Physics in Medicine & Biology

PAPER • **OPEN ACCESS**

Determination of output factor for CyberKnife using scintillation dosimetry and deep learning

To cite this article: Jeremy Ocampo *et al* 2024 *Phys. Med. Biol.* **69** 025024

IOP Publishing Phys. Med. Biol. 69 (2024) 025024 https://doi.org/10.1088/1361-6560/ad1b69

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Determination of output factor for CyberKnife using scintillation dosimetry and deep learning

Jeremy Ocampo[1], **Geoff Heyes**[2], **Hamid Dehghani**[3], **Tim Scanlon**[1], **Simon Jolly**[1] and **Adam Gibson**[4,*]

[1] UCL Physics and Astronomy, London, WC1E 6BT, United Kingdom
[2] Radiotherapy Physics, University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2TH, United Kingdom
[3] School of Computer Science, University of Birmingham, Birmingham B15 2TT, United Kingdom
[4] UCL Medical Physics & Biomedical Engineering, London, WC1E 6BT, United Kingdom
* Author to whom any correspondence should be addressed.

E-mail: adam.gibson@ucl.ac.uk

Keywords: small field dosimetry, output factor, cyberknife, solid plastic scintillator, photography, deep learning

OPEN ACCESS

RECEIVED
18 September 2023

REVISED
19 December 2023

ACCEPTED FOR PUBLICATION
5 January 2024

PUBLISHED
22 January 2024

Abstract

Objective. Small-field dosimetry is an ongoing challenge in radiotherapy quality assurance (QA) especially for radiosurgery systems such as CyberKnife™. The objective of this work is to demonstrate the use of a plastic scintillator imaged with a commercial camera to measure the output factor of a CyberKnife system. The output factor describes the dose on the central axis as a function of collimator size, and is a fundamental part of CyberKnife QA and integral to the data used in the treatment planning system. *Approach*. A self-contained device consisting of a solid plastic scintillator and a camera was build in a portable Pelicase. Photographs were analysed using classical methods and with convolutional neural networks (CNN) to predict beam parameters which were then compared to measurements. *Main results*. Initial results using classical image processing to determine standard QA parameters such as percentage depth dose (PDD) were unsuccessful, with 34% of points failing to meet the Gamma criterion (which measures the distance between corresponding points and the relative difference in dose) of 2 mm/2%. However, when images were processed using a CNN trained on simulated data and a green scintillator sheet, 92% of PDD curves agreed with measurements with a microdiamond detector to within 2 mm/2% and 78% to 1%/1 mm. The mean difference between the output factors measured using this system and a microdiamond detector was 1.1%. Confidence in the results was enhanced by using the algorithm to predict the known collimator sizes from the photographs which it was able to do with an accuracy of less than 1 mm. *Significance*. With refinement, a full output factor curve could be measured in less than an hour, offering a new approach for rapid, convenient small-field dosimetry.

1. Introduction

Quality assurance (QA) of radiotherapy treatment machines is a necessary part of the treatment pathway. It is governed by Codes of Practice (Eaton *et al* 2020) which specify formal measurement protocols for reference dosimetry that are traceable back to a primary standard. Relative dosimetry is less exacting and may be presented as a ratio which is normalised to a peak. As it is less time-consuming, relative dosimetry is performed more frequently than reference dosimetry. Since changes to relative dosimetry may be used as an indicator for a change in reference data (and therefore signify a deviation from the treatment planning system model), relative dosimetry still must be performed with rigour and appropriate expertise.

There are particular challenges for dosimetry of small field sizes. An ionisation chamber with a collection volume larger than the beam size will suffer from partial volume effects and could significantly perturb the dose distribution. Ionisation chambers therefore require small-field corrections and so solid-state detectors based on diodes or diamonds in a water tank are used for relative dosimetry, with some degree of uncertainty (Chalkley and Heyes 2014). The CyberKnife™ (Accuray Inc., Sunnyvale, CA) system is a robot-mounted linear accelerator

that delivers 6 MV radiation with a field size determined by secondary fixed collimators of 5–60 mm diameter. It does not use a flattening filter, meaning that the dose rate is relatively high and the beam profile is not uniform. As well as being inherently a small-field device, the CyberKnife is also inherently four-dimensional—it can deliver radiation from any angle around the patient as a function of time.

Perhaps the most fundamental metric for relative dosimetry is the *percentage depth dose curve (PDD)* which plots the dose deposited by the beam as it varies with depth along the central axis, normalised to d_{max}, the maximum dose on the central axis. Similarly, the *profile* measures the dose perpendicular to the central axis, normalised to d_{max}. Together, these curves, and parameters calculated from them, characterise many of the beam properties in a visual and intuitive manner. These data will be used by the treatment planning system to calculate the dose within a patient at points distant from the isocentre. A fundamental dataset within the planning system is the *relative output factor* (ROF) which describes the change in dose on the central axis with changing beam size as determined by the secondary collimator. A larger beam size leads to more scattered radiation and so a higher dose on the central axis even as the number of monitor units delivered remains the same. The output factor is defined for Cyberknife as the dose on the central axis at a depth of 15 mm for each collimator at a source-surface distance of 785 mm (800 mm source to axis distance), normalised to the equivalent dose for the largest (60 mm diameter) collimator. This is currently measured using an ionisation chamber or solid-state detector in a water tank. Measurements to fully characterise the machine using a variety of chambers is labour intensive and requires significant expertise in small field dosimetry. Detector choice, volume, orientation, initial alignment to the beam central axis, and positioning accuracy within the beam all affect the uncertainty of the measurement, and minimising these uncertainties is time consuming. Changes to machine running parameters (for example beam steering adjustments, or changes to the beam energy) will affect the accuracy of the dose calculation within the patient, and as such the data in the planning system requires periodic validation, and scheduled measurements to verify following a repair or manufacturer intervention.

Verifying the entire 2D or 3D dose distribution is complex. One metric that is used is the *Gamma Index* which combines the relative distance between corresponding points and the relative difference in dose (Low and Dempsey 2003, Hussein *et al* 2017). A threshold of acceptability is generally quoted, for example 3%/3 mm, and the percentage of pixels that exceed this threshold is quoted.

An ideal dosimeter for small-field dosimetry would have certain characteristics:

- Water equivalence to allow direct translation to clinical practice;
- A spatial resolution better than 0.5 mm to allow high dose gradients to be plotted;
- Linear response with dose and dose rate to minimise the need for corrections;
- Able to detect from multiple points simultaneously or rapidly so that a PDD or profile can be obtained efficiently;
- Able to be automated so that multiple collimators can be assessed efficiently;
- Capable of recording Gamma Index to 1%/1 mm accuracy; and
- Ease of use.

