, including the raw images and the corresponding reference label masks (Section 2.2.1), can be found in the corresponding data repository (Schmidt et al., 2025b).

Table 1
Overview of irradiation setups for mono-energetic neutron irradiations at PTB (Training) and broad-energy spectrum from an ^{241}Am-Be irradiation at PSI (Test).

Dataset	Neutron energy / MeV	$H^*(10)$ / mSv
Training	1.2; 2.5; 5.0; 6.5; 14.8; 19.0	From 2.72 to 15.3
Test	Mean 4.05; maximum 11.0	0; 1; 5; 10; 15; 50; 100

2.2. Deep learning with the nnU-Net

2.2.1. Reference and prediction

Fig. 2 visualises the pipeline for obtaining the reference and prediction data within this study. As a starting point, the 16-bit greyscale fluorescent raw images X acquired with the FNTD reader were used as input for a greyscale-based segmentation script (Schmidt et al., 2025a). This deterministic algorithm made use of different image corrections and spot filtering procedures based on intensity and morphology features combined with Principal Component Analysis. It converted two-dimensional (2D) raw images X into 2D binary label masks Y. In these binary label masks Y, individual pixels were classified as either label "1" corresponding to a recoil proton candidate or label "0" corresponding to background. These binary label masks were used as reference when comparing to the network output.

A supervised learning approach was employed to train the deep learning network for semantic segmentation, using pairs of fluorescent raw images X and reference label masks Y as input. The network outputs were prediction label masks $\hat{Y}$ which were also 2D binary label masks with pixels classified as before, namely with "1" assigned to a recoil proton candidate and "0" to the background. The performance of binary segmentation was then compared between the prediction label masks $\hat{Y}$ and their corresponding reference label masks Y.

To compare the dosimetry performance between prediction and reference with dosimetric quantities, a binary-based track counting script was applied. This binary-based track counting script considered similar morphological algorithms as the one presented before (Schmidt et al., 2025a), excluding the greyscale-based features. It was applied

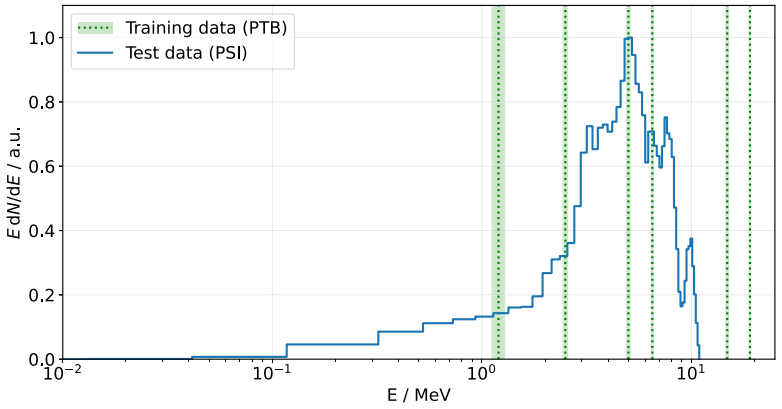

Fig. 1. Neutron spectra of training and test datasets in normalised lethargy representation. Mono-energetic neutron fields as training data and ^{241}Am-Be neutron spectrum as test data.

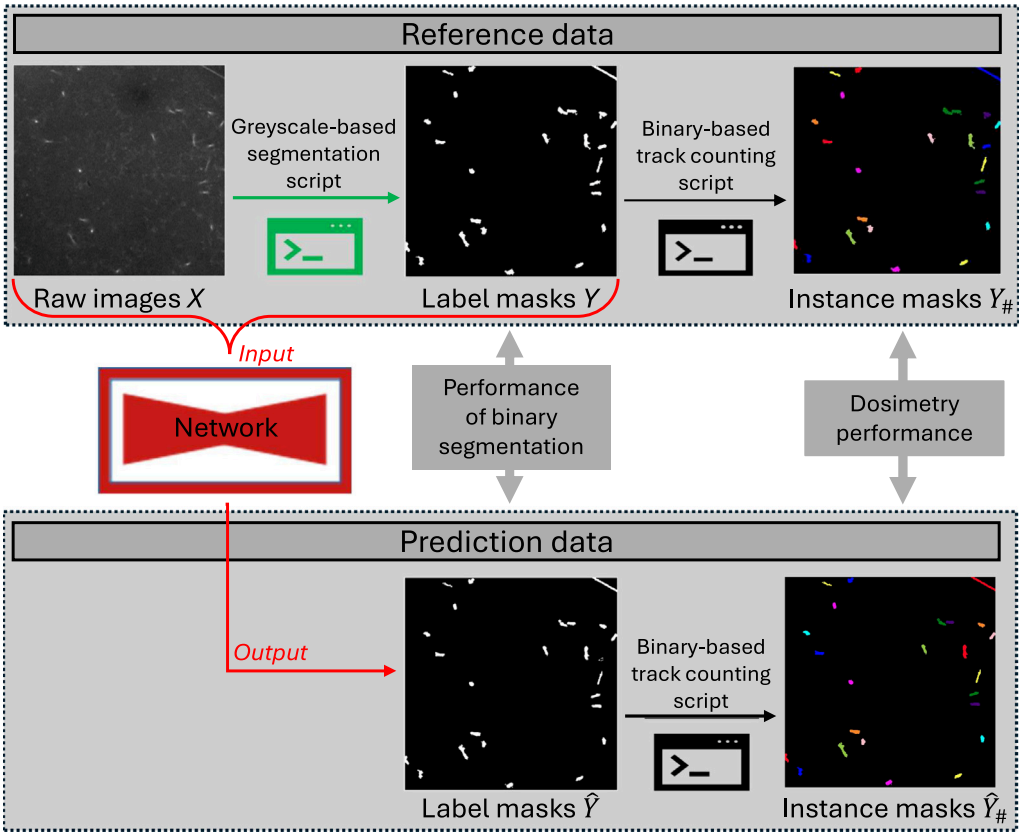

Fig. 2. Schematic data pipeline for obtaining the reference label masks Y and prediction label masks $\hat{Y}$, as well as the reference instance masks $Y_{\#}$ and prediction instance masks $\hat{Y}_{\#}$. The network was trained with raw images X, and reference label masks Y obtained with a greyscale-based segmentation script. The predicted label masks $\hat{Y}$ were compared to the reference label masks Y to assess the performance of binary segmentation. The dosimetry performance of the network was assessed by comparing derived quantities from reference instance masks $Y_{\#}$ and predicted instance masks $\hat{Y}_{\#}$, both obtained by using the same binary-based track counting script.

