2nd Symposium on New Trends in Nuclear and Medical Physics, Poland, September 24–26, 2025

Usage of Deep Learning-Based Portal Dose Images for Treatment Error Detection with Transit Dosimetry

E. Uwitonze[a,b], R. Lanzillotta[a,c], C. Mozzi[d,e], L. Marini[a,c], M. Avanzo[f], F. Lizzi[a], L. Marrazzo[d,e,g], I. Meattini[d,g], S. Pallotta[d,e,g], G. Pirrone[f], A. Retico[a], C. Talamonti[d,e,g] and A.C. Kraan[a,*]

[a] *National Institute for Nuclear Physics (INFN), Section of Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy*
[b] *Rwanda Cancer Centre, KK 739 Street, Kicukiro, Kigali, Rwanda*
[c] *University of Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy*
[d] *University of Florence, Viale Morgagni 50, 50134 Florence, Italy*
[e] *National Institute for Nuclear Physics (INFN), Section of Florence, Via Sansone 1, 50019 Sesto Fiorentino, Italy*
[f] *Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Via Franco Gallini 2, 33081 Aviano, Italy*
[g] *Careggi University Hospital, Largo Giovanni Alessandro Brambilla 3, 50134 Florence, Italy*

Doi: 10.12693/APhysPolA.148.S74 *e-mail: aafke.kraan@pi.infn.it

In-vivo dose monitoring with electronic portal imaging devices (EPIDs) in radiotherapy can be performed by comparing a recorded EPID image with a reference expected image, either directly or by first converting it into water-equivalent dose (portal dose). We developed a deep learning model that transforms EPID images into water-equivalent dose images. An analysis framework was created, which compares the portal dose images with expectations. In this work, we test whether the framework, based on the gamma analysis, can detect changes in monitor units. We irradiated an inhomogeneous phantom using a geometric field and acquired EPID images with a planned number of monitor units, as well as with a controlled excess of monitor units. These EPID images were transformed into portal dose images using our deep learning model. Then we evaluated how the gamma passing rate varied with delivered monitor units, investigating several tolerance criteria and thresholds. The errors in monitor units could be well identified by our deep learning-based portal dose analysis framework. The choice of the gamma index threshold influenced the action level. Using a dose threshold of 10% and a tolerance of 3%/3 mm, the gamma passing rate dropped below 95% when the error in monitor units was larger than 102.4 and 102.3 for global and local gamma index calculations, respectively. For an 80% threshold, the values decreased to 102.2 for global and local calculations. When using a newly developed analysis framework, based on comparison of deep learning-based portal dose images with the gamma index, we confirmed that we can detect errors in monitor units. As a starting point for an alert system, the global gamma index analysis can be used with an 80% threshold and 3%/3 mm or stricter criterion.

topics: portal dose (PD) images, transit dosimetry, radiotherapy, deep learning

1. Introduction

In-vivo dosimetry can be performed using electronic portal imaging devices (EPIDs) in radiotherapy by performing dosimetric comparisons between predicted and measured images [1–8]. In forward transit dosimetry, EPID images can be measured after passing through the patient (transit images) and compared with expectations. These comparisons can be done directly, in which case an acquired EPID image is compared with an expected EPID image (in greyscale), or the comparisons can be done after first converting the EPID image to water-equivalent dose at EPID level (portal dose (PD), in units of Gy). The latter has the advantage that quantitative (dose) comparisons can be done.

EPID-based in-vivo transit measurements can be used in alert systems and have proven to have a high effectiveness in detecting treatment errors, such as patient setup errors, machine malfunctions, and other treatment delivery deviations [9]. Examples of commercially available alert systems for in-vivo dose verification based on EPID measurements are summarized in AAPM Task Group 307 report [7]. The comparison strategies applied include both direct EPID image comparison and quantitative portal-dose comparisons.

