

Development of a deep learning-based
patient-specific target contour
prediction model for markerless tumor
positioning

（マーカーレス腫瘍位置決めを目的とした深
層学習に基づく患者固有標的輪郭予測モデル
の開発）

周 徳軍

RESEARCH ARTICLE

MEDICAL PHYSICS

Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning

Dejun Zhou¹ | Mitsuhiro Nakamura^{1,2} | Nobutaka Mukumoto² |
Michio Yoshimura² | Takashi Mizowaki²

¹ Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto, Japan

² Department of Radiation Oncology and Image-Applied Therapy, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto, Japan

Correspondence

Mitsuhiro Nakamura, Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.
Email: m_nkmr@kuhp.kyoto-u.ac.jp

Abstract

Purpose: For pancreatic cancer patients, image guided radiation therapy and real-time tumor tracking (RTTT) techniques can deliver radiation to the target accurately. Currently, for the radiation therapy machine with kV X-ray imaging systems, internal markers must be implemented to facilitate tumor tracking. The purpose of this study was to develop a markerless deep learning-based pancreatic tumor positioning procedure for real-time tumor tracking with a kV X-ray imaging system.

Methods and materials: Fourteen pancreatic cancer patients treated with intensity-modulated radiation therapy from six fixed gantry angles with a gimbal-head radiotherapy system were included in this study. For a gimbal-head radiotherapy system, the three-dimensional (3D) intrafraction target position can be determined using an orthogonal kV X-ray imaging system. All patients underwent four-dimensional computed tomography (4DCT) simulations for treatment planning, which were divided into 10 respiratory phases. After a patient's 4DCT was acquired, for each X-ray tube angle, 10 digitally reconstructed radiograph (DRR) images were obtained. Then, a data augmentation procedure was conducted. The data augmentation procedure first rotated the CT volume around the superior–inferior and anterior–posterior directions from -3° to 3° in 1.5° intervals. Then, the Super-SloMo model was adapted to interpolate 10 frames between respiratory phases. In total, the data augmentation procedure expanded the data scale 250-fold. In this study, for each patient, 12 datasets containing the DRR images from each specific X-ray tube angle based on the radiation therapy plan were obtained. The augmented dataset was randomly divided into training and testing datasets. The training dataset contained 2000 DRR images with clinical target volume (CTV) contours labeled for fine-tuning the pre-trained target contour prediction model. After the fine-tuning, the patient and X-ray tube angle-specific CTV contour prediction model was acquired. The testing dataset contained the remaining 500 images to evaluate the performance of the CTV contour prediction model. The dice similarity coefficient (DSC) between the area enclosed by the CTV contour and predicted contour was calculated to evaluate the model's contour prediction performance. The 3D position of the CTV was calculated based on the centroid of the contour in the orthogonal DRR images, and the 3D error of the prediction position was calculated to evaluate the CTV positioning performance. For each patient, the DSC results from 12 X-ray tube angles and 3D error from 6 gantry angles were calculated, representing the novelty of this study.

Results: The mean and standard deviation (SD) of all patients' DSCs were 0.98 and 0.015, respectively. The mean and SD of the 3D error were 0.29 mm and 0.14 mm, respectively. The global maximum 3D error was 1.66 mm, and the global minimum DSC was 0.81. The mean calculation time for CTV contour prediction was 55 ms per image. This fulfills the requirement of RTTT.

Conclusions: Regarding the positioning accuracy and calculation efficiency, the presented procedure can provide a solution for markerless real-time tumor tracking for pancreatic cancer patients.

KEYWORDS

deep learning, markerless tumor positioning, pancreatic cancer, real time tumor tracking, target contour prediction

1 | INTRODUCTION

Pancreatic cancer is one of the leading causes of cancer death worldwide.¹ Although surgical resection is recommended for complete eradication, most pancreatic cancer patients have unresectable disease states at diagnosis. Thus, radiotherapy combined with systemic chemotherapy has become the standard treatment.² To deliver the radiation dose to the target accurately, real-time tumor tracking (RTTT) is a useful approach.³

Multiple techniques have been adapted to track the motion of tumors, including kV X-ray imaging,⁴ magnetic resonance imaging,⁵ and ultrasound imaging.⁶ Currently, in clinical practice, the kV X-ray imaging technique is the most widely used. Because of the physical characteristics of kV X-ray, the contrast of soft tissue is low in kV X-ray projection images. Currently, to track the motion of soft targets accurately, internal markers, such as metallic fiducials, are implanted into the tumor or its surrounding tissue before a planning computed tomography (pCT) scan.⁷ In pCT, the geometric correlation between the internal markers and tumor is identified. During treatment, by detecting the positions of internal markers, the position of the target can be calculated using the geometric correlation determined previously. In clinical practice, the use of internal markers has proven to be effective, but some areas for improvement have also been highlighted. First, the implementation of internal markers is invasive. It prolongs the treatment period and causes risk to the patient, such as tumor bleeding. Second, the geometric correlation between the markers and target may change between the pCT scan and time of treatment.⁸ During treatment, the geometric correlation between the markers and target may also shift with respiration. Such marker-induced error causes a decrease in the accuracy of tumor tracking. Third, the markers may cause metal artifacts in the pCT, degrading the image quality.⁹ To overcome these limitations, a solution that realizes markerless tumor positioning is required. The solution should have the ability to directly predict the position of the soft tissue target in kV X-ray projection images without the assistance of markers.

Deep learning is one possible solution. Deep learning is a flourishing area in the field of medical physics.¹⁰ The application of deep learning-based models has improved in the field of radiation oncology treatment, and the improvement of graphics processing units (GPUs) ensures that deep learning models are capable of real-time tumor tracking.

Some researchers have aimed to adapt deep learning to markerless target localization. Zhao et al. researched deep learning-based markerless tumor localization in the pancreas¹¹ and prostate.¹² Their model is based on a region proposal network¹³ and predicts the bounding box of the tumor on digitally reconstructed radiograph (DRR) images. They calculated the mean absolute difference (MAD) between the ground truth and predicted top-left corner of the bounding box to evaluate the localization performance. In their work, two-dimensional (2D) localization was not sufficient for tumor tracking, and the appropriateness of localization with the top-left corner of the bounding box to evaluate performance is worth discussing. In contrast, several investigators developed a three-dimensional (3D) localization model.^{14–16} Their model is based on a fully convolutional neural network¹⁷ and predicts the target position based on the center of the target probability map. By predicting the target in images from two-view X-ray fluoroscopy images, the 3D position could be determined. However, the target was either spherical or ovoid, which is far from a clinical situation.

This paper presents a deep learning-based procedure consisting of data augmentation and training of a patient-specific target contour prediction model for markerless tumor positioning. The data augmentation method and contour prediction model are patient- and kV X-ray tube angle-specific. No historical data are required, and the procedure is independent of institution protocol. Beyond the bounding box and simple shape, the clinical target volume (CTV) contour of the pancreatic tumor is predicted from DRR images, and the 3D position of the tumor is calculated based on the centroid of the CTV contour. The performances of CTV contour and 3D positioning prediction were analyzed and evaluated.

TABLE 1 Patient characteristics

Patient ID	Age (y)	Sex	Stage	CTV size (cm ³)	3D motion range (mm)
1	59	F	cT4N0M0	140.69	13
2	64	F	cT4N0M0	61.07	14
3	59	M	cT4N0M0	87.71	10
4	64	M	cT4N1M0	105.55	15
5	69	F	cT4N0M0	66.15	11
6	73	M	cT4N0M0	107.79	11
7	62	M	cT4N0M0	184.9	16
8	76	F	cT3N0M0	61.92	10
9	75	F	cT3N0M0	91.63	10
10	72	M	cT4N0M0	114.43	10
11	74	M	cT4N0M0	153.28	11
12	73	M	cT4N0M0	133.9	10
13	67	F	cT4N1M0	102.18	13
14	67	F	cT4N0M0	146.56	10

2 | MATERIALS AND METHODS

2.1 | Vero4DRT system

The Vero4DRT system has a gimbaled 6 MV X-ray head and an orthogonal kV X-ray imaging system.¹⁸ The gimbaled head and imaging system were installed with a compact accelerator in an O-ring gantry. The O-ring gantry can rotate $\pm 60^\circ$ about its vertical axis and $\pm 185^\circ$ along its O-shaped structure. The orthogonal kV X-ray imaging system consists of two pairs of X-ray tubes and flat panel detectors (FPDs). The X-ray source to isocenter distance and FPD distance are 1000 and 1836 mm, respectively. The orthogonal kV X-ray imaging system rotates simultaneously with the gimbaled head, and the orthogonal X-ray tube angles are $\pm 45^\circ$ from the axis of the gimbaled head. The maximum irradiation field of FPDs at isocenter level is 222 mm $\times$ 168 mm, with a spatial resolution of 0.211 mm at the isocenter level. The pixel array size of an FPD is 1024 $\times$ 768. During treatment, the imaging system can continuously take orthogonal X-ray projection images.