to both the reference label masks Y and the prediction label masks $\hat{Y}$ to guarantee the identical post-processing methodology. Using the label masks Y and $\hat{Y}$ as input, the proton track candidates were differentiated into individual proton tracks and background. The resulting outputs were instance masks $Y_{\#}$ and $\hat{Y}_{\#}$. In these non-binary 2D masks each pixel was classified to either background with the label "0" or an individual proton track of unique integer label ("1", "2", ...). Every individual track was painted with a different colour in Fig. 2. All identified tracks were counted and the track density was compared between reference $Y_{\#}$ and prediction $\hat{Y}_{\#}$ to evaluate the dosimetry performance of the network predictions.

2.2.2. Training and model selection

In this study, the nnU-Net (Version 2.6.1) was used with the recommended minimal settings (Isensee et al., 2021), along with the 2D configuration and the Residual Encoder Preset 'ResEncL' of 24 GB VRAM (Isensee et al., 2024). The recommended minimal settings to

guarantee the state-of-the-art performance include: (i) a five-fold cross-validation (CV) training scheme, where the five folds were ensembled into one CV model; (ii) 1000 epochs with 250 training iterations per epoch; (iii) a stochastic gradient descent optimiser with an initial learning rate of 0.01 together with the 'polyLR' schedule, and a Nesterov's momentum of 0.99; (iv) the mean of Dice and cross-entropy as loss function. Based on the dataset "fingerprint", rule-based parameters recommended by the nnU-Net during pre-processing were used directly, including a patch size of 512 × 512, a batch size of 35, and z-score intensity normalisation. Empirical parameters, such as post-processing, were adjusted as detailed in Section 2.3.

The nnU-Net was trained on the bwUniCluster 3.0.[1] For this purpose, a container was built with Singularity 4.1.1 and Python 3.12, including PyTorch with CUDA 11.8 support, to integrate the nnU-Net into the cluster. All five folds of the nnU-Net were trained on the bwUniCluster 3.0. Depending on the availability of the nodes, a NVIDIA H100 GPU or a NVIDIA H100 Ice Lake GPU was used. The average RAM usage per fold ranged from 19 GB to 22 GB, resulting in a total training time of approximately 17 h to 20 h. All the used files, including the container, the cluster-relevant codes, the workflow to get from the raw data input to the final output, and the Python scripts for evaluation, can be found in the corresponding software repository (Thai and Vedelago, 2025).

2.3. Performance metrics of the deep learning predictions

The performance of the nnU-Net-based model was quantified by binary segmentation and dosimetry metrics. For the performance of binary segmentation, the binary reference label masks Y were compared with their corresponding binary prediction label masks $\hat{Y}$.

Each of the six mono-energetic and seven Am-Be subsets comprised of $N = 300$ images. Let i denote the ith image out of N images. All raw images and corresponding label and instance masks have $M = 504^2$ pixels. Let j denote the jth pixel out of M pixels.

2.3.1. Binary segmentation metrics

The jth pixel of the ith image is considered as a true positive (TP) if the reference label mask indicated a proton ($Y_{i,j} = 1$) and the network correctly predicted this proton ($\hat{Y}_{i,j} = 1$). The total number of TP in the ith image is then defined as:

$$TP_i = |\{j \mid Y_{i,j} = 1 \ \wedge \ \hat{Y}_{i,j} = 1\}|, \tag{1}$$

where $|\ .\ |$ is the cardinality. Similarly, true negative (TN), false positive (FP) and false negative (FN) are defined for the ith image as:

$$TN_i = |\{j \mid Y_{i,j} = 0 \ \wedge \ \hat{Y}_{i,j} = 0\}|, \tag{2}$$

$$FP_i = |\{j \mid Y_{i,j} = 1 \ \wedge \ \hat{Y}_{i,j} = 0\}|, \tag{3}$$

$$FN_i = |\{j \mid Y_{i,j} = 0 \ \wedge \ \hat{Y}_{i,j} = 1\}|. \tag{4}$$

$n_{ref,i}$ is defined as the number of proton pixels in the ith reference label mask, and $n_{pred,i}$ is defined as the number of proton pixels in the ith prediction label mask.

From the previous quantities, traditionally used quantities such as Dice, Intersection over Union (IoU), and ratio can be derived (see Supplementary Section S2). However, since the nnU-Net framework was originally developed for biomedical datasets, these quantities were adapted for the prediction of recoil protons. Therefore, for the special case of a correct classified background image without any proton pixels ($TN_i = M \Leftrightarrow TP_i = FP_i = FN_i = 0 \Leftrightarrow n_{ref,i} = n_{pred,i} = 0$), the original definitions, also used by the nnU-Net, would result in NaN. Thus, this background image would have been ignored in the calculation of the mean and standard error of the mean (SEM), although the classification was 100 % right. Therefore, the definitions of the adapted Dice (D_{ad}), adapted Intersection over Union (IoU_{ad}) and adapted ratio (r_{ad}) were introduced to account for these correctly classified background images.