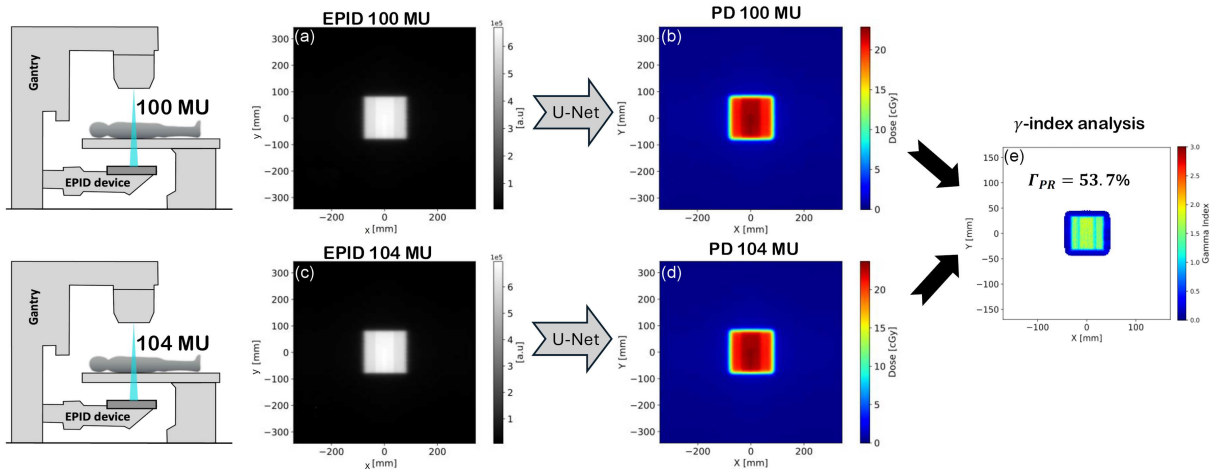

Fig. 1. An EPID (a) and PD (b) image for the MultiPlug phantom acquired with 100 MU (reference situation). Panels (c) and (d) show the EPID and PD image for the MultiPlug phantom acquired with 104 MU. The global gamma index map is given for a 2%/2 mm tolerance and 10% threshold when comparing the 100 and 104 MU PD images (e).

We have developed an in-house deep learning (DL)-based model to convert measured EPID images into portal dose images [10]. This model was trained and validated with a large dataset of EPID measurements coupled with simulated portal dose images [11]. The fast calculation of a quantitative portal dose image from an EPID image makes the framework appropriate for application in a real-time alert framework. We are currently developing such a framework.

The goal of the present work is to test whether the developed analysis framework is sensitive to changes in monitor units. This will be done by analyzing several measurements that were taken while intentionally introducing errors. Particularly, we will investigate the gamma index analysis using different tolerances and thresholds and applying both local and global gamma index calculations. As far as we know, these aspects have not been investigated previously in a systematic manner with DL-based portal dose images.

2. Materials and methods

2.1. Artificial intelligence model

We recently created a deep-learning model, based on a U-net [10], that converts a transit dosimetry EPID image into a portal dose (PD) image. The principle of transit EPID dosimetry is shown in Fig. 1 (left side of panels (a) and (c)). The EPID image approximately measures the X-ray fluence that exits the patient. The model was based on a dataset of 220 measured EPID images on different phantoms, each paired with a corresponding PD image. Here, 90% was used for training/validation, and 10% for testing. More details are given in [11].

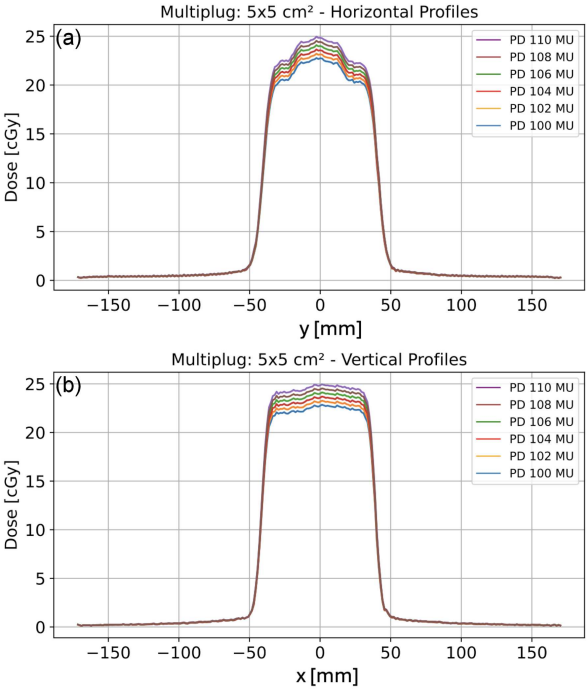

Fig. 2. Horizontal and vertical portal dose profiles for the MultiPlug phantom, for a 5×5 cm^2 field acquired with 100 (blue), 102 (orange), 104 (red), 106 (green), 108 (brown), and 110 (purple) MUs.