2.2 | Patient characteristics and treatment planning

4DCT scans of 14 patients with locally advanced pancreatic cancer were included in this study (Table 1). CT simulations were performed using the BodyFix system (Elekta AB, Stockholm, Sweden) with arms raised overhead. The 4DCT scans were acquired from the LightSpeed RT 16-slice CT simulator (General Electric Medical Systems, Waukesha, WI, USA) or SOMATOM Definition AS (Siemens Medical Systems,

Erlangen, Germany) with a real-time positioning management system (Varian Medical System, Palo Alto, CA, USA). The entire respiratory period was divided into 10 respiratory phases.

The gross tumor volume (GTV) included the tumor and metastatic lymph nodes. The CTV included the GTV plus a 5 mm margin, the retropancreatic space, and the paraaortic lymph nodes between the celiac axis and superior mesenteric artery. The radiation oncologist contoured the GTV and CTV for each phase.

All patients underwent six-port coplanar static intensity-modulated radiation therapy (IMRT) with RTTT. The gantry angles were 80° , 120° , 160° , 200° , 240° , and 280° . Correspondingly, the clockwise and counterclockwise X-ray tube angles were the gantry angles plus and minus 45° , respectively. The prescribed dose was 48 Gy in 15 fractions.

2.3 | Data augmentation procedure for labeled dataset generation

A large labeled dataset was required to both train and evaluate the performance of the deep learning model. This section presents a data augmentation method aimed at generating sufficient labeled data from the original data from one X-ray tube angle of a patient.

The starting point of the data augmentation was a patient's 4DCT, which contained the CT volumes of 10 respiratory phases. To obtain the ground truth of the CTV contour from the DRR image, CTV-only CT volumes were generated by an in-house program. The program extracted CTV contours from the DICOM structure storage file and set the voxel value outside the CTV contours to zero. Then, the open-source program Plastimatch was applied to generate DRR images for both the original and corresponding CTV-only CT volumes. The geometry settings for DRR generation were taken from the Vero4DRT system. As a result, 10 paired DRR and CTV-only DRR images corresponding to each respiratory phase were acquired.

The first step for augmentation was to tilt the gantry angle and ring angle from -3° to 3° in 1.5° intervals for DRR image generation. This has the same effect as the CT volumes being rotated around the superior-inferior and anterior-posterior directions. As the gantry angle and ring angle were tilted five times for each respiratory phase, the data scale was augmented 25-fold after this step. After this step, 250 paired DRR and CTV-only DRR images were acquired.

The second step was to use the Super-SloMo model to interpolate frames between the respiratory phases. The Super-SloMo model was proposed by Nvidia, and the details of the model were introduced in a previous paper.¹⁹ In this work, the Super-SloMo model trained on the Adobe 240-fps dataset was adapted to interpolate images between the original DRR images

and CTV-only DRR images of consecutive respiratory phases. In this step, to ensure that the CTV contours were still valid after interpolation, it was important to ensure that the interpolated motion of the CTV-only DRR images was the same as that of the original DRR images. Here, the hyperparameters acquired based on the original DRR images were directly used to interpolate the CTV-only DRR images. With the same set of hyperparameters, the pixels in the original DRR images and CTV-only DRR images moved simultaneously. Nine frames were interpolated between consecutive phases, producing a 10-fold augmentation. Thus, a total of 2500 paired DRR images and CTV-only DRR images were obtained for one X-ray tube angle of a patient. This data augmentation range and scale were considered to be appropriate by referring to previous work by Zhao et al.¹¹

By extracting the mask of the CTV from the CTV-only DRR image and overlapping it with the corresponding original DRR image, labeled DRR images were obtained. Subsequently, the 2500 labeled DRR images were randomly divided into two subsets: a training dataset of 2000 images to fine-tune a pre-trained target contour prediction model and a testing dataset of 500 images to evaluate the performance of the patient-specific CTV contour prediction model. The same dataset splitting approach was used in previous works for similar purposes.^{11,12,16}

2.4 | Patient-specific deep learning model for target contour prediction

For each patient and each X-ray tube angle, a specific target contour prediction model was trained and tested with the corresponding DRR image dataset. The workflow for training and testing a patient-specific deep learning model for target contour prediction is shown in Figure 1. The target of this study was the CTV. After data augmentation, a pre-trained target contour prediction model was fine-tuned with the training dataset to predict the CTV contours from the DRR images. The details of the model can be found in a previous paper done by He et al.²⁰ In brief, this model can predict the contour of the target in each region of interest (RoI). In this work, the backbone for feature extraction was ResNet with a depth of 50²¹ and a feature pyramid network (FPN).²² The reason for this choice was based on the consideration of accuracy and calculation time. The model was pre-trained on the COCO dataset.²³ The loss function of the model during training was multitasking. For each RoI, the loss function contained loss on classification, bounding box,²⁴ and contour.²⁰ The training and testing frameworks were both detectron2, provided by Facebook AI Research.²⁵ The base learning rate was set to 0.02. The pre-trained model was fine-tuned for 2000 iterations.

2.5 | Performance evaluation

For one kV X-ray tube angle of a patient, a testing dataset containing 500 DRR images was used to evaluate the performance of the deep learning model.

The dice similarity coefficient (DSC) was introduced to evaluate the performance of the target contour prediction as follows:

$$DSC = \frac{2 |A_{\text{truth}} \cap A_{\text{predict}}|}{A_{\text{truth}} + A_{\text{predict}}}, \quad (1)$$

where A_{truth} and A_{predict} are the areas enclosed by the ground truth contour and predicted contour, respectively. The intersection symbol indicates the area where A_{truth} and A_{predict} overlap.

During the testing, the original DRR images were inputted into the fine-tuned model. The output was DRR images with the predicted CTV contours. Then, the ground truth and predicted CTV centroid were calculated based on the corresponding CTV contours. With the CTV centroids on the orthogonal paired DRR images, the 3D position was defined as the midpoint of the shortest intersection vector between the vectors connecting the orthogonal kV X-ray source and corresponding 2D position on the FPDs. The distance between the 3D position of a labeled CTV and predicted CTV was calculated to evaluate the performance of the model for tumor positioning.

3 | RESULTS

The calculation efficiency of the tumor contouring model was found to be high. The high-performance computer (HPC) used in this study had an Intel Core i7 9800X CPU, four Quadro GV100 32 GB GPUs, and eight 16 GB RAM modules. The training time with one GPU was 2.5 s per iteration. For 2000 iterations, the training time for one X-ray tube angle of a patient was approximately 1.4 h. As the HPC had four GPUs and each patient had 12 X-ray tube angles, the total training time for one patient was approximately 4.2 h. The averaged contour prediction time with one GPU for one image was 55.17 ms. As the orthogonal images can be calculated parallelly on two GPUs, the calculation efficiency is sufficient for real-time tumor tracking.

The predicted contour of CTV and its centroid visually matched the ground truth with high accuracy. Figure 2 shows examples of predicted contours and centroids overlaying the ground truth results from two orthogonal X-ray tube angles of a patient in different respiratory phases equal to end-inhalation, end-exhalation, and middle-respiratory phase. The ground truth contour of a CTV was obtained by overlapping the CTV-only

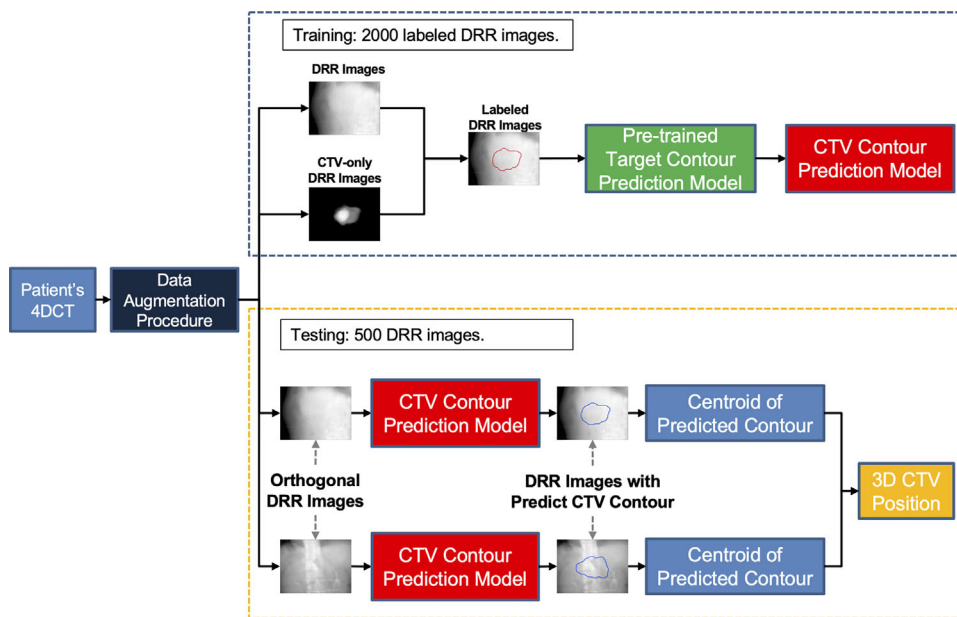


FIGURE 1 Overall workflow of the deep learning-based markerless tumor positioning procedure