The adapted Dice for the ith image is defined as:

$$D_{ad,i} = \begin{cases} 1, & \text{if } TP_i = FP_i = FN_i = 0, \\ \frac{2TP_i}{2TP_i + FP_i + FN_i}, & \text{else.} \end{cases} \tag{5}$$

The adapted IoU for the ith image is defined as:

$$IoU_{ad,i} = \begin{cases} 1, & \text{if } TP_i = FP_i = FN_i = 0, \\ \frac{TP_i}{TP_i + FP_i + FN_i}, & \text{else.} \end{cases} \tag{6}$$

The ratio is the predicted proton pixels $n_{pred,i}$ divided by the reference proton pixels $n_{ref,i}$. If there are no recoil proton tracks in the reference ($n_{ref,i} = 0$), which is the case in some images of the training and test data, the ratio is undefined, since it would mean dividing $n_{pred,i}$ by 0. To amend this and extend the definition of ratio, three cases are considered for defining the adapted ratio. First, when $n_{ref,i} = 0$ and $n_{pred,i} = 0$, it means the network properly predicted a background image, and therefore the ratio is set to 1. Second, when $n_{ref,i} = 0$ but $n_{pred,i} > 0$, the ratio is penalised by comparing $n_{pred,i}$ with the mean number of proton pixels over all images of a certain subset $\bar{n}_{pred}$. If $n_{pred,i} \ll \bar{n}_{pred}$, it actually means that the network is not that far from the reference, thus assigning to the ratio a value approaching 1 when $n_{pred,i}$ over $\bar{n}_{pred}$ approaches 0. This is implemented by a linear interpolation between 1 and 0. Instead, if $n_{pred,i} \geq \bar{n}_{pred}$, it is severely penalised by setting the ratio to 0, which means a failure of the network prediction. The third case considers all the scenarios when $n_{ref,i} \neq 0$, with the ratio defined as $n_{pred,i}$ divided by $n_{ref,i}$. These three cases lead to the following definition of the adapted ratio for the ith image:

$$r_{ad,i} = \begin{cases} 1, & \text{if } n_{ref,i} = 0 \wedge n_{pred,i} = 0, \\ \max(1 - \frac{n_{pred,i}}{\bar{n}_{pred}}, 0), & \text{if } n_{ref,i} = 0 \wedge n_{pred,i} > 0, \\ \frac{n_{pred,i}}{n_{ref,i}}, & \text{else.} \end{cases} \tag{7}$$

The functional behaviour of the first two cases ($n_{ref,i} = 0$) of Eq. (7) is illustrated in Supplementary Figure S1.

The binary segmentation metrics were calculated for the training dataset as well as for the test dataset. For validation of the training, the mean and SEM were calculated based on the fold-specific validation sub-dataset for the individual folds, while the whole training dataset was used for the CV model. For the test dataset, the mean and SEM were calculated for each $H^*(10)$ subset.

Besides metrics evaluating the performance of binary segmentations, metrics assessing the dosimetry performance of the deep learning predictions can be determined. Here, quantities extracted from the reference instance masks $Y_\#$ and the prediction instance masks $\hat{Y}_\#$ were compared.

2.3.2. Dosimetry metrics

Let t_i be the number of proton tracks in the ith reference or prediction instance mask. By averaging over all N images of a subset, $\bar{t}$ represents the mean number of tracks per image. The track density $\rho = \bar{t}/A$ can be calculated by normalisation of $\bar{t}$ with the area A of an image. With the used acquisition settings, A is approximately 9.7×10^{-5} cm^2. The mean and SEM of the track density were calculated for each $H^*(10)$ subset. The background (0 mSv) was then subtracted from each of the six $H^*(10)$ subsets.

Fitting ρ (in tracks cm^{-2}) as a function of $H^*(10)$ (in mSv) gives the dose-response curve. A weighted linear regression with a fit going through the origin of the form $\hat{\rho} = S \cdot H^*(10)$ was applied with the weights $w_i = 1/\sigma_{rel,i}^2$ and $\sigma_{rel,i} = \Delta\rho_i/\rho_i$. The slope S (in tracks mSv^{-1} cm^{-2}) can be extracted from this fit and represents the sensitivity of the FNTDs.

[1] BwUniCluster 3.0. https://wiki.bwhpc.de/e/BwUniCluster3.0 (accessed 26 September 2025).

3. Results

3.1. Training and validation of the deep learning network

The training and validation loss of the 1. fold of the nnU-Net is shown in Fig. 3(a). The losses of the other four folds can be seen in Supplementary Figure S2. All folds showed the same trend, continuously decreasing with increasing number of epochs, supporting a non-overfitting, converged model.

The performance of the binary segmentation by the nnU-Net quantified with the modified metrics can be seen in Fig. 3(b) for the five individual folds, as well as for the CV model. D_{ad} ranges from 0.9443 to 0.9553 for the five different folds. For the CV model, D_{ad} is 0.9505 ± 0.0031. IoU_{ad} covers a range between 0.9165 to 0.9299 for the individual folds and results in 0.9239 ± 0.0035 for the CV model. r_{ad} goes from 0.992 to 1.105 for the individual folds, while for the CV model, r_{ad} gives 1.000 ± 0.004.

3.2. Performance of binary segmentation on the test dataset

The CV model trained on images of FNTDs irradiated with mono-energetic neutrons was used for inference on the ^{241}Am-Be test images. Two examples of the test subset corresponding to $H^*(10) = 100$ mSv are visualised in Fig. 4, depicting a raw image, its corresponding reference label mask, and the network predicted label mask. By comparing the reference in (b) with the prediction of the network in (c) for the first example, and the reference in (e) with the prediction in (f) for the second example, the predicted label masks closely resemble the reference label masks. The predicted proton spots appear slightly smaller in terms of area, but most of them match the reference proton spots.