The EPID images were obtained by irradiating different phantoms with many treatment fields in different conditions, using 6 MV photon beams from a Versa HD Elekta Linac at Careggi Hospital in Florence. The paired PD images were obtained using simulations from the Monaco treatment planning system.

TABLE I

Values where the GPR drops below 95% for global and local gamma index calculations, using thresholds of 10% and 80%.

Calculation	Threshold	2%/2 mm	3%/3 mm	4%/4 mm	5%/5 mm
Global	10%	102.2	102.4	104.3	104.8
Local	10%	102.2	102.3	104.3	104.5
Global	80%	102.1	102.2	104.2	104.5
Local	80%	102.1	102.2	104.2	104.3

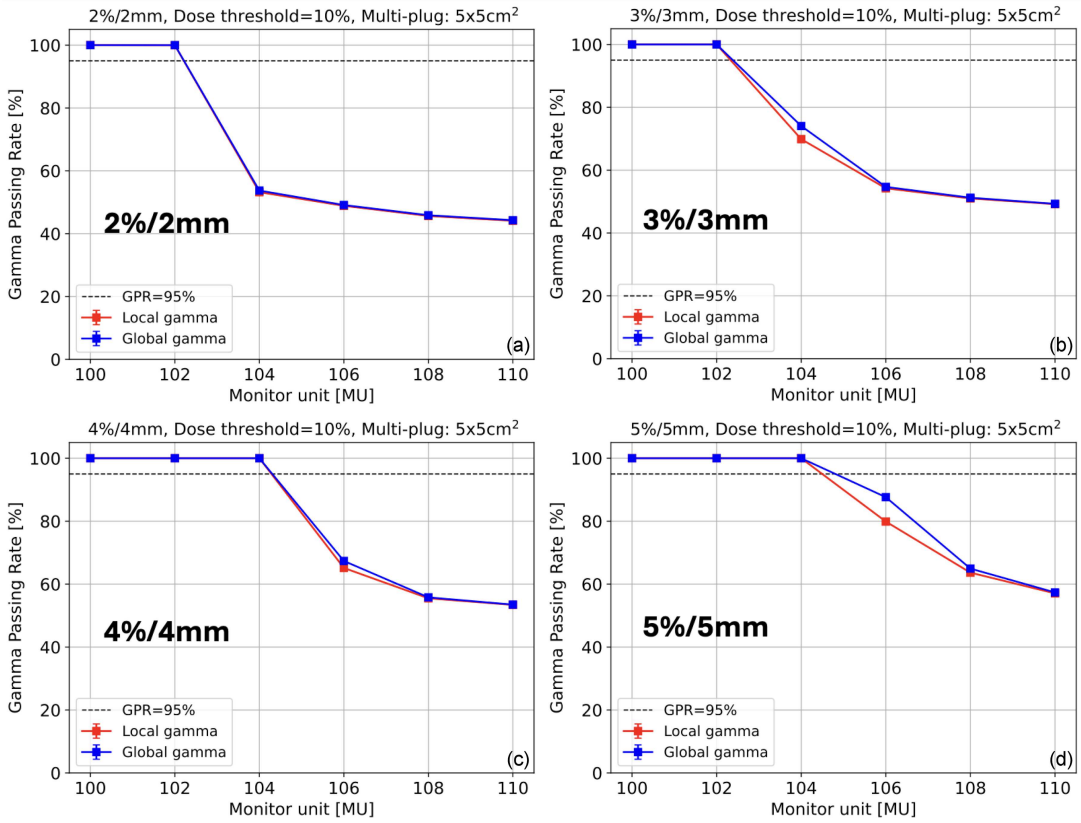

Fig. 3. Gamma passing rate versus monitoring units, for global (blue) and local (red) gamma index analysis with 2%/2 mm (a), 3%/3 mm (b), 4%/4 mm (c), and 5%/5 mm (d), all with a 10% threshold, for the MultiPlug phantom.