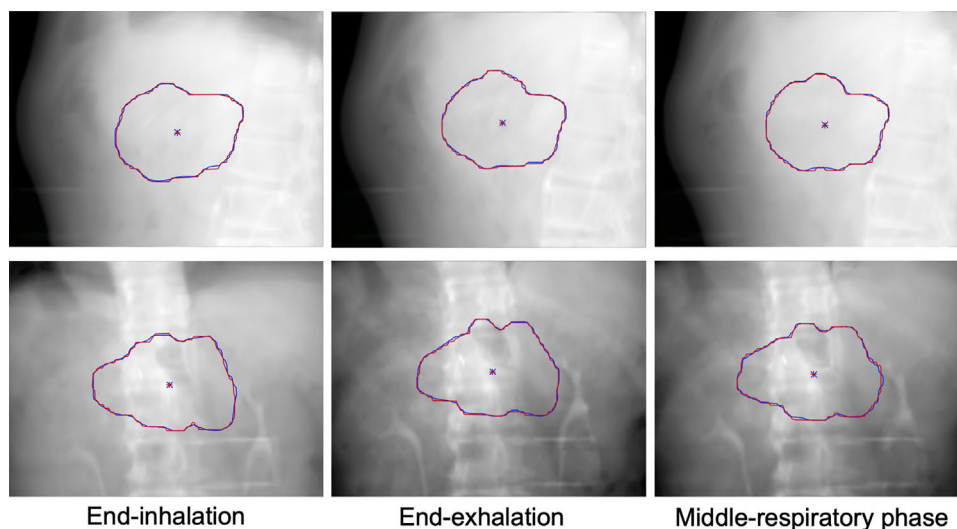


FIGURE 2 Examples of pancreatic tumor CTV contours and centroids predicted by the deep learning model (blue lines and $\times$ symbols) and their corresponding ground truth labels (red lines and $+$ symbols) overlaid on the patient's digitally reconstructed radiographs. The first, second, and third columns represent end-inhalation, end-exhalation, and middle-respiratory phase, respectively. The kV X-ray tube angles for the first and second rows are 115° and 205° , respectively

DRR image with its corresponding DRR image. During the frame interpolation between respiratory phases, the weight parameters were generated by the DRR images, and the same weight parameter set was used to interpolate the frames of CTV-only DRR images to ensure that the motion was identical between the CTV-only DRR images and DRR images.

For one X-ray tube angle of a patient, 500 testing images were imported to evaluate the performance of the model. The present study included 14 patients

treated with six fixed gantry angles, which corresponds to 12 X-ray tube angles. A total of 84 000 results were included in the CTV contouring evaluation, and 42 000 results were included in the 3D positioning evaluation.

DSC was introduced to evaluate the model's contouring performance. The calculation of DSC was based on single DRR images, and the results versus X-ray tube angle and patient were compared. Figure 3 shows the statistical DSC results regarding X-ray tube angle

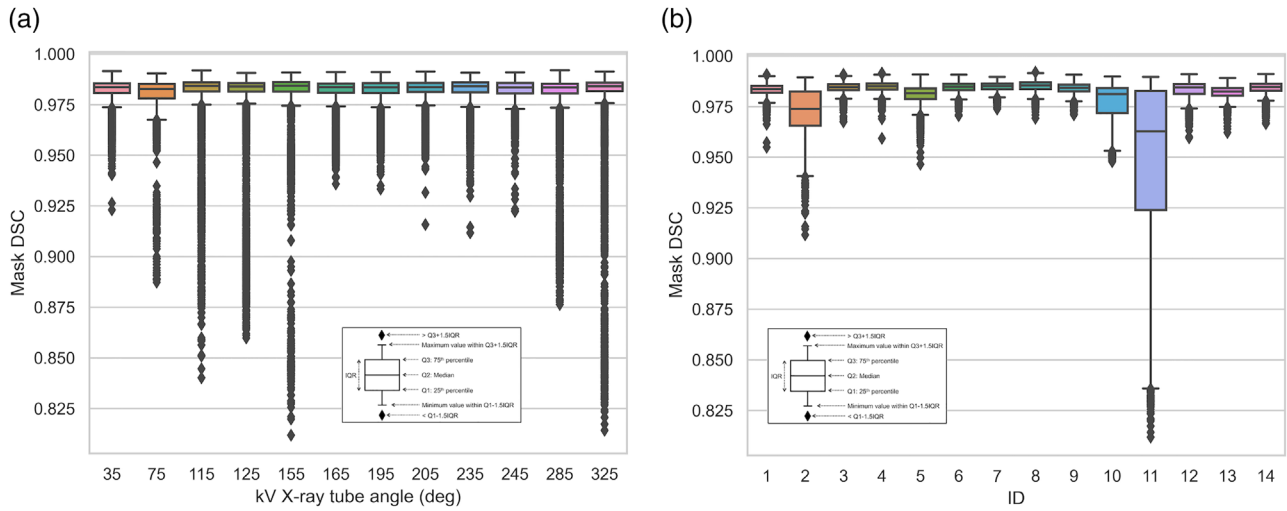


FIGURE 3 DSCs regarding (a) X-ray tube angle and (b) patient. The orthogonal X-ray tube angles were $\pm 45^\circ$ from the gantry angle

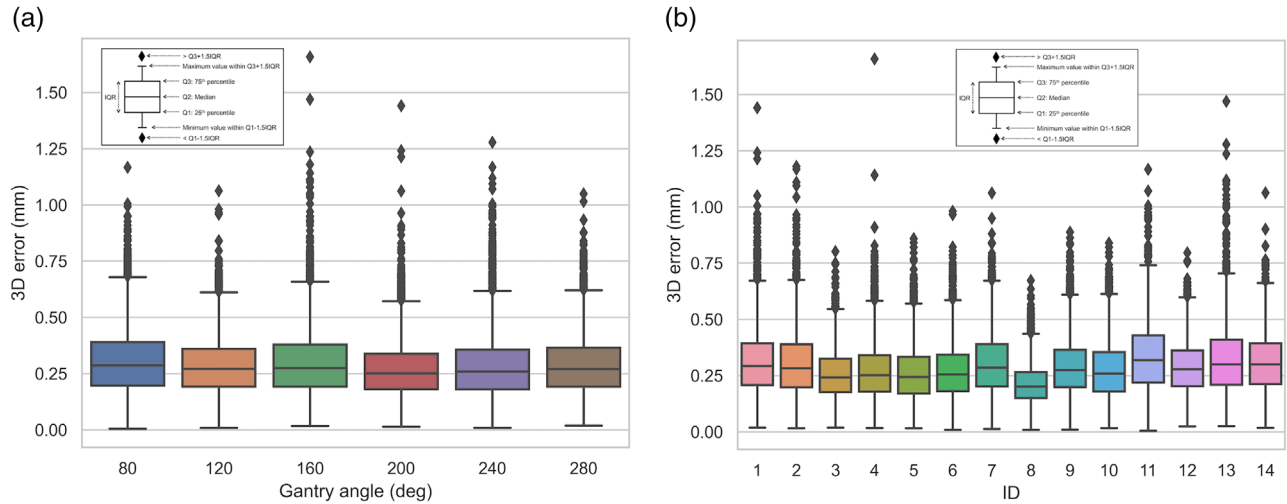


FIGURE 4 3D error (in mm) regarding (a) gantry angle and (b) patient

(Figure 3a) and patient (Figure 3b). The mean and standard deviation (SD) for the 84 000 results were 0.98 and 0.015, respectively. The global minimum DSC was 0.81. In Figure 3b, it is noticeable that for patients nos. 2, 10, and 11, the statistical DSC results are not as good as those for the other patients.

The 3D positioning error was calculated to evaluate the model's target positioning performance. The calculation of 3D position was based on paired orthogonal DRR images, and the 3D positioning error versus gantry angle and patient was compared. Figure 4 shows a box plot of the 3D error regarding gantry angle (Figure 4a) and patient (Figure 4b). For the total of 42 000 3D error results, the mean value and SD were 0.29 mm and 0.14 mm, respectively. The global maximum 3D error was 1.66 mm. Figure 4b shows that for patient no. 8, the deep learning-based procedure achieved the best performance regarding 3D error.

4 | DISCUSSION

To the best of our knowledge, the present study is the first to realize contour prediction of pancreatic tumors from DRR images and to predict 3D tumor position. As the procedure is patient- and kV X-ray tube angle-specific, it is not oncologist or institution dependent. Its application can be easily extended to different tumor sites and institutions.