The performance of the nnU-Net-based model on the test dataset quantified with the modified metrics is shown in Fig. 5 for all ^{241}Am-Be $H^*(10)$ subsets. For the 0 mSv subset, all three metrics show a value of at least 0.985. For the 1 mSv subset, D_{ad} and IoU_{ad} are at least 0.951, while r_{ad} is 0.976. For the subsets within the training dataset range, D_{ad} ranges from 0.935 to 0.956, IoU_{ad} from 0.898 to

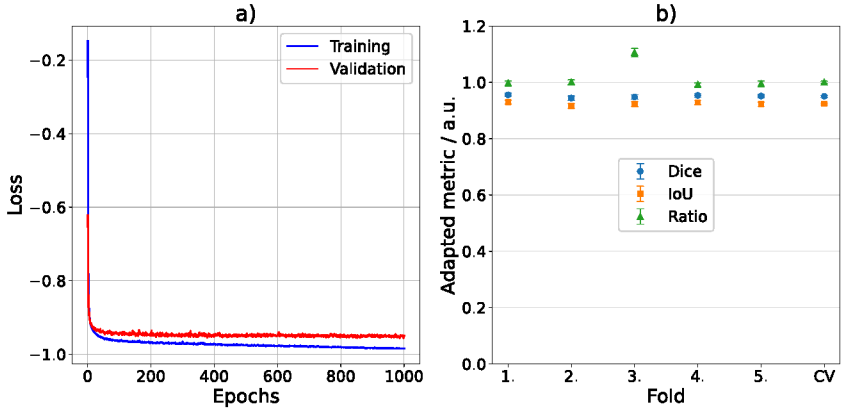

Fig. 3. Training and validation performance of the network. (a) Performance quantified by the training and validation loss of the 1. fold for 1000 epochs of training. (b) Performance metrics of binary segmentation for the validation of the five individual folds and the CV model.

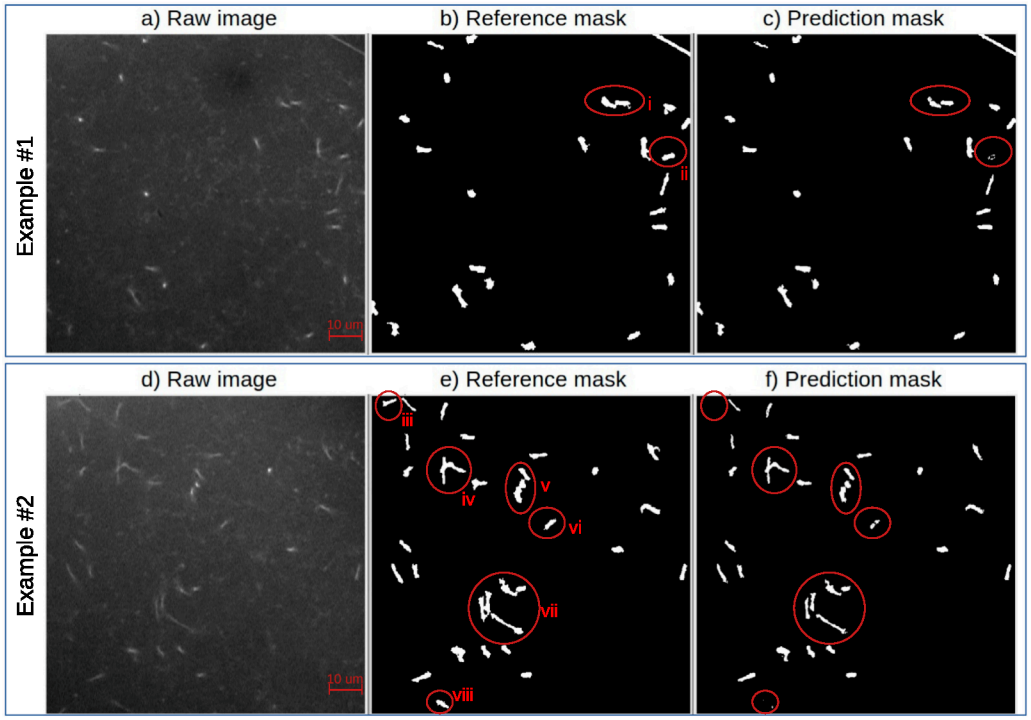

Fig. 4. Illustration of two examples, including the raw images (a, d), the reference label masks (b, e), and the network predicted label masks (c, f).

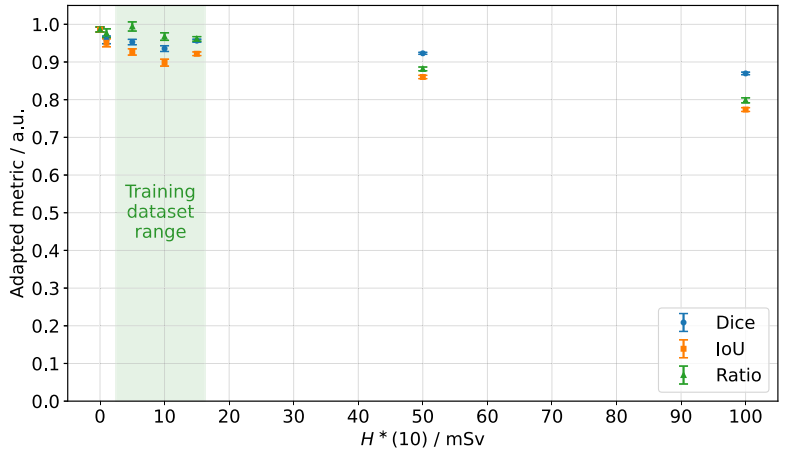

Fig. 5. Performance metrics of the network predictions for the broad-spectrum ^{241}Am-Be dataset when trained with the mono-energetic dataset.

0.926 and r_{ad} from 0.961 to 0.994. All three metrics show decreasing values towards high $H^*(10)$. While D_{ad} reduced to 0.923 and 0.870, IoU_{ad} shrank to 0.860 and 0.774, respectively, for the subsets 50 mSv and 100 mSv. This reflects the larger becoming mismatch of predicted and reference proton pixels towards high $H^*(10)$ values. In addition, r_{ad} decreased to 0.882 and 0.798, respectively, for the subsets 50 mSv and 100 mSv. The mean metrics over the whole ^{241}Am-Be test dataset are: $\overline{D}_{ad} = 0.9404 \pm 0.0053$, $\overline{IoU}_{ad} = 0.9025 \pm 0.0099$ and $\overline{r}_{ad} = 0.938 \pm 0.010$.