2.2. Measurements

We irradiated a MultiPlug phantom (Model 1220, P/N 1220000-3Z, Sun Nuclear Corporation, Melbourne, FL, USA) at the center of the phantom with a field with zero gantry angle of 5×5 cm^2 (at isocenter) with 100 MU (reference situation). Then, we introduced controlled errors in monitoring units (MUs), keeping all circumstances the same except for the MUs, which were varied from 100 MU to 102, 104, 106, 108, and 110 MU. In Fig. 1a, we show a reference EPID image, $\mathrm{EPID}_{\mathrm{ref}}$, together with corresponding reference portal dose image, $\mathrm{PD}_{\mathrm{ref}}$, in Fig. 1b, acquired with 100 MU (no error). Figure 1c shows an EPID image, together with a corresponding portal dose image, PD_i, in Fig. 1d, acquired with an error, in this case 104 MU. Indeed, slightly higher portal dose values are obtained with 104 MUs, as can be seen in Fig. 1 from the color scale bars on the right.

Figure 2 shows the central horizontal (a) and vertical (b) profiles of the DL-based portal dose predictions, from 100 to 110 MU, for the MultiPlug phantom irradiated with a 5×5 cm^2 field.

2.3. Analysis

A metric that is recommended for EPID and portal dose comparisons is the gamma index analysis [7, 12, 13]. It combines both spatial and

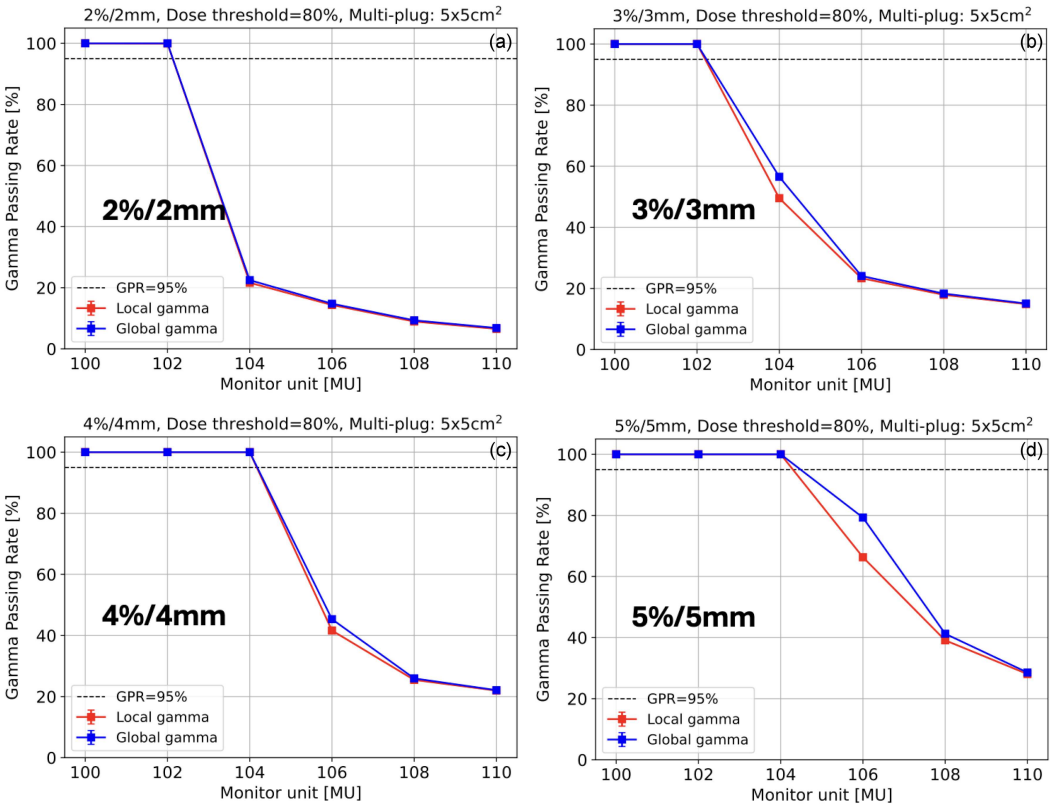

Fig. 4. Gamma passing rate versus Monitor Units, for global (blue) and local (red) gamma index analysis with 2%/2 mm (a), 3%/3 mm (b), 4%/4 mm (c), and 5%/5 mm (d), all with an 80% threshold, for the MultiPlug phantom, irradiated with a 5×5 cm^2 field.