Clinically, regarding real-time tumor tracking, the accuracy of 3D positioning is the main focus. The results clearly show that the deep learning-based procedure can accurately predict the 3D position of a pancreatic tumor without the assistance of internal markers. The mean $\pm$ SD of 3D error regarding the 42 000 results was 0.29 ± 0.14 mm, and the maximum 3D error was 1.66 mm. The mean 3D error was close to the FPD pixel size at isocenter level, indicating that the accuracy is

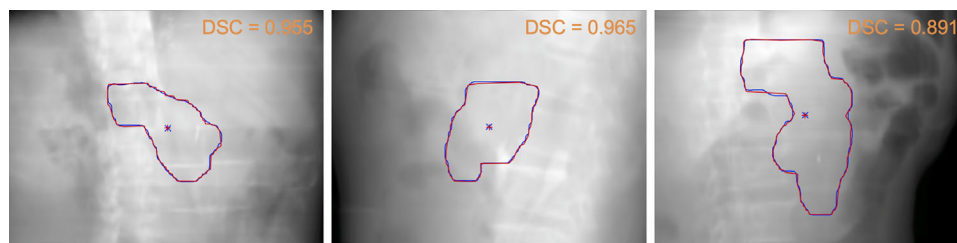


FIGURE 5 Examples of three cases with low DSC. The blue lines and $\times$ symbols represent the predicted CTV contours and centroids, respectively. The red lines and $+$ symbols represent the ground truth labels. The left is from patient no. 2 with X-ray tube angle 205° (DSC is 0.955). The middle is from patient no. 10 with X-ray tube angle 115° (DSC is 0.965). The right is from patient no. 11 with X-ray tube angle 325° (DSC is 0.891)

close to the machine limitation and fulfills clinical practice requirements. Compared with the work of Zhao et al. with a similar tumor site,¹¹ their 2D mean absolute errors were found to be only less than 2.6 mm, which illustrates the novelty of this work.

The box plots of each gantry angle Figure 4a are similar, so it can be concluded that the 3D positioning error was independent of the gantry angle. The second row of Figure 2 demonstrates that the 3D positioning performance was stable, even when the CTV in the DRR images was overlapping with vertical bone in a large area. From Figure 4b, it is noticeable that the 3D error for patient no. 8 is smaller than that for the other patient. This is because among the patients with tumor motion equal to 10 mm, the CTV size of patient no. 8 was the smallest. This indicates that the deep learning-based procedure and model may have better positioning performance on targets with small volume and small motion range.

In the future, the prediction of target contour may assist in treatment, and as such, target contour prediction analysis was also conducted in this study. As shown in Figure 3a, the DSC box plots are similar for different X-ray tube angles. Therefore, the target contour prediction of the deep learning-based procedure was stable for different X-ray tube angles. As shown in Figure 3b, the DSCs for patients nos. 2, 10, and 11 are notably lower compared to those of the other patients. Figure 5 shows the labeled and predicted DRR images of these three patients from the X-ray tube angles with the lowest mean DSC values. These target contours have two common features. The first is asymmetry. It can be seen from the DRR images that the contour of the CTV is far from symmetrical. This asymmetry indicates the existence of sudden changes in the contour, and it is difficult to predict the contour accurately in regions containing sudden changes. The second common feature is smoothness. When the contour is not smooth, it is difficult to predict it accurately. Meanwhile, as shown in Figure 5, although the inaccurate contour prediction would cause a relatively low DSC value, at this level of prediction accuracy, the centroid of prediction and ground truth CTV contour still matched well. This can explain why the DSC value

for patient no. 11 in Figure 3b was lower than the others, but in Figure 4b, the 3D error was not particularly higher compared to that for other patients.

The calculation time per image is critical for clinical implementation. With the assistance of a GPU and the latest deep learning framework, the mean contour prediction time of an image was 55 ms. According to the American Association of Physicists in Medicine Task Group 76,²⁶ to realize RTTT, the total system latency should be less than 500 ms. For the Vero4DRT system, the system lag was determined to be less than 48 ms.²⁷ In addition to the image processing time achieved in this work, the deep learning-based markerless tumor positioning procedure fulfills the requirement for RTTT.

Compared with the previous works,^{11,16} this work shows advantages in the following aspects. First, our work predicted the contour of the tumor and determined the tumor position based on the centroid of the contour. This definition may be more in line with clinical needs. Second, the calculation times of Zhao et al. and Takahashi et al. were less than 200¹¹ and 32.5 ms¹⁶ per image, respectively. Therefore, the work presented here is much quicker than Zhao et al.'s and comparable with Takahashi et al.'s. However, considering that our work can predict the contour of a tumor with a similar calculation time, compared with Takahashi et al.'s work, the calculation time is still superior. The third and most important aspect is the positioning accuracy. For pancreatic tumor positioning, the presented procedure offers the best positioning results.

Currently, a pre-trained target contour model is required to predict target position because less data are required for fine-tuning a pre-trained deep learning model than initially training a deep learning model. According to the results, the presented data augmentation procedure and patient-specific training data scale were sufficient to fine-tune a pre-trained target contour model. Ideally, a deep learning model that predicts 3D position directly from orthogonal images would be better. Meanwhile, to initially train a deep learning model, much more than 2000 images are required. As there is currently no pre-trained model for this purpose, the present

work may contribute to the development of a practical approach in the field.

Although this paper presented a procedure to generate a patient- and X-ray tube angle-specific model that can detect the contour and centroid of a pancreatic tumor CTV from DRR images with high accuracy, two main limitations exist regarding the implementation of this procedure in clinical practice. First, because it is currently impossible to detect and contour pancreatic tumors from real X-ray projection images obtained during radiation therapy, the synthetic DRR images of this study were used for both training and testing the target contour prediction model. Although the synthetic DRR images were realistic, there were still differences from the real ones. During treatment, the MV beam scatters may be detected by the FPDs and generate noise in the X-ray projection images.²⁸ This may decrease the performance of the model. For the X-ray imaging system of Vero4DRT, X-ray projection images are recorded as raw data recorded by the FPDs. Conversion based on contrast, gamma correction, image window level, and window width is required to render X-ray projection images similar to synthetic DRR images. The model performance may also depend on the accuracy of the conversion. In the future, the model performance will be evaluated with real kV X-ray projection images of patients with pancreatic tumors who underwent RTTT treatment. For pancreatic tumors, the marker-implanted RTTT results were recorded and published.²⁹ Once the presented work is expanded and successfully predicts target positions in real kV X-ray images of the recorded patients, the validation of the model performance will be presented by comparing clinical RTTT records and the predicted 3D target positions of the model. That is the next milestone for the authors.

Second, patient anatomy and pancreatic tumor motion may change between 4DCT simulation and the time of treatment.³⁰ Therefore, the procedure must augment the data and train the target contour prediction model using the DRR images generated from patients' 4DCT scans. For the institutional procedure under which this study was conducted, the patient's 4DCT was simulated approximately 7 to 14 days before the treatment started. The patients were treated in 15 fractions within a median of 21 days.⁷ During this period, the interfractional change in tumor shape and surrounding normal tissues will have a negative effect on the model performance. To overcome this limitation, analysis of the performance on real kV X-ray projection images is required. If the decrease in positioning performance is identified as clinically significant, an update of the model based on the DRR images from cone beam computed tomography volume or another 4DCT simulation may be required during the patient's period of treatment. This is within the scope of adaptive radiotherapy.

5 | CONCLUSION

A novel deep learning-based procedure for markerless tumor positioning was presented in this paper. For 14 pancreatic tumor patients treated with six-gantry-angle IMRT, the mean $\pm$ SD of 3D error was 0.29 ± 0.14 mm. The CTV contour prediction time per image was approximately 55 ms. This procedure provides a potential solution to markerless real-time tumor tracking for multiple tumor sites or organs of interest. Future studies should focus on testing and evaluating the performance of the presented procedure using real X-ray projection images. After retrospective studies and validations with radiation oncologists' contours on X-ray projection images, a phantom study will be conducted, and clinical implementation will be considered.

ACKNOWLEDGMENTS

We sincerely appreciate all the staff and members of the Medical Physics Laboratory of Kyoto University Graduate School of Medicine, Department of Information Technology and Medical Engineering, and Human Health Sciences (<http://medicalphysics.hs.med.kyoto-u.ac.jp/>) for their excellent technical support and valuable comments on this study.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.
2. Tempero MA, Arnoletti JP, Behrman SW, et al. Pancreatic adenocarcinoma, version 2.2012. featured updates to the NCCN guidelines. *J Natl Compr Canc Netw*. 2012;10(6):703-713.
3. Matsuo Y, Onishi H, Nakagawa K, et al. Guidelines for respiratory motion management in radiation therapy. *J Radiat Res*. 2013;54(3):561-568.
4. Shirato H, Shimizu S, Kunieda T, et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2000;48(4):1187-1195.
5. Acharya S, Fischer-Valuck BW, Kashani R, et al. Online magnetic resonance image guided adaptive radiation therapy: first clinical applications. *Int J Radiat Oncol Biol Phys*. 2016;94(2):394-403.
6. Omari EA, Erickson B, Ehlers C, et al. Preliminary results on the feasibility of using ultrasound to monitor intrafractional motion during radiation therapy for pancreatic cancer. *Med Phys*. 2016;43(9):5252.
7. Nakamura A, Hiraoka M, Itasaka S, et al. Evaluation of dynamic tumor-tracking intensity-modulated radiotherapy for locally advanced pancreatic cancer. *Sci Rep*. 2018;8(1):17096.
8. Ashida R, Goto Y, Akimoto M, et al. Evaluation of the positional differences between bone matching and fiducial marker matching in target localization of the expiratory breath-hold radiation therapy for pancreatic cancer. *Am J Clin Oncol*. 2018;36(4):517.
9. Brook OR, Gourtsoyianni S, Brook A, Mahadevan A, Wilcox C, Raptopoulos V. Spectral CT with metal artifacts reduction software for improvement of tumor visibility in the vicinity of gold fiducial markers. *Radiol*. 2012;263(3):696-705.