Scintillating materials can offer an alternate approach for relative dosimetry. This has been most commonly proposed using plastic scintillating fibres coupled to a sensitive light detector (Beddar 2006, Kam *et al* 2022). These devices meet some of the criteria above, but can only sample a limited number of points in the radiation field. Photography of a solid scintillation block has been proposed for relative dosimetry of photon, electron and proton beams (Helo *et al* 2014, Beaulieu and Beddar 2016, Almurayshid *et al* 2017) but has not yet been accepted as a reliable method for providing robust, quantitative relative measurements. Cherenkov emissions from photon beams can also be imaged in conjunction with scintillation (Frelin *et al* 2005, 2008) or in its own right for dosimetry (Pogue *et al* 2017) and for patient verification (Roussakis *et al* 2015), but like scintillation imaging has not yet been accepted as a clinical tool.

Here, we investigate photography of a solid plastic scintillator block as a potential alternative device for small-field relative dosimetry. A plastic scintillator is close to water equivalent; the spatial resolution is determined by the camera; dose and dose rate responses are linear; and an entire image is acquired rapidly from which PDD, profile, output factor and other parameters can be determined. We show that such a device analysed using convolutional neural networks (CNN) can record dose distribution with sufficient accuracy for relative dosimetry. Due to the ease of use of the system, plus the potential for automation, this would increase the clinical availability of the system as less non-clinical time would be required for scheduled verification measurements, and less time would be needed to verify the clinical accuracy of the machine after a repair intervention. We do

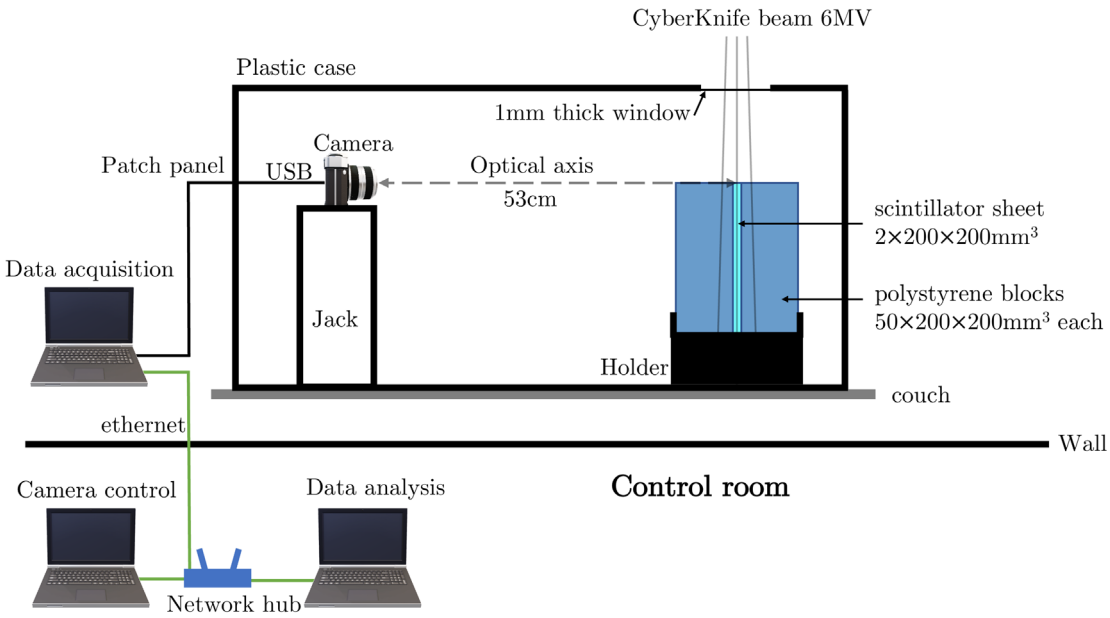

Figure 1. Schematic diagram showing the detector in a Peli case in the treatment room, with control and analysis computers in the control room.

not propose this as a solution for absolute dosimetry, or as a substitute for the initial 'gold standard' commissioning data with a watertank and detector, however it could be used to provide a verification dataset for routine and scheduled quality assurance.

2. Method

2.1. Device configuration

The treatment beam excites a volume from which it is challenging to isolate the central axis, which is required to reproduce the PDD and output factor. This problem was solved by using a thin scintillating sheet sandwiched between two non-scintillating polystyrene blocks. The two blocks provide realistic water-equivalent scatter while the thin scintillating sheet isolates a plane through the central axis, perpendicular to the optical axis of the camera (figure 1). However, as the beam size increases, it excites both scintillation from the thin sheet in proportion to the diameter of the beam and also Cherenkov radiation from the volume in proportion to the area of the beam, meaning that the contribution from Cherenkov radiation quickly dominates.

Figure 1 is a schematic diagram of the final device. A 2 mm thick polystyrene scintillator sheet, sandwiched between two 200 × 200 mm non-scintillating polystyrene blocks, is mounted inside a Peli case (Peli, UK) together with a Nikon D7500 SLR camera with 20 mm lens and 20.9 MPixel detector. The Peli case has a thin plastic window, ensuring the system is light-tight while minimising beam distortion due to build-up. It is placed on the CyberKnife couch and aligned using the CyberKnife room lasers (these are set to the centre of the room: a point in space upon which the in-room kV imaging system is aligned). A patch panel allows for easy communication with a control laptop which communicates by Ethernet to a network hub in the control room, allowing multiple users to connect for simultaneous control and analysis.

The polystyrene blocks have density 1.03 g cm^{-3} and refractive index 1.57. Two different scintillating sheets were used, both polystyrene, a standard blue scintillator (SP32 from Nuvia, Czech Republic with a peak emission at 425 nm) and a novel green scintillator (SP33, Nuvia, (Hamel *et al* 2020), with peak emission at 503 nm). Cherenkov emission is predominantly blue, so it was anticipated that a green scintillator would maximise the scintillation-to-Cherenkov ratio in the green channel of the camera. The central axis of the camera was aligned with the top edge of the plastic blocks and the distance from the camera to the plastic block fixed so that the entire width of the 20 cm block was in frame. If the camera pointed at the centre of the plastic blocks, there was a significant artefact from light generated in the scintillator, reflecting off the underside of the plastic block and obscuring the entry point.