physical dose information [14]. The main free parameters are the dose difference (DD) tolerance, distance-to-agreement (DTA) tolerance, and dose threshold (TH). Their recommended values depend on the treatment center and the patient group [7, 15]. As a threshold value, 10% of the maximum reference dose is widely used [16, 17], but values up to 80% have been reported for dose comparisons [18]. As tolerance criteria DD/DTA, we tested 2%/2 mm, 3%/3 mm, 4%/4 mm, and 5%/5 mm. While 2%/2 mm is stricter than what is typically applied in in-vivo dosimetry, it has been applied in pre-treatment and phantom analyses [7]. The gamma analysis can be global, where the DD is normalized to a single maximum dose (from the reference PD), or local, where the DD is normalized to its own expected dose. The resulting gamma index distributions were all visually inspected. An example is given in Fig. 1e, where two PD images are compared. Based on the gamma index distribution, the gamma passing rate (GPR) is calculated as

$$\mathrm{PR} = \frac{\sum_{v,\mathrm{pass}} \Gamma_v^{\mathrm{pass}}}{\sum_{v,\mathrm{tot}} \Gamma_v^{\mathrm{valid}}}, \qquad (1)$$

where Γ_v is a book value for voxel v, $\Gamma_v^{\mathrm{pass}} = 1$ if the pixel gamma index value is below 1 and above threshold, $\Gamma_v^{\mathrm{pass}} = 0$ otherwise; and $\Gamma_v^{\mathrm{valid}} = 1$ if above threshold and $\Gamma_v^{\mathrm{valid}} = 0$ otherwise. We assume two distributions to be clinically equivalent when PR is at least 0.95, in other words, when at least 95% of all pixels pass the tolerance.

The workflow of our analysis was as follows. First, for all PD distributions, we tested how GPR varies with the delivered MUs, using different combinations of thresholds and tolerances. The analysis was performed in Python, and we used the gamma index function from PyMedPhys [19, 20]. Second, we evaluated at what MU value PR drops below the 95% limit, and extracted the corresponding value, $\mathrm{MU}_{\mathrm{lim}}$, by numerical interpolation. Third, we investigated the difference between global and local gamma index analysis, performing PD comparisons, again using various thresholds and tolerances.

3. Results

3.1. Portal dose gamma analysis

Figure 1e shows the gamma index distribution that was obtained with 2%/2 mm and a 10% threshold when comparing the PD distribution obtained with 104 MUs with the reference. The corresponding GPR was 53.7%. Hence, this kind of error would be detected in this case.