3.3. Dosimetry for a broad-spectrum irradiation with the mono-energetic trained network

The track counting on the nnU-Net predicted binary label masks leads to the examples shown in Fig. 6. The top row shows the first example, while the bottom row illustrates the second example, both corresponding to the ^{241}Am-Be 100 mSv subset. The raw images can be seen in (a, d) for the two examples. The counted proton tracks are shown in the corresponding predicted instance masks in (b, e). An overlay image of the raw image and its corresponding instance mask is shown in (c, f) for each example. These two examples illustrate the capabilities of the track counting from the network prediction.

For the first example shown in Fig. 6(b), the two closely lying tracks in (i) were classified as two separate tracks, although in the binary reference mask, those two spots were connected (Fig. 4b). The second example shown in Fig. 6(e) is characterised by multiple regions where tracks are located in proximity. For instance, in (vii), multiple tracks lying close to each other were classified as five separate tracks, while in the binary reference label mask, the upper two spots as well as the outer-left two spots were connected (Fig. 4e). As can be seen in Fig. 6(b) for (ii) and Fig. 6(e) for (iii) and (viii), these tracks were not counted, although the binary reference label masks (Fig. 4b and e) showed spots.

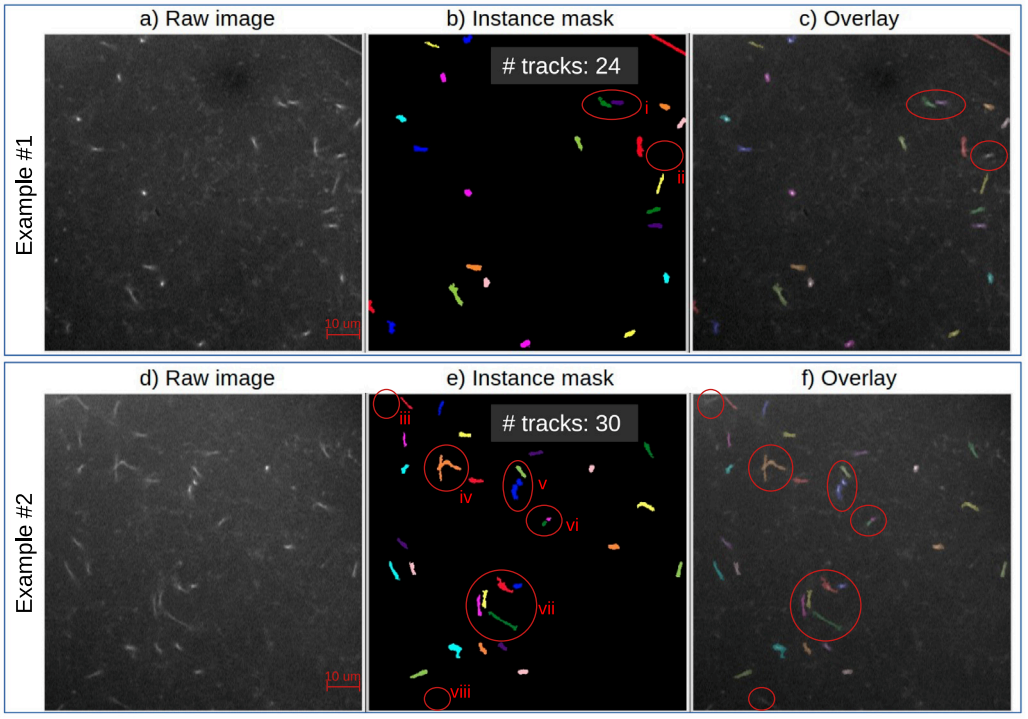

Fig. 6. Illustration of two examples for 100 mSv, including the raw images (a, d), the instance masks (b, e), and the overlay images (c, f).

The dose-response function of the network predicted tracks is depicted in Fig. 7, along with the reference. The linear fits yielded a sensitivity for the reference of S_{ref} = (2420 ± 20) tracks mSv^{-1} cm^{-2} with R^2_{ref} = 0.9998, and a sensitivity for the prediction of S_{pred} = (2280 ± 20) tracks mSv^{-1} cm^{-2} with R^2_{pred} = 0.9995. The sensitivities are remarkably very similar, presenting a deviation of 5.8 % for the mean values.

When the linear fits are confined to the ^{241}Am-Be subsets within the $H^*(10)$ range of the training dataset, the fits yielded a sensitivity for the reference of S_{ref} = (2530 ± 60) tracks mSv^{-1} cm^{-2} with R^2_{ref} = 0.9989, and a sensitivity for the prediction of S_{pred} = (2500 ± 60) tracks mSv^{-1} cm^{-2} with R^2_{pred} = 0.9990. In this case, the difference between the mean values of the sensitivities is reduced to 1.2 %, further supporting the agreement of the network prediction with the reference.

4. Discussion

4.1. Network training and validation

The loss curves of all five folds follow the same trend, indicating a reasonable dataset size and functioning random subset selection. From Fig. 3(a) and Supplementary Figure S2, it can be seen that the curves do not indicate overfitting and stagnating losses seem to appear already after 100 epochs. From the loss functions only, it is not possible to conclude that a model trained on fewer epochs will also perform well regarding other metrics (Isensee et al., 2021).

Regarding the validation performance of the CV model in Fig. 3(b), D_{ad} shows a value over 0.95, indicating a large overlap between prediction and reference. Also the IoU_{ad}, a more strict quantity for penalising under- and over-segmentation compared to D_{ad}, yields a value over 0.92, additionally supporting highly matching predictions. The mean value of 1.000 for r_{ad} concludes that the amount of proton pixels in the predictions matches the references, independent of the location in the reference and prediction. It can be concluded that this trained CV model predicts with high performance over the whole training dataset.