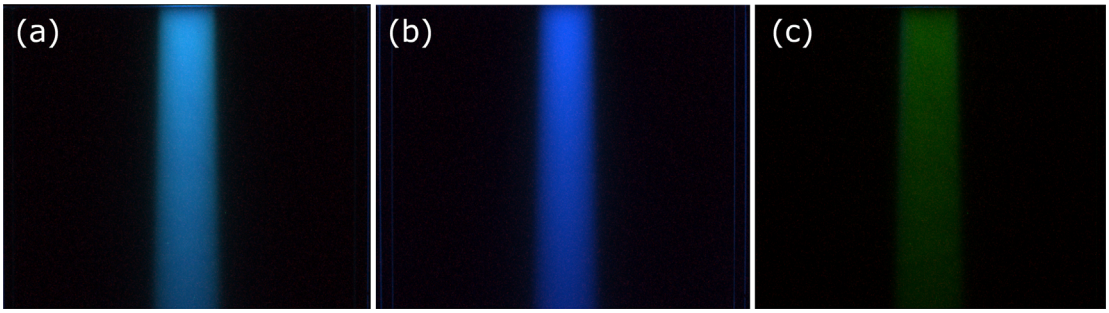

Figure 2. Example images for the 30 mm collimator, all normalised to the same intensity. (a) is an original photograph from the green scintillator-polystyrene block; (b) shows Cherenkov light form the polystyrene block only; (c) is the subtracted image showing scintillation light only.

2.2. Experimental method

The 6 MV treatment beam emits 1000 ± 15 monitor units (MU)/min and delivers 10 Gy min^{-1} at 15 mm depth with the 60 mm collimator beam. The machine was aligned orthogonally with the upper surface of the plastic block, and centred onto the scintillating sheet. CyberKnife output factors are measured at 785 mm source to surface distance (SSD) with the measurement point at 15 mm deep. The treatment planning system utilises tissue phantom ratio (TPR) measurements, which is the ratio of absorbed dose at a given point to the dose at a fixed reference depth using a constant source to axis distance by varying the SSD. For comparative data, and ease of subsequent experimental setup, a PDD dataset can be measured alongside the TPR. PDD uses a constant SSD, simplifying the measurement technique in a watertank. A full set of dosimetry measurements (PDD, ROF and Off Centre Ratios) were made with a PTW-60019 synthetic single-crystal microDiamond^{TM} (PTW, Germany) detector (repeating the method of Chalkley and Heyes 2014) at an SSD of 785 mm, with scintillation photography performed at the same SSD for direct comparison. The camera ISO and aperture were set to 400ISO and f/5.6, and the exposure time adjusted from 10 s to 30 s depending on the beam diameter to ensure the images were well exposed without saturation. Photographs were taken while the beam was on, with the timing determined by the camera shutter and were stored in Nikon's NEF raw data format.

Every image taken with the polystyrene blocks and 2 mm thick scintillator sheet was paired with one without the scintillator so that Cherenkov light could be identified and subtracted to generate an image showing scintillation alone (figure 2). However, for large field sizes, the Cherenkov light was of a similar order of magnitude to the scintillation light, so even after the Cherenkov image was subtracted, bias and statistical errors remained in the scintillation image. Each measurement consisted of five identical repeated images. Dark background images were acquired with the beam off and x-ray scatter was measured from the top half of the image where no scintillator was present. The dark background was typically 0.03% of the maximum intensity and the x-ray scatter was 1%–5% depending on the exposure time.

2.3. Monte Carlo simulation

The experimental pipeline was tested and verified with a Monte Carlo model implemented in GEANT4 (Allison *et al* 2016) using the *G4EmLivermorePhysics* and *G4OpticalPhysics* classes and reproducing the geometry shown in figure 1. The dose deposited in the polystyrene block was firstly modelled, with the resulting scintillation and Cherenkov photons then simulated, propagated and focussed onto an image plane using a thin lens approximation with a lens aperture of 4.1 cm, which ensured that the entire depth of the polystyrene block was in focus. The speed of the simulation was increased by killing photons which were repeatedly internally reflected and by simulating only those optical photons which are emitted in a solid angle that encompasses the lens. These improvements decreased the time required to detect 10^5 simulated optical photons from about a week to about 10 h using an Intel Xeon E5-2430 2.7 GHz CPU.

The model was validated by simulating dose distributions and comparing them to the dose measured with a microdiamond detector. The mean absolute difference (MAD) between simulated and measured dose was less than 1% for the profile and output factor. For the PDD, this increased to 2% as the simulated dose distribution increasingly underestimated the dose compared to microdiamond measurements as the depth increased. When simulated and measured photographs were compared at different field sizes, the MAD was generally less than 2%, but increased to about 3% as the PDD got deeper and to about 5% at the steep edges of the profile. We believe this was due to the scintillation light being scattered within the polystyrene block.

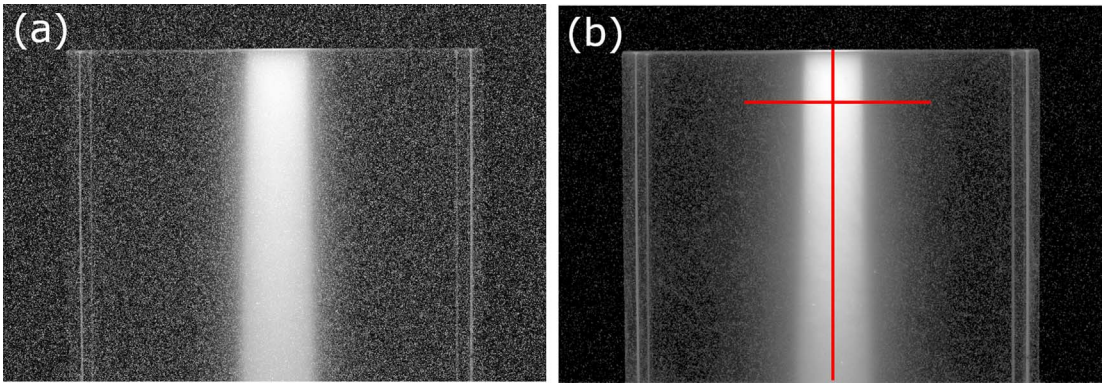

Figure 3. The green channel of photographs of the scintillating block (a) before and (b) after application of the noise reduction process described in section 2.4. To emphasise the difference, contrast has been stretched by histogram equalisation. The lines along which the PDD and profile are extracted are shown in (b).

2.4. Classical image processing

The greatest source of noise in the image was due to photons from the scattered x-rays, leading to a mean noise-to-signal ratio of 0.60. The most effective way to reduce this was to apply two median filters. First, each pixel was adjusted so that it became the median value of that pixel across the five images. Second, a spatial median filter with a 3×3 pixel kernel was applied to the image five times. This reduced the mean noise-to-signal ratio to 0.04, without distorting the spatial distribution of the image (figure 3).

The polystyrene blocks were photographed with a ruler to provide an absolute length scale which confirmed the spatial resolution was 90 μm. All photographs were rotated so that an image showing the alignment laser would appear vertical.