10. Cui S, Tseng HH, Pakela J, Ten Haken RK, El Naqa I. Introduction to machine and deep learning for medical physicists. *Med Phys*. 2020;47(5):e127-e147.
11. Zhao W, Shen L, Han B, et al. Markerless pancreatic tumor target localization enabled by deep learning. *Int J Radiat Oncol Biol Phys*. 2019;105(2):432-439.
12. Zhao W, Han B, Yang Y, et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiother Oncol*. 2019;140:167-174.
13. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6):1137-1149.
14. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Phys Med*. 2019;59:22-29.
15. Mori S. A real-time, single-exposure, dual-energy subtraction mask for markerless tumor tracking in radiotherapy: proof of concept. *Phys Med*. 2019;63:63-69.
16. Takahashi W, Oshikawa S, Mori S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: proof of concept by phantom study. *Br J Radiol*. 2020;93(1109):20190420.
17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2015;7:3431-3440.
18. Hiraoka M, Mizowaki T, Matsuo Y, Nakamura M, Verellen D. The gimbaled-head radiotherapy system: rise and downfall of a dedicated system for dynamic tumor tracking with real-time monitoring and dynamic WaveArc. *Radiother Oncol*. 2020;153:311-318.
19. Jiang HZ, Sun DQ, Jampani V, Yang MH, Learned-Miller E, Kautz J. Super slo-mo: high quality estimation of multiple intermediate frames for video interpolation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2018:9000-9008.
20. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(2):386-397.
21. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. *IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016:770-778.
22. Lin TY, Dollar P, Girshick R, He KM, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017. IEEE; 2017:936-944.
23. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. *Lect Notes Comput Sci*. 2014;8693(V):740-755.
24. Girshick R. Fast R-CNN. *IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2015;1440-1448.
25. Wu Y, Kirillov A, Massa F, Lo W-Y, Girshick R. Detectron2. Accessed July 2, 2021. <https://github.com/facebookresearch/detectron2>
26. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a. *Med Phys*. 2006;33(10):3874-3900.
27. Depuydt T, Verellen D, Haas O, et al. Geometric accuracy of a novel gimbals based radiation therapy tumor tracking system. *Radiother Oncol*. 2011;98(3):365-372.
28. Iramina H, Nakamura M, Mizowaki T. Direct measurement and correction of both megavoltage and kilovoltage scattered X-rays for orthogonal kilovoltage imaging subsystems with dual flat panel detectors. *J Appl Clin Med Phys*. 2020;21(9):143-154.
29. Akimoto M, Nakamura M, Nakamura A, et al. Inter- and intrafractional variation in the 3-dimensional positions of pancreatic tumors due to respiration under real-time monitoring. *Int J Radiat Oncol Biol Phys*. 2017;98(5):1204-1211.
30. Minn AY, Schellenberg D, Maxim P, et al. Pancreatic tumor motion on a single planning 4D-CT does not correlate with intrafraction tumor motion during treatment. *Am J Clin Oncol*. 2009;32(4):364-368.

How to cite this article: Zhou D, Nakamura M, Mukumoto N, Yoshimura M, Mizowaki T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Med Phys*. 2022;1–9. <https://doi.org/10.1002/mp.15456>

RESEARCH

Open Access



Development of AI-driven prediction models to realize real-time tumor tracking during radiotherapy

Dejun Zhou¹, Mitsuhiro Nakamura^{1,2*} , Nobutaka Mukumoto², Hiroaki Tanabe³, Yusuke Iizuka², Michio Yoshimura², Masaki Kokubo⁴, Yukinori Matsuo² and Takashi Mizowaki²

Abstract

Background: In infrared reflective (IR) marker-based hybrid real-time tumor tracking (RTTT), the internal target position is predicted with the positions of IR markers attached on the patient's body surface using a prediction model. In this work, we developed two artificial intelligence (AI)-driven prediction models to improve RTTT radiotherapy, namely, a convolutional neural network (CNN) and an adaptive neuro-fuzzy inference system (ANFIS) model. The models aim to improve the accuracy in predicting three-dimensional tumor motion.

Methods: From patients whose respiration-induced motion of the tumor, indicated by the fiducial markers, exceeded 8 mm, 1079 logfiles of IR marker-based hybrid RTTT (IR Tracking) with the gimbal-head radiotherapy system were acquired and randomly divided into two datasets. All the included patients were breathing freely with more than four external IR markers. The historical dataset for the CNN model contained 1003 logfiles, while the remaining 76 logfiles complemented the evaluation dataset. The logfiles recorded the external IR marker positions at a frequency of 60 Hz and fiducial markers as surrogates for the detected target positions every 80–640 ms for 20–40 s. For each logfile in the evaluation dataset, the prediction models were trained based on the data in the first three quarters of the recording period. In the last quarter, the performance of the patient-specific prediction models was tested and evaluated. The overall performance of the AI-driven prediction models was ranked by the percentage of predicted target position within 2 mm of the detected target position. Moreover, the performance of the AI-driven models was compared to a regression prediction model currently implemented in gimbal-head radiotherapy systems.

Results: The percentage of the predicted target position within 2 mm of the detected target position was 95.1%, 92.6% and 85.6% for the CNN, ANFIS, and regression model, respectively. In the evaluation dataset, the CNN, ANFIS, and regression model performed best in 43, 28 and 5 logfiles, respectively.

Conclusions: The proposed AI-driven prediction models outperformed the regression prediction model, and the overall performance of the CNN model was slightly better than that of the ANFIS model on the evaluation dataset.

Keywords: Real-time tumor tracking, Tumor motion prediction, Convolutional neural network, Adaptive neuro-fuzzy inference system

Background

During beam delivery, the targets—particularly those located in the thoracic and abdominal regions—move during respiration [1]. Conventionally, the internal target volume method is the most common approach to

*Correspondence: m_nkmr@kuhp.kyoto-u.ac.jp

¹ Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, 53 Kawahara-Cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

perform radiation therapy for such targets, as it sufficiently covers the range of movement [2]. However, in this approach, along with the target, the surrounding normal tissue is also irradiated at a high dose, which may have adverse consequences.

With recent advances in radiotherapy systems, four-dimensional (4D) radiotherapy can be performed in clinical practice. In this approach, breath-hold, respiratory gating, and real-time tumor tracking (RTTT) techniques can be adopted to reduce the effects of respiratory motion [3]. In particular, through the RTTT, the beam position can be changed with the target position, thereby minimizing the occurrence of the adverse events caused by the motion of the target without burdening the patient's breath or extending the treatment time [4].

In September 2011, we started infrared reflective (IR) marker-based RTTT (IR Tracking) with a gimbal-head radiotherapy system, known as Vero4DRT (Hitachi Ltd., Tokyo, Japan, and Brainlab AG, Feldkirchen, Germany) [5], for lung [6], liver [7], and pancreatic [8] cancer patients. IR Tracking is a hybrid RTTT technique that combines direct positioning and indirect RTTT methods [9]. The IR Tracking method predicts the internal target position with the positions of IR markers placed on the patient's abdominal wall by using the regression-based prediction model. The prediction accuracy of IR Tracking depends considerably on the performance of the prediction model [10]. We have observed that the regression-based prediction model does not represent the tumor motion accurately. In this regard, the existing regression-based prediction model implemented in Vero4DRT can be improved in terms of accuracy [11–13].

Artificial intelligence (AI) techniques are being extensively and rapidly implemented in radiotherapy [14]. In general, support vector regression [15, 16], Gaussian process regression [17], neural networks [18, 19], and fuzzy logic [20, 21] can be applied to predict target positions with external surrogate positions; however, these algorithms are associated with specific limitations when applied to prediction models. The performance of the support vector regression is not satisfactory when data is used from free-breathing humans [16]. Moreover, the Gaussian process regression approach was tested only on a respiratory simulation phantom model with a rubber hot-water bottle [17]. Consequently, the simulation cannot accurately reflect the actual patient breathing and tumor motion. Although neural networks and fuzzy logic exhibit a satisfactory performance, in the research conducted with the use of these techniques, only 3 and 20 patients were tested, respectively [18, 20, 21]. In the work presented by Teo et al., only the tumor motion in the superior–inferior direction was predicted with the input of internal target position with an electronic portal

imaging device (EPID) at the frequency at 7.5 Hz [19, 22]. In addition, their approach cannot be performed if the internal target positions are invisible on EPID. Thus, the approach was rendered unsuitable for intensity-modulated radiation therapy or volumetric modulated arc therapy.