4.2. Performance of binary segmentation

In general, the predicted recoil proton label masks matched the reference label masks quite accurately for all $H^*(10)$ subsets up to the training dataset range. As seen in the examples in Fig. 4, especially for large $H^*(10)$ values such as 100 mSv, the spots segmented by the nnU-Net became smaller and smoother compared to their reference, like (i)-(viii). Shrinkage may be an advantage for separating closely lying tracks occurring at higher $H^*(10)$ values as in (i), (v) and (vii). Another example is illustrated in (v), where three tracks are visible in the raw image, but identified as two tracks in the reference, and also predicted as two tracks (Figs. 4 and 6). This means that the network prediction works and reflects the idea that the performance of the network can only be as good as the reference label masks used for training. A possible solution to improve the separation for cases like (v) would be implementing a deterministic morphological opening transformation within the track counting algorithm (Uchida, 2013).

The metrics regarding the performance of binary segmentation in Fig. 5 yielded values between 0.9 and 1.0 for all three adapted metrics for the low $H^*(10)$ subsets, namely from 0 mSv to 15 mSv. This indicates very accurate binary predictions for the training dataset range. For higher $H^*(10)$ subsets, i.e. 50 mSv and 100 mSv, r_{ad} decreased, indicating smaller predicted spots, along with a lower D_{ad} and IoU_{ad} value, but also a larger discrepancy between the latter two metrics, since the under-segmentation becomes more dominant. This trend probably lies in the fact that training was performed on rather low $H^*(10)$ images. Extending the training dataset by high $H^*(10)$ irradiated FNTDs could improve the predictions.

4.3. Dosimetry for a broad-spectrum irradiation with the mono-energetic trained network

As seen in the examples in Fig. 6, many close tracks were separated correctly. Shrinkage of predicted spots compared to the reference spots can result in splitting one single spot into multiple spots, as in (vi) or vanishing spots, shown in (ii), (iii) and (viii), which influenced proper track recognition during post-processing. This could be caused by heterogeneity in illumination across a readout area and shifted pixel columns from image construction by the FNTD reader, which can be corrected by large-sized Gaussian blurring and small-/medium-sized Gaussian blurring, respectively, applied on the fluorescent raw images before training input (Leong, 2003; Uchida, 2013; Devi et al., 2023). Overlapping tracks as in (iv) were counted as one and could be separated by track properties, like in the previous work of Schmidt et al. (2025a), whose deterministic script involves intensity- and direction-based features for overlapping track separation.

To quantify the dosimetric performance of the nnU-Net-based model, Fig. 7 compares the sensitivity between the reference and the prediction. The ratio of the sensitivities S_{pred} over S_{ref} yields 0.9, indicating accurate dosimetry performance by the nnU-Net-based model, according to the recommended limits of a factor from 0.5 to

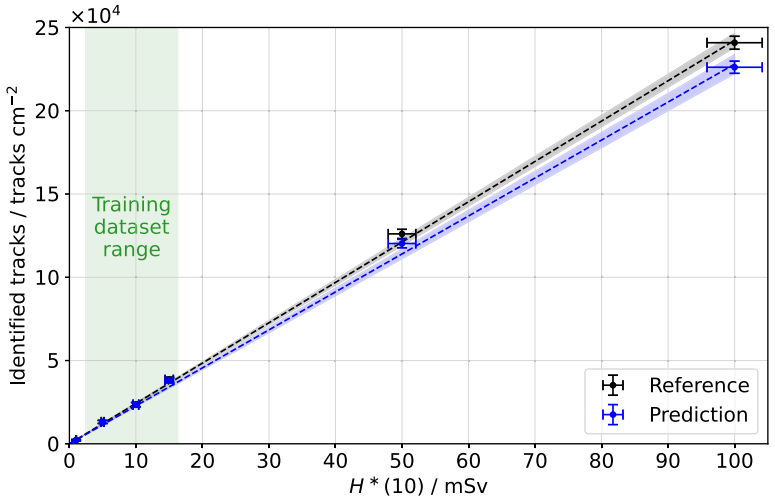

Fig. 7. Dosimetry performance of the network prediction compared with the reference for the broad-spectrum ^{241}Am-Be test dataset. The linear fits are shown with their 95 % confidence interval.

2.0 for neutron dosimetry (ISO 21909-1:2021, 2021). The fit confined to the training dataset range yielded a ratio of 1.0, highlighting even further the dosimetry performance of the network.

Finally, the fitted dose-response can be compared with previous work, where the sensitivity was determined by a deterministic greyscale-based segmentation and track counting script (Schmidt et al., 2025a). This same script was used to obtain the reference label masks, which were used as input for the nnU-Net (Fig. 2). A binary-based track counting script was used afterwards to extract the number of tracks, but the greyscale-based segmentation and track counting script also does the whole procedure, leading to the number of tracks (Supplementary Figure S3). Thus, the binary-based track counting script was benchmarked against the greyscale-based segmentation and track counting script (Supplementary Section S4). Aside from the discussion of benchmarking the track counting script, when comparing the predicted sensitivity to the one previously reported (Schmidt et al., 2025a), the relative deviation resulted in 5.0 %, supporting the potential of the network for neutron dosimetry with FNTDs.

4.4. Potential for its application to higher neutron energies

There are only a few experimental mono-energetic neutron facilities, e.g., the iThemba Laboratory for Accelerator Based Sciences (iThemba LABS, Cape Town, South Africa) or the Research Center for Nuclear Physics (RCNP, Osaka, Japan), that provide fast mono-energetic neutron beams, or the High Intensity Proton Accelerator Facility (HIPA) at the PSI, which produces a broad-spectrum (Nolte et al., 2004; Mosconi et al., 2010; Iwamoto et al., 2011; Hohmann et al., 2025). It was shown that neutron dosimetry with deep learning networks, such as the nnU-Net, shows highly performing outcomes regarding binary segmentation and dosimetry-related metrics. The concept of training with mono-energetic neutron fields and predicting on broad neutron spectra, in this study on ^{241}Am-Be, was demonstrated with success. An extension would be to correlate experimental data as a reference with Monte Carlo simulations. Afterwards, the proven concept can be extended to cover different energy gaps and predict recoil proton tracks generated by higher-energy neutron sources. For secondary neutrons produced during ion beam therapy, this would open the possibility to train on mono-energetic neutron data generated with Monte Carlo simulations, and use the prediction potential of the deep learning for the broad-spectrum high-energy secondary neutrons.