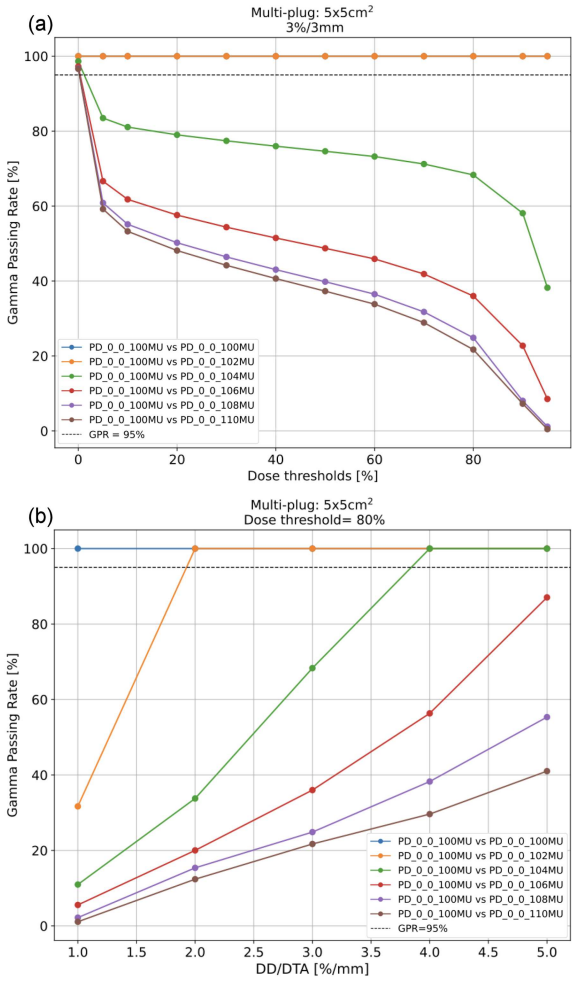

Fig. 5. (a) GPR versus dose threshold chosen in the gamma index analysis, when comparing several acquisitions with different excess in MU (100, 102, 104, 106, 108, and 110 MU) with the reference situation of 100 MU. (b) The GPR versus the DD/DTA criteria, for several acquisitions with different excess in MU (100, 102, 104, 106, 108, and 110 MU) compared with the reference situation at 100 MU.

In Fig. 3, the GPR is given as a function of the delivered monitor units for the local (red) and global (blue) gamma calculation, applying tolerances of 2%/2 mm (a), 3%/3 mm (b), 4%/4 mm (c), and 5%/5 mm (d). A threshold of 10% of the maximum reference dose value was applied here. Figure 3 shows that the GPR decreases with increasing error, as expected. The difference between the global and local gamma index analysis was small: the largest discrepancy found was at 5%/5 mm criteria, where at 106 MU, the difference between the GPR based on a local and global calculation was found to be roughly 8% for the 10% dose threshold.

The same was repeated for a threshold of 80%, with the result given in Fig. 4. The largest difference between the global and local gamma index analysis was found with the 5%/5 mm criteria again at 106 MU, where it was 13%.

The values for MU where the PR crossed the 95% level were evaluated and are displayed in Table I for the local and global gamma analyses with tolerances of 2%/2 mm (third column), 3%/3 mm (fourth column), 4%/4 mm (fifth column), and 5%/5 mm (sixth column) for a threshold of 10% and 80%.

For the 3%/3 mm, we investigated in more detail the dependence of the GPR on the dose cut-off. Figure 5a shows the GPR as a function of the applied threshold. We can see that there is a strong dependence of the GPR on the dose threshold chosen, as expected. If the dose threshold is high, the GPR is smaller, and thus the alert framework will be more sensitive to errors. Figure 2, showing wide dose tails, confirmed that it is indeed reasonable to apply moderate to high dose thresholds, e.g., 80% in these cases, so low-dose tails are excluded. Figure 5b shows the GPR as a function of the DD/DTA criteria for the different acquisitions, including also the 1%/1 mm values.

4. Discussion

Our framework, based on PD distributions obtained from our DL model, was able to catch errors in excess in MUs. In all cases, the GPR decreases with increasing value of the introduced error, as expected. The gamma analysis only weakly depended on the type of gamma index calculation chosen (local or global). Theoretically, the global gamma calculation results in lower gamma index values (thus higher GPR results). The global gamma theoretically yields lower values (higher GPR results) but may mask errors in low-dose regions [16–18], while local gamma can be too sensitive there. To align with some other alert frameworks [7], we propose using global gamma by default. In any case, the differences are small.

Regarding the threshold, 80% of the reference maximum is probably preferred over 10%. It ensures that the PD analysis is focused on the high-dose region and automatically leads to somewhat lower passing rates. Crossing the alert value of GPR $< 95\%$ occurs thus at smaller MU values when a high threshold is applied than when a low threshold is applied, leading to more conservative alert values.

Clinically, it is desirable to be able to detect deviations of about 3% in MUs [7], or even less. From Figs. 3 and 4, we can see that for the Multi-Plug phantom, this requires tolerances of 3%/3 mm or less. For the development of a patient alert system, patient-specific values may apply. Values like 3%/3 mm or 2%/2 mm may be a good starting point in the first tests of a future, more advanced alert system. This is similar to the values that are

reported for several patient groups in the AAPM Task Group 307 report [7]. New measurements with other phantoms and fields would be needed to confirm this, including also different types of treatment errors, such as positioning, gantry angles, or anatomical errors.