Considering such aspects, two AI-driven prediction models, expected to have enhanced accuracy, were developed in this work. A convolutional neural network (CNN)-driven model with fine-tuning, and a model driven by an adaptive neuro-fuzzy inference system (ANFIS) with a pattern search algorithm were used. Compared to the regression-based prediction model, the AI-driven prediction models can better predict the internal target position in three dimensions (3D) using the external marker position without changes in the current workflow. Compared to the CNN model, the ANFIS model contained fewer layers, does not require building a reference model, and does not need too much data for training. With two different approaches based on AI, it may be possible to have more alternatives for future research. By comparing and analyzing the performance of the different prediction models in the same scenarios, we can obtain a better understanding of their characteristics. The performance of the proposed prediction models was further compared to that of the regression model and was evaluated to enhance the understanding of the difference between the AI mechanisms and the conventional approach involved in developing prediction models.

Methods

IR tracking procedure of Vero4DRT

The details of the RTTT procedure for Vero4DRT were described in a previous study [10].

Before the treatment beam delivery, an IR camera mounted on the ceiling of the treatment room monitors the motion of the one-dimensional (1D) IR markers placed on the abdominal wall every 16.7 ms. In addition, the orthogonal kV X-ray imaging subsystem implemented in Vero4DRT detects the fiducial markers as surrogates for the detected target positions (P_{detect}) every 80–640 ms. These motions are monitored for 20–40 s. After monitoring, a regression-based prediction model $f(P_{\text{IR}}, v_{\text{IR}})$ is built as follows:

$$f(P_{\text{IR}}, v_{\text{IR}}) = aP_{\text{IR}}^2 + bP_{\text{IR}} + c + dv_{\text{IR}}^2 + ev_{\text{IR}}, \quad (1)$$

where P_{IR} is the averaged 1D IR marker position of multiple IR markers' measurements, v_{IR} is the averaged vertical velocity of the IR markers. The positions of the IR markers are predicted from the past position (25 ms before) to compensate for the system delay. The parameters from a to e are optimized by linear regression.

During the treatment beam delivery, the future 3D target position (P_{predict}) is predicted from the position and velocity of IR markers with the use of Eq. (1). Additionally, the internal target position is monitored every 1 s to verify the results produced by the prediction model.

Data characteristics

This research was performed in accordance with the Declaration of Helsinki and was approved by the institutional review board. A total of 1079 logfiles were extracted from Vero4DRT. These logfiles were obtained from lung, liver, and pancreatic cancer patients whose 3D respiration-induced motion of the tumor, as indicated by the fiducial markers, exceeded 8 mm during IR Tracking. We selected 8 mm based on previous studies, considering the adaptation of respiratory motion management techniques [23] and significance of long- and short-term tumor motion variability [24]. All the patients were breathing freely, and more than four external IR markers were attached on the abdominal walls of each patient. The logfiles were acquired when building the prediction model. In our clinical practice protocol, we recorded IR marker positions during an interval of 20–40 s at a frequency of 60 Hz before treatment beam delivery and detected target positions indicated by the implanted markers for the same period at intervals ranging from 80 to 640 ms. During this period, the regression prediction model was constructed [10–12].

The 1079 logfiles were randomly divided into two datasets. To improve the prediction accuracy, as many datasets as possible are required. In this study, the historical dataset for the CNN model contained 1003 logfiles, and the remaining 76 logfiles complemented the evaluation dataset. The evaluation dataset was used to evaluate the

performance of the prediction models. For each logfile in the evaluation dataset, the first three quarters were used as the training periods for transfer learning to build the patient-specific prediction models. The last quarter was the testing period and was used to test the performance of the prediction model.

Table 1 summarizes the IR marker motion patterns for the logfiles in the evaluation dataset. The mean and standard deviation (SD) values of the peak-to-peak motion range (R), the breathing period (T), and the 90th percentile of the respiratory velocity (v_{90}) during the training and testing periods were calculated separately. The absolute difference of each value was calculated to show whether the respiratory motion was smooth and stable. Table 2 shows the summary of tumor motion range in three directions. The mean and SD values of the detected target motion ranges in the right–left, superior–inferior, and anterior–posterior directions during the training and testing period were calculated.

CNN-driven prediction model

The schema for the CNN model is shown in Fig. 1. The CNN model was constructed to have nineteen layers in total, with eight convolution layers, five batch normalization layers, three dropout layers, a flatten layer, and two dense layers. In this study, the model was implemented in Python 3.6.4 and Keras 2.1.2. The Adam optimizer was employed, and the loss function was the mean value of the absolute differences between the detected and predicted target positions presented by the CNN model.

The CNN model consisted of training based on a large patient population and conducting patient-specific transfer learning. Initially, a single reference model was developed with the use of the historical dataset. In this

Table 1 Summary of infrared reflective (IR) marker motion characteristics for evaluation dataset

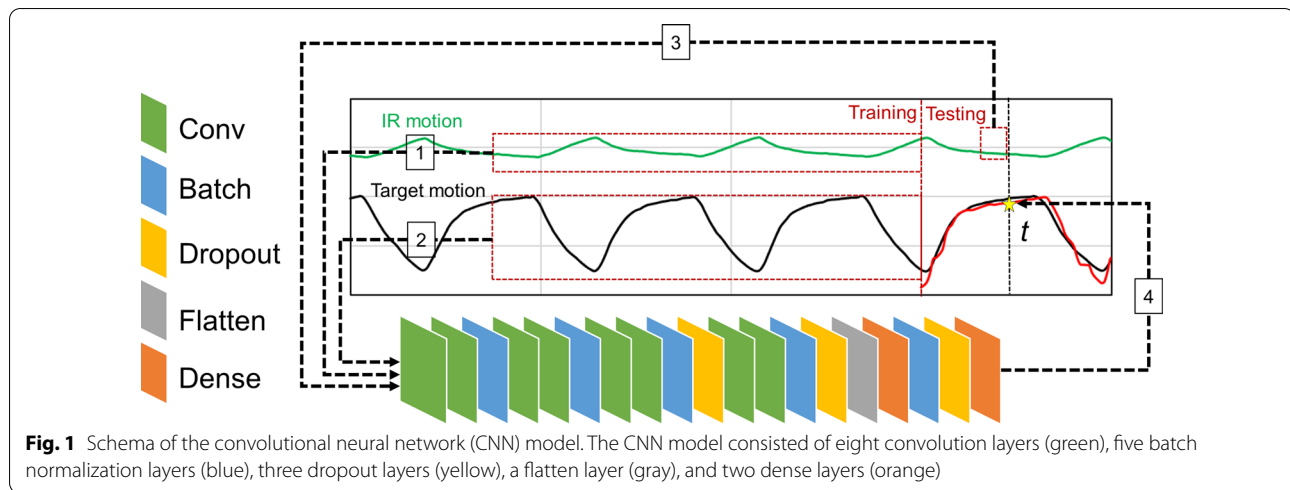
	Training period	Testing period	Absolute difference	<i>p</i> -value
R (mm)	7.2 ± 2.9 [2.9–15.5]	7.1 ± 3.0 [2.9–16.5]	1.2 ± 1.7 [0.0–10.4]	0.87
T (s)	4.2 ± 1.5 [2.5–8.2]	4.2 ± 1.4 [1.1–8.4]	0.6 ± 0.9 [0.0–5.7]	0.97
v_{90} (mm/s)	7.3 ± 2.0 [4.0–13.5]	7.4 ± 2.2 [3.8–17.5]	0.8 ± 0.9 [0.0–5.0]	0.74

R , Peak-to-peak motion range; T , period; v_{90} , 90th percentile of respiratory velocity. Values are presented as means $\pm$ standard deviations (SD) [range, min–max]. *P*-values are the paired *t*-test results between training period and testing period

Table 2 Summary of detected target motion ranges in three directions for the evaluation dataset

	Training period	Testing period	Absolute difference	<i>p</i> -value
Right–left (mm)	2.1 ± 1.7 [0.5–9.1]	2.1 ± 1.9 [0.4–10.8]	0.4 ± 0.4 [0.0–2.2]	0.76
Superior–inferior (mm)	16.4 ± 8.1 [7.6–37.1]	16.5 ± 8.6 [7.4–46.0]	2.4 ± 3.8 [0.0–27.1]	0.92
Anterior–posterior (mm)	3.0 ± 1.3 [0.8–6.4]	3.1 ± 2.4 [0.8–19.9]	0.7 ± 1.9 [0.0–16.7]	0.64

Values are shown in means $\pm$ SD [range, min–max]. *P*-values are the paired *t*-test results between training period and testing period



process, CNN could learn and acquire knowledge from the dataset. The reference model was trained for 20 epochs with a learning rate of 0.001. The reason for the setting of parameters was based on the consideration that the reference model was fine-tuned later. The construction of the reference model involved the following steps:

1. Randomly extract the data for 12 s P_{IR} (720 positions) for IR marker No. 1 from a single historical logfile (dashed line 1 in Fig. 1).
2. Extract 25 P_{detect} values from the same period at equal intervals (dashed line 2 in Fig. 1).
3. Extract 50 P_{IR} immediately before time t in the last quarter of a single historical logfile (dashed line 3 in Fig. 1).
4. Calculate $P_{predict}$ at time t [$P_{predict}(t)$] in the last quarter of the single historical logfile with the data from steps 1–3 (dashed line 4 in Fig. 1).
5. Train the reference model and learn the weights based on the aforementioned steps.
6. Repeat the steps until all the IR markers, time intervals, and entire historical dataset are covered.
3. Extract 50 P_{IR} immediately before time t in the training period.
4. Tune the reference model with $P_{detect}(t)$ and the data acquired in steps 1–3 until all the time points in the training period are covered.
5. Calculate $P_{predict}(t)$ one at a time. The input was 50 P_{IR} immediately before time t , with the last 720 P_{IR} and 25 P_{detect} values of the training period.
6. Repeat step 5 until all the markers are covered.
7. Calculate the average value of $P_{predict}(t)$ calculated with each IR marker. This result is the final prediction result.
8. Repeat step 7 until all the time t in testing period is covered.