Each photograph of the scintillator sheet includes components from Cherenkov and scintillation light (figure 2). We want an image that shows only the scintillation signal so the Cherenkov signal (taken from the corrected photograph obtained from the plastic block with no scintillator) is subtracted from the photograph showing the combined signal, as implemented by Beddar *et al* (1992), Yogo *et al* (2017) and others. Finally, corrections are applied for refraction (with the application of Snell's law), light intensity (with a $1/\text{distance}^2$ factor) and exposure time (to ensure that pixel intensity is linear with exposure).

2.5. Deep learning methods

The classical image processing described above failed to meet clinical requirements (see section 3.1), so an alternative approach based on CNN was implemented. A neural network is a layered structure of connected nodes known as neurons where the strengths of the connections are adjusted to minimise an objective function. In a CNN, one or more of the layers includes a convolution operator which extracts simplified features of the image, meaning that raw pixel data can be processed (see, for example, Shin *et al* 2016). CNNs have become the method of choice for machine learning image analysis particularly for extracting information from noisy images in a wide range of applications including medical imaging (Chen *et al* 2020) and in radiotherapy (Tomori *et al* 2018, Nyflot *et al* 2019).

Here, the training data consisted of pairs of simulated images where each pair consisted of the simulated dose in the polystyrene block and the simulated photograph. We therefore trained the CNN to predict the dose given the measured photograph (figure 4). The algorithm was validated on different images from those used to train the model and throughout, the model was blind to the measured data. The model was implemented by extending a standard ResNet CNN architecture provided by the python library TensorFlow (Abadi *et al* 2016).

In practice, different CNNs were implemented in order to predict PDD, output factor and other variables, but the basic theory and structure of the models were the same as was the underlying Monte Carlo model. Implementing different deep learning models to solve for different parameters might appear counterintuitive, but to train a single CNN to fit for all the unknown parameters risks overfitting because it would need to be very carefully trained and tuned for a specific configuration. It would then not be robust to any unexpected changes that it has not been trained for. In addition, some of the parameters are correlated (for example, increasing the beam width reduces the scintillation/Cherenkov ratio). This means that a combined CNN might learn that correlation rather than learning the mapping between the image and the dose, which is what we want. Having separate models means that the combined method is be more robust to unexpected changes. Tables 1 and 2 summarise details of the structure and performance of the algorithms used to calculate output factor and PDD. The software is available on Gitlab http://gitlab.com/jeremyocampo/scintinets.

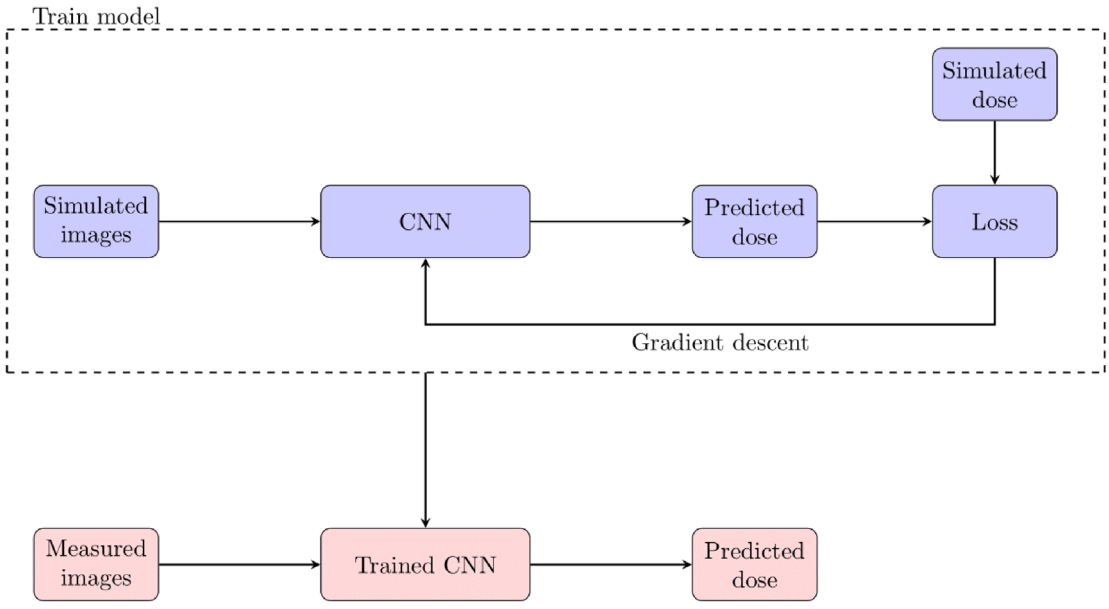

Figure 4. Flow chart showing supervised learning approach.

Table 1. Model performance and size for predicting output factors on the validation dataset. The absolute error between each predicted and conventionally measured output factor was computed, from which the mean absolute difference was calculated over all field sizes.

No. of layers	No. of trainable parameters	Mean absolute difference in output factor
3	5×10^4	5.4 ± 7.2
12	2×10^6	8.4 ± 4.0
24 (ResNet)	4×10^6	1.0 ± 1.2

Table 2. Model performance and size for predicting PDD on the validation dataset. The absolute error between each predicted and conventionally measured PDD was computed, from which the mean absolute difference was calculated over all field sizes.

No. of layers	No. of trainable parameters	Mean absolute difference in PDD (%)
3	7×10^4	2.1 ± 3.0
11 (Autoencoder)	4×10^6	0.68 ± 0.65

One well-known disadvantage of deep learning methods is that they require a large training dataset (typically 10^4–10^6 images). We do not have enough measured data to train a CNN, so instead we simulated a dataset using the GEANT4 model described in section 2.3. To generate many images, the simulation parameters were varied randomly within realistic ranges using a technique called domain randomisation (Tobin *et al* 2017). We chose to randomly vary the beam energy, camera position, scintillator thickness, SSD, beam direction, beam width and intensity, as shown in table 3. This provided 1546 training images and 386 other images that were used for validation. Additional training images were created by randomly rotating, blurring, cropping and adding noise to these images. The scintillation to Cherenkov ratio was also varied. This second round of adjustments could be applied much more quickly than the simulations, allowing a large number of training images to be built up efficiently. In total, 75 000 images were used to train the algorithm that was used to calculate PDD and 90 000 for the output factor. Full details of the development of the algorithm can be found in Ocampo (2023).

2.6. Validation of deep learning algorithm

One of the criticisms of deep learning in safety-critical applications such as radiotherapy is that it is not always easy to have confidence in the results, given that it is hard to understand how decisions are made. In an attempt

Table 3. Domain randomisation parameters for the output factor and PDD models. Mean and standard deviation are shown for normally distributed parameters; mean, maximum and minimum are shown for parameters drawn from a uniform distribution. The peak energy for the PDD model was drawn from a log-uniform distribution between $6 \times 10^{-1.5}$ and $6 \times 10^{1.5}$ MV (equivalent to 0.2–190 MeV). A log-uniform distribution ensures that the selection of energies is distributed evenly over length scales while the broad range of energies was required to provide the CNN with data with sufficient variation. With a narrower range, the PDDs were too similar for the algorithm to learn the differences.