5. Conclusion

A deep learning network was successfully utilised for neutron dosimetry with FNTDs. For the first time, the nnU-Net framework was adapted for track detection for FNTD images. FNTDs irradiated by mono-energetic neutrons were used for model training. The validation metrics for the binary segmentation yielded mean values very close to the ideal 1, namely 0.9505, 0.9239 and 1.000, proving the high performance of the network. This trained model recognised the tracks of the test dataset irradiated with an ^{241}Am-Be energy spectrum, also supported by the test metrics. The dosimetry performance of the network for the whole $H^*(10)$ range yielded a sensitivity of (2280 ± 20) tracks mSv^{-1} cm^{-2} with a deviation of 5.8 % from the reference. For the trained $H^*(10)$ range the sensitivity was (2500 ± 60) tracks mSv^{-1} cm^{-2}, deviating 1.2 % from the reference. The shown results proved the concept of training a deep learning network on mono-energetic irradiated FNTDs, and using this trained network to detect tracks on FNTDs exposed to an ^{241}Am-Be neutron source.

CRediT authorship contribution statement

Long-Yang Jan Thai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stefan Schmidt:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Alexandra Walter:** Writing – review & editing, Methodology, Conceptualization. **Richard V. Häcker:** Writing – review & editing, Software. **Kristina Giske:** Writing – review & editing, Conceptualization. **José Vedelago:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the corresponding author used ChatGPT in order to assist in the development of Python scripts for data analysis. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project was partially financed by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; Project number 495217943). The authors acknowledge support by the state of Baden-Württemberg through bwHPC. The authors are grateful to José Robledo, Marie Piraud and Karol Szustakowski from Helmholtz AI for the fruitful discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.radmeas.2026.107662.

Data availability

The data used for this work is available at the data repository Schmidt et al. (2025b) and the software repository Thai and Vedelago (2025).

References

Akselrod, G., Akselrod, M., Benton, E., Yasuda, N., 2006. A novel Al2O3 fluorescent nuclear track detector for heavy charged particles and neutrons. Nucl. Instrum. Methods Phys. Res. B 247 (2), 295–306. http://dx.doi.org/10.1016/j.nimb.2006.01.056.

Akselrod, M.S., Fomenko, V.V., Bartz, J.A., Haslett, T.L., 2014. Automatic neutron dosimetry system based on fluorescent nuclear track detector technology. Radiat. Prot. Dosim. 161 (1–4), 86–91. http://dx.doi.org/10.1093/rpd/nct293.

Akselrod, M., Fomenko, V., Harrison, J., 2020. Latest advances in FNTD technology and instrumentation. Radiat. Meas. 133, 106302. http://dx.doi.org/10.1016/j.radmeas.2020.106302.

Akselrod, M., Kouwenberg, J., 2018. Fluorescent nuclear track detectors – Review of past, present and future of the technology. Radiat. Meas. 117, 35–51. http://dx.doi.org/10.1016/j.radmeas.2018.07.005.

Baiocco, G., Barbieri, S., Babini, G., Morini, J., Alloni, D., Friedland, W., Kundrát, P., Schmitt, E., Puchalska, M., Sihver, L., Ottolenghi, A., 2016. The origin of neutron biological effectiveness as a function of energy. Sci. Rep. 6 (1), 34033. http://dx.doi.org/10.1038/srep34033.

Bolzonella, M., Ambrožová, I., Caresana, M., Gibbens, N., Gilvin, P., Mariotti, F., Savary, A., Stabilini, A., Vittoria, F., Yukihara, E., Chevallier, M.-A., 2022. Neutron personal dosimetry using polyallyl diglycol carbonate (PADC): Current status, best practices and proposed research. Phys. Open 12, 100114. http://dx.doi.org/10.1016/j.physo.2022.100114.

Dahham, A.M., Kadhim, N.F., Mohammed, R.S., 2025. Utilizing plasma interaction as a novel etching technique for CR-39 nuclear track detector. Eur. Phys. J. Plus 140 (4), 287. http://dx.doi.org/10.1140/epjp/s13360-025-06224-0.

Devi, T.G., Patil, N., Rai, S., Philipose, C.S., 2023. Gaussian blurring technique for detecting and classifying acute lymphoblastic leukemia cancer cells from microscopic biopsy images. Life 13 (2), 348. http://dx.doi.org/10.3390/life13020348.

Gómez-Ros, J., Bedogni, R., Domingo, C., 2023. Personal neutron dosimetry: State-of-the-art and new technologies. Radiat. Meas. 161, 106908. http://dx.doi.org/10.1016/j.radmeas.2023.106908.

Guo, X., Liu, T., Yang, Y., Dai, J., Wang, L., Tang, D., Sun, H., 2024. Automatic segmentation of type a aortic dissection on computed tomography images using deep learning approach. Diagnostics 14 (13), 1332. http://dx.doi.org/10.3390/diagnostics14131332.

Hohmann, E., Galeev, R., Harzmann, S., Reginatto, M., Walter, N., Mayer, S., 2025. Investigation of neutron stray fields close to the two meson targets of the PSI proton accelerator facility. Radiat. Prot. Dosim. 201 (17), 1148–1151. http://dx.doi.org/10.1093/rpd/ncaf093.

Howell, R.M., Burgett, E.A., 2014. Secondary neutron spectrum from 250-MeV passively scattered proton therapy: Measurement with an extended-range Bonner sphere system. Med. Phys. 41 (9), 092104. http://dx.doi.org/10.1118/1.4892929.