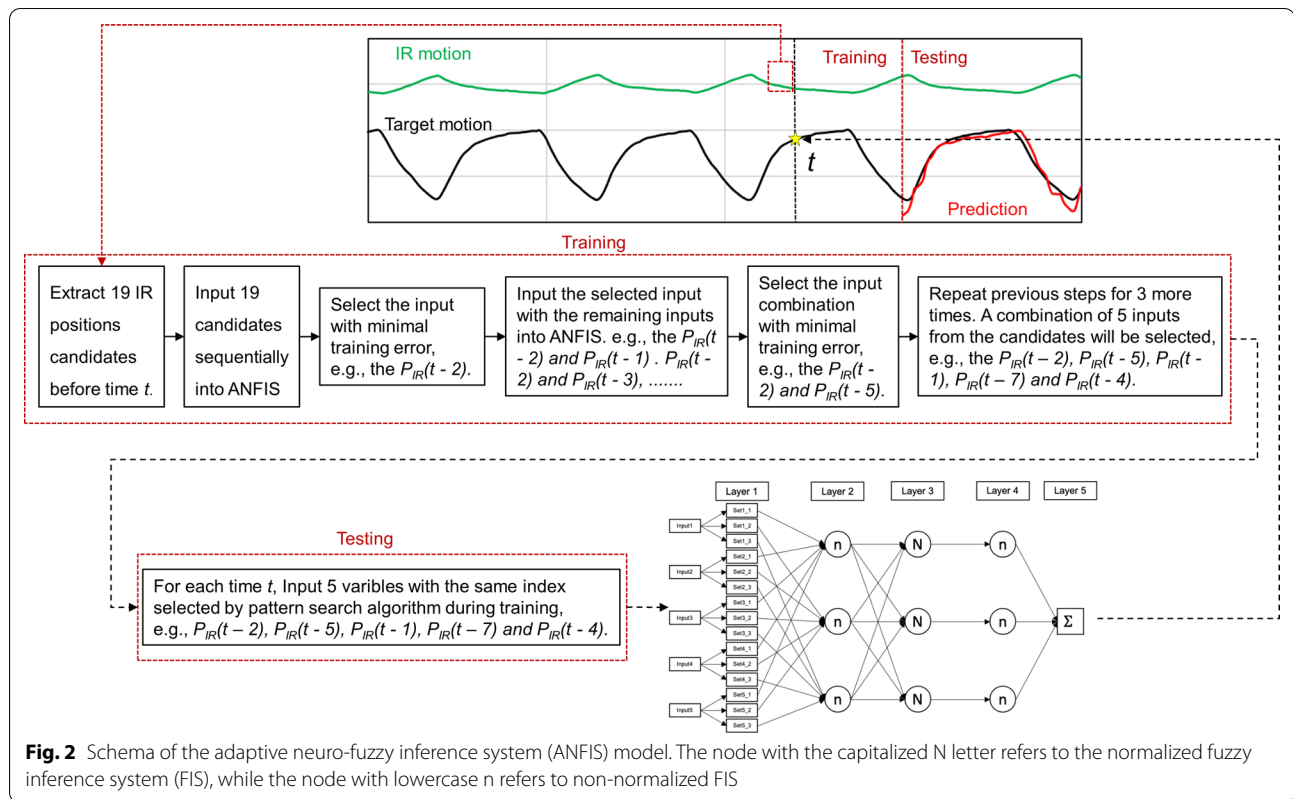
For each logfile in the evaluation dataset, the reference model was tuned using the data in the first three-quarters of the logfile (training period). The tuned reference model was trained for five epochs, with the learning rate ranging linearly from 0.0005 to 0.0001. For each logfile, the CNN model calculated $P_{predict}$ using the following steps:

1. Extract the data for the last 12 s P_{IR} (720 positions) of IR marker No. 1 from the training period.
2. Extract 25 P_{detect} values from the same period with equal intervals.

ANFIS-driven model

The ANFIS technique combined the adaptive neural network and fuzzy inference system (FIS). The FIS used the fuzzy set theory and fuzzy rules to map the inputs to the outputs. The fuzzy set was generated through a clustering algorithm, and the mapping was performed by considering the membership function and fuzzy rules. Subsequently, a five-layer adaptive neural network was adapted as a machine learning approach to tune the FIS parameters. The detailed information regarding ANFIS can be found in [25] and [26].

The schema for the ANFIS model presented in this work is shown in Fig. 2. The input of the model was the IR marker positions, and the output was the predicted target position in 3D. The ANFIS model predicted the target position one at a time, similar to the CNN and regression-based models. In this work, FIS and ANFIS were implemented in MATLAB (R2020a, MathWorks, Natick, MA, USA) using the Fuzzy Logic and ANFIS Toolboxes.



The Fuzzy Logic Toolbox provided the Mamdani and Sugeno FIS types. The Sugeno-type FIS was adopted for the ANFIS model because of its higher computational efficiency compared to that of the Mamdani-type FIS. The hybrid method was selected as the optimization method in the ANFIS Toolbox. In particular, the hybrid method combined the backpropagation and least-squares estimation techniques for the parameters of the input and output membership functions, respectively. In the ANFIS model, the ANFIS was implemented with a pattern search algorithm, which can sequentially select the input data from the candidates to optimize the total squared error of the ANFIS during the training. For each logfile, the ANFIS model calculated P_{predict} based on the following steps:

1. Extract P_{detect} and P_{IR} of IR marker No. 1 from the training period of each logfile.
2. For each $P_{\text{detect}}(t)$, prepare 11 input candidates of P_{IR} (herein, values of $P_{\text{IR}}(t)$ to $P_{\text{IR}}(t-10)$ were selected owing to their proximity to $P_{\text{detect}}(t)$). Eight additional candidates, specifically, $P_{\text{IR}}(t-15)$, $P_{\text{IR}}(t-20)$, $P_{\text{IR}}(t-25)$, $P_{\text{IR}}(t-30)$, $P_{\text{IR}}(t-35)$, $P_{\text{IR}}(t-40)$, $P_{\text{IR}}(t-45)$, and $P_{\text{IR}}(t-50)$, were selected as they may influence $P_{\text{detect}}(t)$; the numbers refer to the index in the array of the IR marker motion data). The correspond-

ing P_{detect} values of the 11 input candidates were not extracted.

3. Process the 19 input candidates sequentially and select the candidate with the minimum training error in the ANFIS.
4. Sequentially process the remaining input candidates with the selected candidates and repeat steps 3 and 4 until five inputs are selected from the 19 candidates. These five inputs were considered to be the most relevant patterns of P_{IR} with $P_{\text{detect}}(t)$.
5. Train the model with $P_{\text{predict}}(t)$ and the most relevant pattern of P_{IR} during the training period for four epochs.
6. Calculate $P_{\text{predict}}(t)$ with the most relevant pattern of P_{IR} during the testing period.
7. Repeat step 6 until all the IR marks are covered.
8. The average value of $P_{\text{predict}}(t)$ calculated using each IR marker is the final prediction result.
9. Repeat step 8 until all the time t in testing period is covered.

Data analysis

The proposed prediction models processed each of the randomly selected 76 logfiles by using the aforementioned procedure. For each logfile, the CNN model

fine-tuned the reference model during the training period and yielded the prediction results for the testing period. The ANFIS model used the pattern search algorithm and developed the ANFIS for each logfile during the training period and predicted the target positions during the testing period. To enable a comparison, a regression model was constructed during the training period [13], and P_{predict} values were calculated during the testing period for each of the 76 logfiles.

During data analysis, the detected target position was considered as the ground truth of the prediction. The overall performance of the prediction model was ranked by the percentage of P_{predict} within 2 mm of P_{detect} at each recorded time. Furthermore, the cumulative percentage curve of 3D prediction positional error for the three models was analyzed.