An important advantage with respect to most previously published works is that the DL-based alert framework eliminates the need for traditional, complex correction and calibration procedures, making it a highly valuable tool for clinical in-vivo dose verification that improves patient safety.

5. Conclusions

The setting of the tolerance criteria is one of the main challenges in the development of alert systems and needs to take many factors into account. This is true especially for DL-based predictions. In this work, we identified a starting set of tolerances and thresholds by testing the detection of errors in MUs for a MultiPlug phantom. We found that a global gamma calculation, 80% threshold, and tolerances of 3%/3 mm or stricter could be applied. The presented results are highly useful as a starting point for more specific studies for the detection of treatment errors, using DL-based portal dose images. An alert system based on DL-based dose comparisons could improve patient safety while requiring minimal workload.

Acknowledgments

This research was partly funded by the InTrEPID project (MUR PRIN 2022CWXR8K — CUP I53D23000520006) and the ARTEMIS project (INFN, CSN5).

References

[1] W. van Elmpt, L. McDermott, S. Nijsten, M. Wendling, P. Lambin, B. Mijnheer, *Radiother. Oncol.* **88**, 289 (2008).

[2] B. McCurdy, P. Greer, J. Bedford, in: *Clinical 3D Dosimetry in Modern Radiation Therapy*, Ed. Ben Mijnheer, Boca Raton, CRC Press, 2017.

[3] B.M.C. McCurdy, *J. Phys. Conf. Ser.* **2630**, 012009 (2023).

[4] B. Mijnheer, S. Beddar, J. Izewska, C. Reft, *Med Phys.* **40**, 070903 (2013).

[5] B. Mijnheer, *J. Phys. Conf. Ser.* **1305**, 012061 (2019).

[6] S. Celi, E. Costa, C. Wessels, A. Mazal, A. Fourquet, P. Francois, *J. Appl. Clin. Med. Phys.* **17**, 262 (2016).

[7] N. Dogan, B.J. Mijnheer, K. Padgett et al., *Med Phys.* **50**, e865 (2023).

[8] M. Miften, A. Olch, D. Mihailidis et al., *Med. Phys.* **45**, e53 (2018).

[9] E.C. Ford, S. Terezakis, A. Souranis, K. Harris, H. Gay, S. Mutic, *Int. J. Radiat. Oncol. Biol. Phys.* **84**, e263 (2012).

[10] L. Marini, M. Avanzo, A.C. Kraan, F. Lizzi, C. Mozzi, A. Retico, C. Talamonti, *Nucl. Instrum. Methods Phys. Res. A* **1069**, 169908 (2024).

[11] C. Mozzi, L. Marini, M. Avanzo, A. Kraan, F. Lizzi, L. Marrazzo, I. Meattini, A. Retico, C. Talamonti, *Radiother. Oncol.* **206**, S2616 (2025).

[12] E. Marco, R. Baldoni, E. Bossuyt et al., *Phys. Imaging Radiat. Oncol.* **32**, 100666 (2024).

[13] Y. Xia, J. Adamson, Y. Zlateva, W. Giles, *J. Radiosurg. SBRT* **7**, 135 (2020).

[14] D.A. Low, W.B. Harms, S. Mutic, J.A. Purdy, *Med. Phys.* **25**, 656 (1998).

[15] I. Olaciregui-Ruiz, R. Rozendaal, B. Mijnheer, A. Mans, *Med Phys.* **46**, 45 (2019).

[16] M. Miften, A. Olch, D. Mihailidis et al., *Med. Phys.* **45**, e53 (2018).

[17] D. Zhang, B. Wang, G. Zhang, C. Ma, X. Deng, *J. Appl. Clin. Med. Phys.* **19**, 230 (2018).

[18] E. Timakova, S.F. Zavgorodni, *J. Appl. Clin. Med. Phys.* **25**, e14459 (2024).

[19] M. Wendling, L.J. Zijp, L.N. McDermott, E.J. Smit, J.-J. Sonke, B.J. Mijnheer, M. van Herk, *Med. Phys.* **34**, 1647 (2007).

[20] S. Biggs, M. Jennings, S. Swerdloff et al., *PyMedPhys* **7**, 4555 (2022).