Domain randomisation for output factor				
	Mean	Standard deviation	Minimum	Maximum
Beam width (mm)	25.8		4	62
Energy (MeV)	6.0	1.50		
Lens to beam distance (cm)	52.1		32	70
Lens depth offset (mm)	100.1		92	108
Lens lateral offset (mm)	1.0		−10	10
Sheet thickness (mm)	2.7		1	6.5
SSD (cm)	78.4	1.02		
Domain Randomisation for Percentage Depth Dose				
	Mean	Standard deviation	Minimum	Maximum
Beam width (mm)	27.3		5	60
Energy (MeV)	14.9		2	150
Lens to beam distance (cm)	49.7		30	70
Lens depth offset (mm)	100.0		90	110
Lens lateral offset (mm)	−0.2		−10	10
Sheet thickness (mm)	3.7		1	6.5
SSD (cm)	78.5	0.97		

to mitigate this concern, the model was used to predict some parameters that are known *a priori*, such as the peak energy (which is fixed at 6 MV in the CyberKnife), the beam width (which is determined by the fixed collimator diameters) and the scintillation to Cherenkov ratio (which can be measured). A correct beam energy will determine the position of maximum dose on the PDD curve, and also the ratio of dose at two points on the curve (for example the ratio of dose at 20 cm deep to that at 10 cm deep). Beam width at depth, as predicted by the model can be compared against the measured data which is defined by the fixed collimator diameter. As before, to obtain the best performance, slightly different models were used to predict each parameter but the underlying Monte Carlo model and fundamental model properties remained the same.

3. Results

The results presented below are taken from measured data. Results calculated from photographs are compared to measurements taken with a microdiamond detector using standard radiotherapy quality assurance protocols.

3.1. Classical image processing

In all cases, images taken with the green scintillator agreed more closely to the microdiamond measurements than those with the blue scintillator (figure 5). This supports our hypothesis that because Cherenkov light is strongest in the blue, the scintillation-to-Cherenkov ratio is greatest in the green channel. Nevertheless, the MAD between PDDs measured with the green scintillator sheet and the microdiamond detector were generally around 3% for all collimator sizes (MAD was 12%–17% for the blue scintillator sheet), which is inadequate for radiotherapy QA where we need to assess the Gamma criterion to 2%/2 mm or better. A similarly poor agreement was seen for profiles and the output factor curve, motivating the development of the machine learning methods.

This lack of accuracy is likely caused by two main factors. First, the noise is propagated additively when the Cherenkov-only image is subtracted from the Cherenkov-plus-scintillation image, and second light is present in the tails of the profiles in photographs in regions where the measured dose was zero, confirming the presence of scattered light mentioned in section 2.3 and observed by others (Petric *et al* 2006).

3.2. Deep learning methods

After processing photographs using the green scintillator with the CNN, the MAD between PDDs measured with the CNN and those measured with the microdiamond detector was around 1%. Across all collimator widths from 5 to 60 mm, 95% of points met the Gamma 3%/3 mm criterion, 92% met the 2%/2 mm criterion

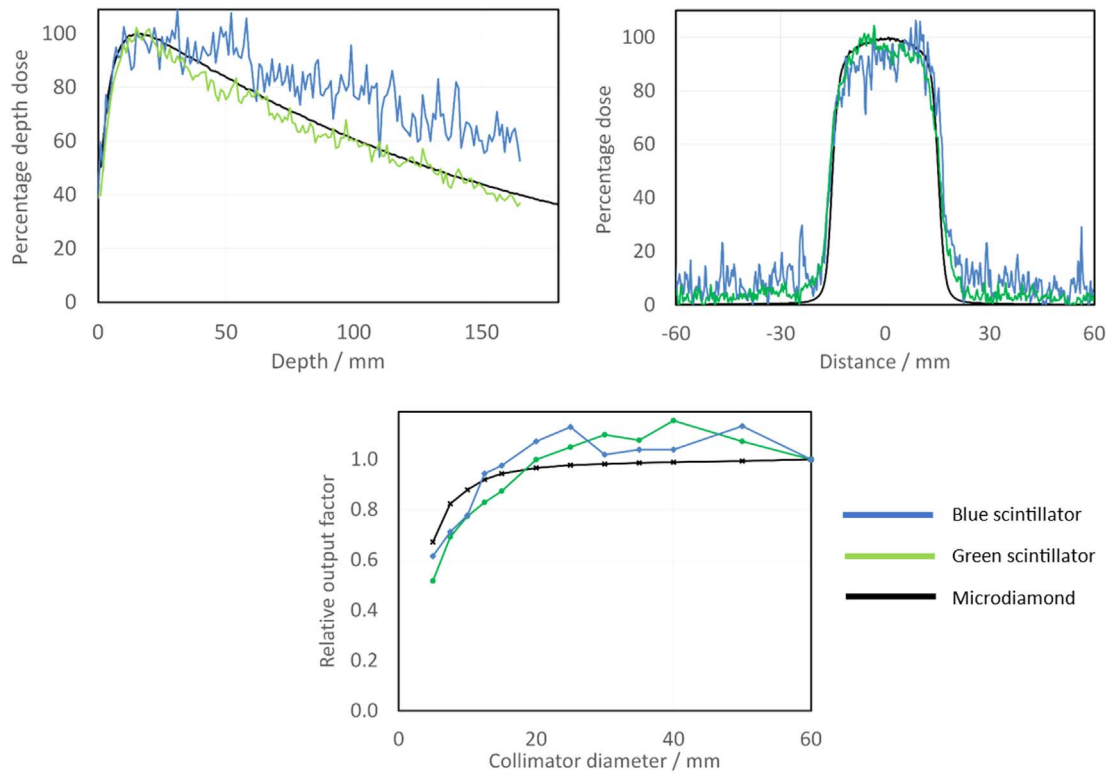

Figure 5. The PDD and profile with the 30 mm collimator measured using the classical approach, and the output factor measurements. Measurements are shown with the blue scintillator (in blue with diamonds), green scintillator (in green with circles) and microdiamond detector (in black with crosses). This demonstrates the noise in the raw measurements especially with the blue scintillator, and the resulting poor determination of output factor.

and 78% met the 1%/1 mm criterion (compared to 41%, 34% and 24% respectively for PDDs determined classically). The PDDs taken with the 20 mm collimator are shown in figure 6.