ICRP 103, 2007. ICRP publication 103: The 2007 recommendations of the international commission on radiological protection. Chapters 3 and 4. Ann. ICRP 37 (2–4), 49–79. http://dx.doi.org/10.1016/j.icrp.2007.10.005.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211. http://dx.doi.org/10.1038/s41592-020-01008-z.

Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F., 2024. nnU-net revisited: A call for rigorous validation in 3D medical image segmentation. http://dx.doi.org/10.48550/ARXIV.2404.09556.

ISO 21909-1:2021, 2021. Passive neutron dosimetry systems Part 1: Performance and test requirements for personal dosimetry. URL https://www.iso.org/standard/78756.html.

Iwamoto, Y., Hagiwara, M., Satoh, D., Iwase, H., Yashima, H., Itoga, T., Sato, T., Nakane, Y., Nakashima, H., Sakamoto, Y., Matsumoto, T., Masuda, A., Nishiyama, J., Tamii, A., Hatanaka, K., Theis, C., Feldbaumer, E., Jaegerhofer, L., Pioch, C., Mares, V., Nakamura, T., 2011. Quasi-monoenergetic neutron energy spectra for 246 and 389 MeV 7Li(p,n) reactions at angles from 0° to 30°. Nucl. Instrum. Methods Phys. Res. A 629 (1), 43–49. http://dx.doi.org/10.1016/j.nima.2010.12.022.

Joshi, S., Pant, M., Malhotra, A., Deep, K., Snasel, V., 2025. A nnU-Net-based automatic segmentation of FCD type II lesions in 3D FLAIR MRI images. Front. Artif. Intell. 8, 1601815. http://dx.doi.org/10.3389/frai.2025.1601815.

Leong, F.J.W.-M., 2003. Correction of uneven illumination (vignetting) in digital microscopy images. J. Clin. Pathol. 56 (8), 619–621. http://dx.doi.org/10.1136/jcp.56.8.619.

Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., Zoccoli, A., 2021. Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Sci. Rep. 11 (1), 22920. http://dx.doi.org/10.1038/s41598-021-01929-5.

Mosconi, M., Musonza, E., Buffler, A., Nolte, R., Röttger, S., Smit, F., 2010. Characterisation of the high-energy neutron beam at iThemba LABS. Radiat. Meas. 45 (10), 1342–1345. http://dx.doi.org/10.1016/j.radmeas.2010.06.044.

Nikaido, F., Abe, Y., Minami, T., Kuramoto, K., Yasui, T., Sakai, K., Kanasaki, M., Fukuda, Y., Kiriyama, H., Jao, C.S., Chu, C.M., Wu, K.T., Woon, W.Y., Liu, Y.L., Pikuz, T., Hamaguchi, S., Saura, N., Benkadda, S., Kusumoto, T., Kodaira, S., Kuramitsu, Y., 2024. 3D track extraction from a fluorescent nuclear track detector via machine learning and an application to diagnostics of laser-driven ions. Rev. Sci. Instrum. 95 (10), 103302. http://dx.doi.org/10.1063/5.0219480.

Nolte, R., Allie, M.S., Bottger, R., Brooks, F.D., Buffler, A., Dangendorf, V., Friedrich, H., Guldbakke, S., Klein, H., Meulders, J.P., Schlegel, D., Schuhmacher, H., Smit, F.D., 2004. Quasi-monoenergetic neutron reference fields in the energy range from thermal to 200 MeV. Radiat. Prot. Dosim. 110 (1–4), 97–102. http://dx.doi.org/10.1093/rpd/nch195.

Rayed, M.E., Islam, S.S., Niha, S.I., Jim, J.R., Kabir, M.M., Mridha, M., 2024. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. Inform. Med. Unlocked 47, 101504. http://dx.doi.org/10.1016/j.imu.2024.101504.

Schmidt, S., Christensen, J.B., Lutz, B., Stabilini, A., Yukihara, E.G., Vedelago, J., 2025a. Sensitivity analysis of fluorescent nuclear track detectors for fast and high-energy mono-energetic neutron dosimetry. Med. Phys. 52 (7), e17799. http://dx.doi.org/10.1002/mp.17799.

Schmidt, S., Munoz, I., Yukihara, E., Vedelago, J., 2026. Fluorescent Nuclear Track Detectors for out-of-field neutron dosimetry in proton therapy. Med. Phys. 53 (2), e70303. http://dx.doi.org/10.1002/mp.70303.

Schmidt, S., Thai, L.-Y.J., Vedelago, J., 2025b. Dataset of fluorescent nuclear track detector images for neutron dosimetry [dataset]. Mendeley data, v1. http://dx.doi.org/10.17632/pwh8tph424.1.

Stabilini, A., Akselrod, M., Fomenko, V., Harrison, J., Yukihara, E., 2021. Principal Component Analysis applied to neutron dosimetry based on PADC detectors and FNTDs. Radiat. Meas. 141, 106516. http://dx.doi.org/10.1016/j.radmeas.2021.106516.

Thai, L.-Y.J., Vedelago, J., 2025. Deep learning for neutron dosimetry with FNTDs [software]. Github. URL https://github.com/DKFZ-OpenMedPhys/deep-learning-for-neutron-dosimetry-with-FNTDs.

Uchida, S., 2013. Image processing and recognition for biological images. Dev. Growth & Differ. 55 (4), 523–549. http://dx.doi.org/10.1111/dgd.12054.

Vedelago, J., Schmidt, S., Stengl, C., Karger, C.P., Jäkel, O., 2024. Secondary neutrons in proton and light ion beam therapy: A review of current status, needs and potential solutions. Radiat. Meas. 176, 107214. http://dx.doi.org/10.1016/j.radmeas.2024.107214.

Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., Liu, F., 2024. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. Bioengineering 11 (10), 1034. http://dx.doi.org/10.3390/bioengineering11101034.

Zakalek, P., Gutberlet, T., Brückel, T., 2025. Neutron sources for large scale user facilities: The potential of high current accelerator-driven neutron sources. Prog. Part. Nucl. Phys. 142, 104163. http://dx.doi.org/10.1016/j.ppnp.2025.104163.