The improvement in calculation of output factors is shown in figure 7 which plots the output factor calculated using the microdiamond detector and the CNN. The mean percentage difference between the CNN and measurements is 1.1% (for the green scintillator) and 3.4% (for the blue scintillator) compared to 9% and 7% calculated classically. The deviation of 1.1% is in line with the ~2% difference in measurements made with different solid state detectors reported by Chalkley and Heyes (2014). Furthermore, for most radiotherapy applications, the accepted value should be within 2% for all field sizes (IPEM 2018), which we have achieved except for the 5 mm collimator where the difference is 2.5%.

3.3. Validation of deep learning algorithm

The CNN was validated by predicting energy, scintillation to Cherenkov ratio and beam width. The mean predicted energy was (5.85 ± 1.43) MV compared to the true energy of 6.00 MV. The scintillation ratio was significantly underestimated by 0.201 ± 0.167 (on a scale of 0–1) but the beam width was successfully predicted with a MAD between the prediction and the actual collimator width of 0.92 ± 0.50 mm, or a mean percentage difference of 5.3%.

Figure 8 shows the distributions of the predicted variables, along with those for PDD and output factor for completeness. In almost all cases, the mean of the prediction agrees well with the model. The exception is the prediction of beam width which tends to be overestimated. This is likely to be because the training data was based on the beam widths available with the Cyberknife system and included relatively more data from small beam widths than large beams widths.

4. Discussion

Measurements of PDD, profile and output factor following analysis by CNN suggest that this device could be used for relative dosimetry. The results with the smallest (5 mm) and largest (60 mm) collimator were least

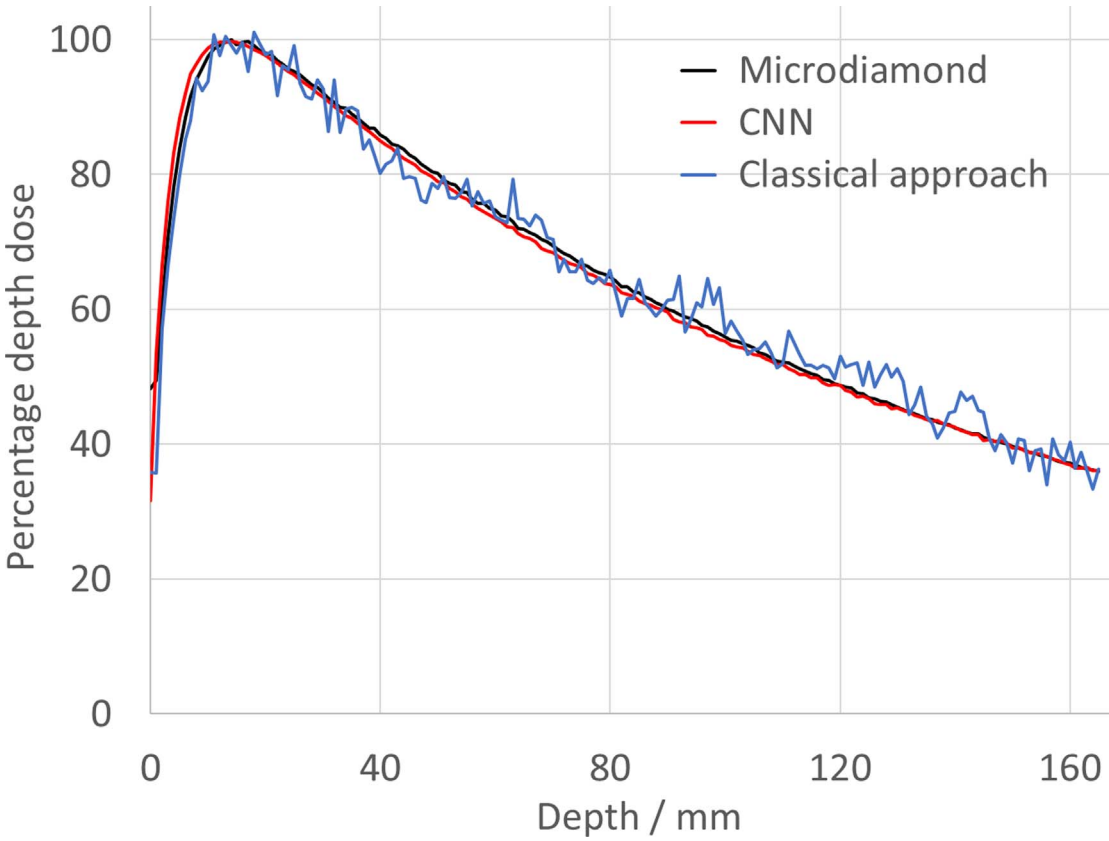

Figure 6. Comparison of PDDs with the 20 mm collimator measured using the microdiamond detector (in black), direct measurements taken from the photograph using the classical approach (in blue) and the prediction from the CNN algorithm (in red).

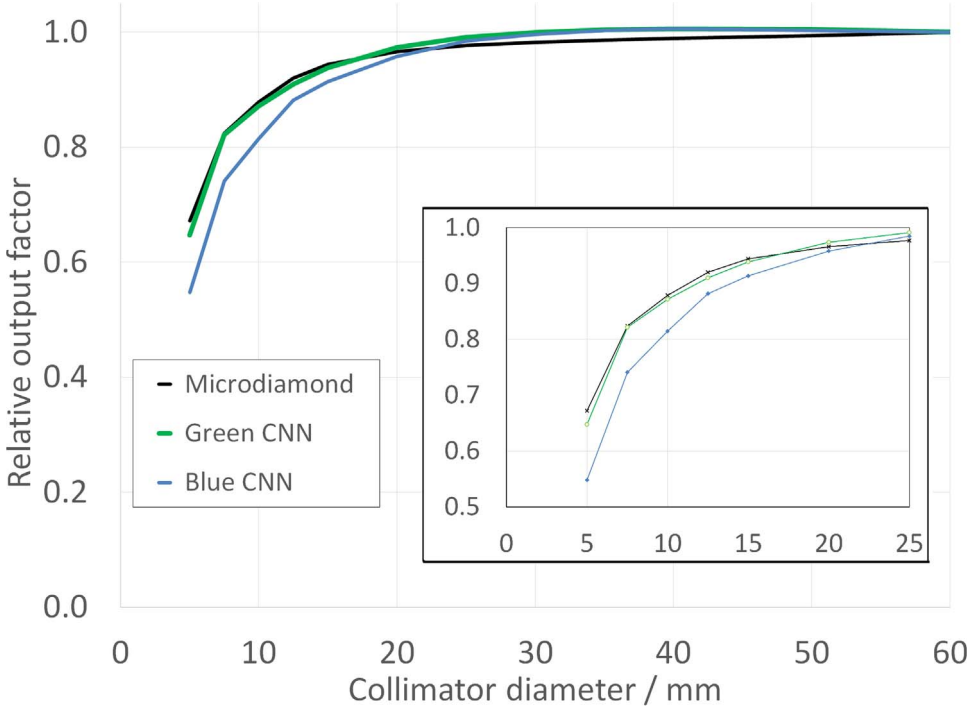

Figure 7. Output factors calculated with the CNN approach. Measurements are shown with the blue scintillator (in blue), green scintillator (in green) and microdiamond detector (in black). Insert shows enlargement of collimator sizes <25 mm.

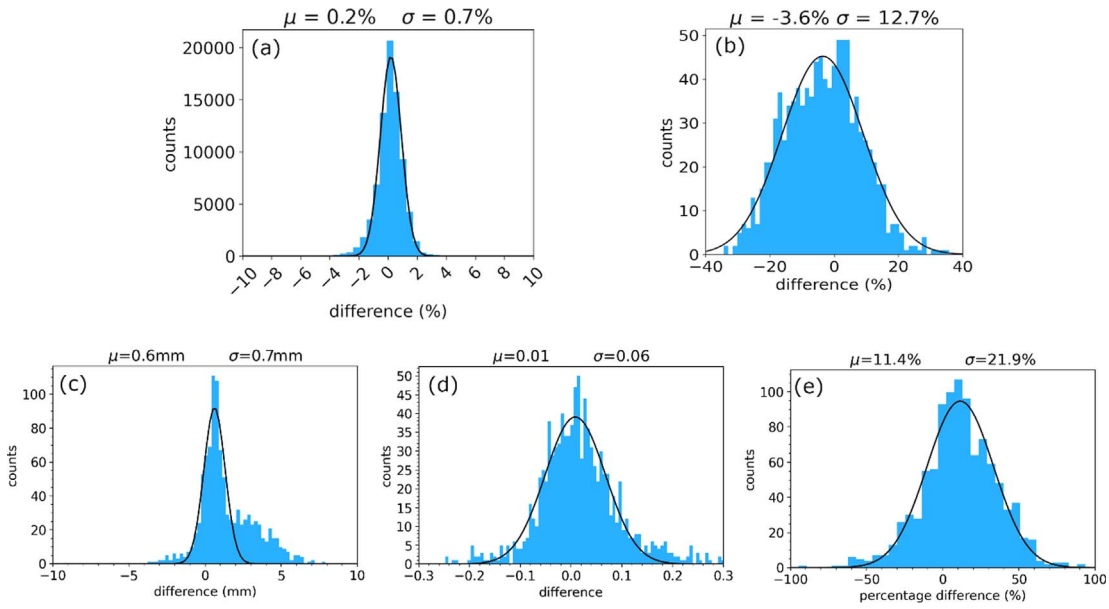

Figure 8. Distributions of the predictions from the CNN models. (a) is the difference between the measured and predicted dose prediction for PDD across all beam sizes. (b) is the difference between the modelled and predicted output factors for the 5 mm beam size (where the agreement between CNN and measurements was poorest). (c) is the difference between modelled and predicted beam width across all beam widths. (d) is the (dimensionless) difference measured and predicted scintillation/Cherenkov ratio across all beam sizes. (e) is the modelled and predicted energies. The mean and standard deviation of the best fit Gaussians are given above each figure.

accurate which we assume is because we did not generate training data outside that range, meaning there is less information for the neural network to interpolate at the extremes.

The scintillation images are noisy, which poses a challenge for classical methods. This is a result of relatively low light levels from the thin scintillator sheet, scattered optical radiation from within the scintillator and secondary x-rays scattered from the scintillator to the camera. When subtracting the Cherenkov-only photograph from the Cherenkov + scintillation photograph, the noise propagates additively, meaning that the images remain noisy even after combining five images and carrying out median filtering.

Noise could be reduced experimentally, for example by using an image intensified camera, or by shielding the camera and viewing the scintillator through a mirror. There are also improvements to be made to the polystyrene blocks which could be optically coupled to each other using a gel or similar. However, we were motivated to build a practical, light, low-cost system and all of these experimental improvements increase complexity. We therefore implemented a deep learning approach which was shown to generate PDDs, profiles and output factors with similar accuracy to those measured using solid state detectors.

The device is still a prototype and collimators need to be changed manually. If the process was automated, a full set of photographs from which an output factor curve could be obtained would take less than an hour to acquire.

Additional confidence in the performance of the CNN was offered by predicting known parameters. The beam energy was predicted reasonably successfully (predicted 5.85 MV compared to the actual energy of 6 MV). The scintillation ratio was underestimated especially at small beam widths where the volume emitting scintillation light and that emitting Cherenkov light become similar. The beam width was predicted with a mean absolute percentage difference of 5.3% which is sufficient to give some confidence in the outputs of the CNN. These discrepancies are likely influenced by the choice of domain randomisation. The purpose of this work was to model PDD and output factor and we chose the range of random variation in the simulations accordingly. For example, we randomly sampled energies from 6×10^{s} MV where s is uniformly sampled from $[-1.5, 1.5]$. This gives a wide range, allowing the algorithm to learn from widely differing PDDs but does not allow good energy resolution around the 6 MV energy of interest. This general approach—of predicting parameters that are known in advance to give confidence in the output of machine learning algorithms—may be more widely applicable in safety-critical applications such as medical imaging and radiotherapy.

5. Conclusion

We have demonstrated that machine learning methods in radiotherapy dosimetry are able to offer quantitative predictions of output factor in small-field dosimetry when classical methods fail. We are aware that it is often

challenging to translate machine learning based methods into clinical practice but, by predicting additional parameters, we believe that we can offer indirect evidence for the quality and robustness of the algorithm which give additional confidence that the method is robust and reliable.

In the introduction, we presented some characteristics of an ideal dosimeter for CyberKnife. Here, we review those characteristics for the device described here:

- **Water equivalence.** The polystyrene blocks have density 1.03 g cm^{-3} and refractive index 1.57. Polystyrene's formula is $(C_8H_8)_n$. According to the NIST tables (NIST 2004), the mass attenuation coefficient of polystyrene at 6 MV is 0.0262 cm^2 g^{-1} while that of water is 0.0277 cm^2 g^{-1}.
- **Spatial resolution**. The pixel resolution was 90 μm which is sufficient to plot the steep descent of the penumbra.
- **Linear response**. Both dose and dose rate showed linear responses ($R^2 > 0.999$).
- **Multiple points**. The system can detect from multiple points simultaneously so that a PDD or profile can be obtained in a single shot.
- **Automation**. In principle, the system could be automated so that multiple collimators can be assessed efficiently
- **Gamma index.** Using our CNN models, 95% of points met a Gamma index of 3 mm/3%, 92% met 2 mm/2% and 78% met 1 mm/1% over all collimator sizes when comparing model predictions to PDD measurements.
- **Ease of use**. As a prototype, the current system is not easy to use, but with further development, it could be mostly automated so that a full output factor curve could be obtained in less than an hour.

Overall, we believe that with some development of hardware and software to aid usability, this system meets most of these requirements and could offer an alternative approach for small-field relative dosimetry.

We have proposed a method for determining PDD, profile and output factor for CyberKnife that is faster, lower cost and easier to use than existing methods and that appears to have equivalent accuracy. We envisage this being used for regular, relative QA to assess whether the linear accelerator's performance has changed since the previous QA test. We do not propose that this replaces absolute measurements with a water tank, but such a system may allow for more frequent QA and more rapid return to clinical use following a service. Further development would allow a movie to be recorded of the time-varying treatment field, potentially allowing verification of the complex four-dimensional treatment plan.

Acknowledgments

JO was funded by the STFC UCL Centre for Doctoral Training in Data Intensive Science (grant no. ST/P006736/1) and the UKRI COVID-19 Grant Extension Allocation. TS is funded by a grant from the Royal Society.

We would like to thank the CyberKnife and radiotherapy physics dosimetry team at University Hospitals Birmingham NHS Foundation Trust, Derek Attree and Connor Godden at UCL Physics and Astronomy for constructing the pelicase and Dr Edward Edmondson for help with managing the HEP computing clusters.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: http://gitlab.com/jeremyocampo/scintinets.

ORCID iDs

Simon Jolly https://orcid.org/0000-0002-2713-0732
Adam Gibson https://orcid.org/0000-0001-9650-5983

References

Abadi M *et al* 2016 TensorFlow: large-scale machine learning on heterogeneous distributed systems arXiv:1603.04467
Allison J *et al* 2016 Recent developments in Geant4 *Nucl. Instrum. Methods Phys. Res.*A **835** 186–225

Almurayshid M, Helo Y, Kacperek A, Griffiths J, Hebden J and Gibson A 2017 Quality assurance in proton beam therapy using a plastic scintillator and a commercially available digital camera *J. Appl. Clin. Med. Phys.* **18** 210–9
Beaulieu L and Beddar S 2016 Review of plastic and liquid scintillation dosimetry for photon, electron, and proton therapy *Phys. Med. Biol.* **61** R305
Beddar A S 2006 Plastic scintillation dosimetry and its application to radiotherapy *Radiat. Meas.* **41** S124–33
Beddar A S, Mackie T R and Attix F H 1992 Cerenkov light generated in optical fibres and other light pipes irradiated by electron beams *Phys. Med. Biol.* **37** 925
Chalkley A and Heyes G 2014 Evaluation of a synthetic single-crystal diamond detector for relative dosimetry measurements on a CyberKnife^{TM} *Br. J. Radiol.* **87** 20130768
Chen X, Men K, Chen B, Tang Y, Zhang T, Wang S, Li Y and Dai J 2020 CNN-based quality assurance for automatic segmentation of breast cancer in radiotherapy *Front. Oncol.* **10** 524
Eaton D J, Bass G, Booker P, Byrne J, Duane S, Frame J, Grattan M, Thomas R A, Thorp N and Nisbet A 2020 IPEM code of practice for high-energy photon therapy dosimetry based on the NPL absorbed dose calibration service *Phys. Med. Biol.* **65** 195006
Frelin A M, Fontbonne J M, Ban G, Colin J, Labalme M, Batalla A, Isambert A, Vela A and Leroux T 2005 Spectral discrimination of Cerenkov radiation in scintillating dosimeters *Med. Phys.* **32** 3000–6
Frelin A M, Fontbonne J M, Ban G, Colin J, Labalme M, Batalla A, Vela A, Boher P, Braud M and Leroux T 2008 The dosimap, a new 2D scintillating dosimeter for IMRT quality assurance: characterization of two Cerenkov discrimination methods *Med. Phys.* **35** 1651–62
Hamel M, Pjatkan R and Burešová H 2020 From the R&D to the commercialization of a new green-emitting plastic scintillator *Nucl. Instrum. Methods Phys. Res.*A **955** 163294
Helo Y, Rosenberg I, D'Souza D, MacDonald L, Speller R, Royle G and Gibson A 2014 Imaging Cerenkov emission as a quality assurance tool in electron radiotherapy *Phys. Med. Biol.* **59** 1963
Hussein M, Clark C H and Nisbet A 2017 Challenges in calculation of the gamma index in radiotherapy – Towards good practice *Phys. Med.* **36** 1–11
IPEM 2018 *Report 81 2nd Edition Physics Aspects of Quality Control in Radiotherapy* https://ipem.ac.uk/resources/books/report-81-2nd-edition-physics-aspects-of-quality-control-in-radiotherapy/
Kam W, Martyn M, Olusoji O J, Woulfe P and O'Keeffe S 2022 Plastic scintillation fibre dosimetry for monitoring of external beam radiation therapy *Optical Sensors and Sensing Congress*STu4C.2
Low D A and Dempsey J F 2003 Evaluation of the gamma dose distribution comparison method *Med. Phys.* **30** 2455–2464
NIST 2004 X-ray mass attenuation coefficients *NIST* https://nist.gov/pml/x-ray-mass-attenuation-coefficients
Nyflot M J, Thammasorn P, Wootton L S, Ford E C and Chaovalitwongse W A 2019 Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks *Med. Phys.* **46** 456–64
Ocampo J 2023 *Designing Convolutional Neural Networks for Scintillation Photography and General Applications (UCL)*
Petric M P, Robar J L and Clark B G 2006 Development and characterization of a tissue equivalent plastic scintillator based dosimetry system *Med. Phys.* **33** 96–105
Pogue B W, Zhang R, Glaser A, Andreozzi J M, Bruza P, Gladstone D J and Jarvis L A 2017 Cherenkov imaging in the potential roles of radiotherapy QA and delivery *J. Phys. Conf. Ser.* **847** 012046
Roussakis Y, Zhang R, Heyes G, Webster G, Mason S, Green S, Pogue B and Dehghani H 2015 Real-time cherenkov emission portal imaging during cyberknife ® radiotherapy *Phys. Med. Biol.* **60** N419
Shin H-C, Roth H R, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D and Summers R M 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning *IEEE Trans. Med. Imaging* **35** 1285–98
Tobin J, Fong R, Ray A, Schneider J, Zaremba W and Abbeel P 2017 Domain randomization for transferring deep neural networks from simulation to the real world *2017 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* pp 23–30
Tomori S, Kadoya N, Takayama Y, Kajikawa T, Shima K, Narazaki K and Jingu K 2018 A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance *Med. Phys.* **45** 4055–65
Yogo K, Tatsuno Y, Tsuneda M, Aono Y, Mochizuki D, Fujisawa Y, Matsushita A, Ishigami M, Ishiyama H and Hayakawa K 2017 Practical use of a plastic scintillator for quality assurance of electron beam therapy *Phys. Med. Biol.* **62** 4551