According to the International Organization of Standardization (ISO) standard 5725-1 [27], the accuracy of a measurement is a combination of the trueness (mean error) and precision (standard deviation of the error, SD). In this study, the performances of the proposed CNN, ANFIS, and regression model on a single logfile were also evaluated in terms of accuracy. The mean absolute error (MAE) and SD between P_{predict} and P_{detect} were calculated for each logfile from the evaluation dataset. The parametric paired t -test was performed to evaluate the statistical significance of MAE between the AI-driven and regression model; the level of significance was set to 0.05.

As the CNN model learnt from the historical dataset and the ANFIS model was trained and tested on a single logfile, the change in the respiratory range, period, and velocity measured by the IR markers for the training and testing periods of the logfile might have influenced the comparison of the proposed prediction models. To quantify these changes, the variables δ_r , δ_p , and δ_v , that indicated the degrees of change in the respiration range, period, velocity between the training and testing periods, respectively, were defined and calculated for each logfile:

$$\delta = \left| \frac{\text{Value}_{\text{test}}}{\text{Value}_{\text{train}}} - 1 \right|, \quad (2)$$

where $\text{Value}_{\text{test}}$ and $\text{Value}_{\text{train}}$ represent the corresponding values during the testing and training periods, respectively. For the range and period, the values were the mean range and period, respectively. For velocity, the values were the 90th percentiles of the IR velocity during the testing period and training period, respectively. A larger δ value indicates a greater change. In particular, for stable respiratory patterns, δ will be close to zero.

Results

The averaged training times of the CNN and ANFIS models for each logfile were approximately 12 s and 95 s, respectively.

The cumulative percentage curve of the 3D prediction positional error is also consistent with the aforementioned result (Fig. 3). As shown in Fig. 3, the CNN and ANFIS models exhibit nearly the same cumulative percentage distribution when the 3D prediction positional error is smaller than 1 mm. When the distance between P_{predict} and P_{detect} ranged from 1 to 3 mm, the CNN model exhibited the highest performance. Overall, the performance of the AI-driven models was better than that of the regression model. The percentages of 3D prediction positional error within 2 mm were 95.1%, 92.6% and 85.6% for the CNN, ANFIS, and regression models, respectively. This indicates that the CNN model showed the best performance among the three models. There were significant differences in MAE between the CNN and regression model ($p < 0.05$) and between the ANFIS and regression model ($p < 0.05$).

The mean $\pm$ SD values of the degrees of change in the respiration range (δ_r), period (δ_p), and velocity (δ_v) between the training and testing periods were 0.17 ± 0.32 (range, 0.00–2.69), 0.13 ± 0.17 (range, 0.00–0.98), and 0.10 ± 0.12 (range, 0.00–2.69), respectively. Figure 4 shows the relationships between δ_r , δ_p , and δ_v , and MAE $\pm 2\text{SD}$. Figure 4a shows that the performances of the ANFIS and CNN models are comparable. Figure 4b and c show that the performance comparison of the CNN and ANFIS models is stable at all ranges for δ_p and δ_v . Meanwhile, the AI-driven models always performed better than the regression model. In the following section, the performance comparison of the prediction models is discussed based on δ_r .

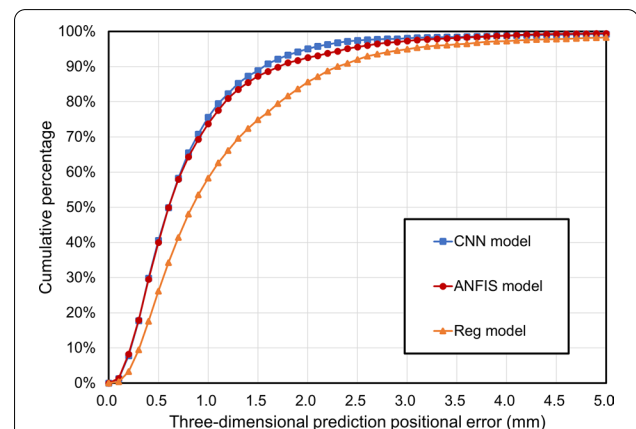


Fig. 3 Cumulative percentage curve regarding P_{predict} within P_{detect} in designated distance

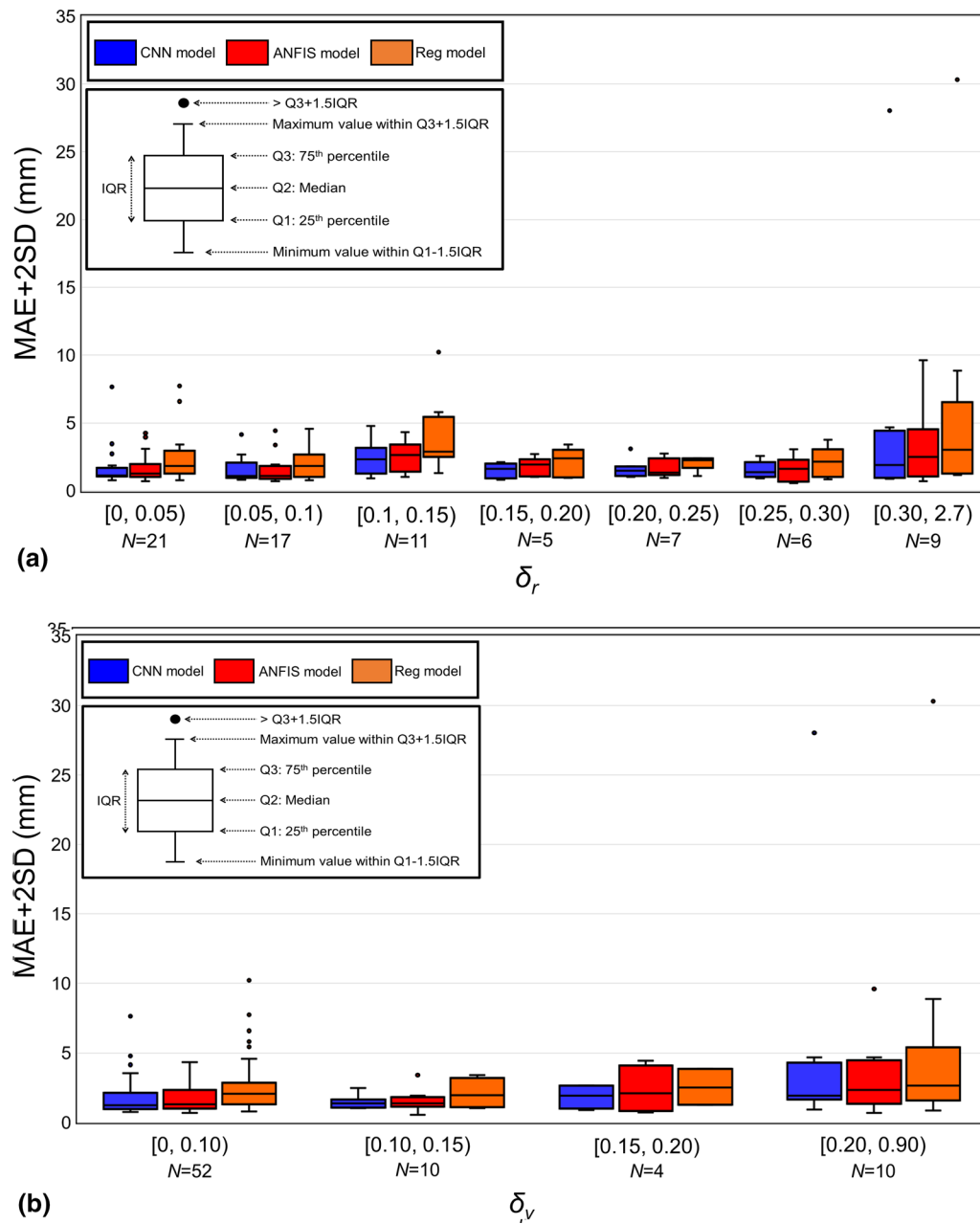


Fig. 4 Mean absolute error (MAE) + 2 standard deviations (SD) between P_{predict} and P_{detect} versus (a) δ_r , (b) δ_v , and (c) δ_r . The boxes represent the interquartile ranges (IQRs). Outliers were above the third quartile plus $1.5 \times \text{IQR}$. Blue, red, and orange represent the CNN, ANFIS, and regression models, respectively. N in the horizon axis refers to the number of logfile regarding to each range

Upon comparison, it was noted that for the 43 logfiles (56.6%) showing that the CNN model outperformed the other models, the median δ_r value was equal to 0.12. In contrast, for the 28 logfiles wherein the ANFIS model outperformed the other models, the median δ_r was observed to be 0.07 (36.8%). Figure 5 shows an example of IR motion with a δ_r value of 0.06. The MAE+2SD

values of the CNN and ANFIS models were 1.29 and 0.71 mm, respectively. In this case, the performance of the ANFIS model was slightly better than that of the CNN model for the respiration range between the training and testing periods. As δ_r increased, the change in the respiration range became significant, and the CNN model outperformed the other models. For instance, in

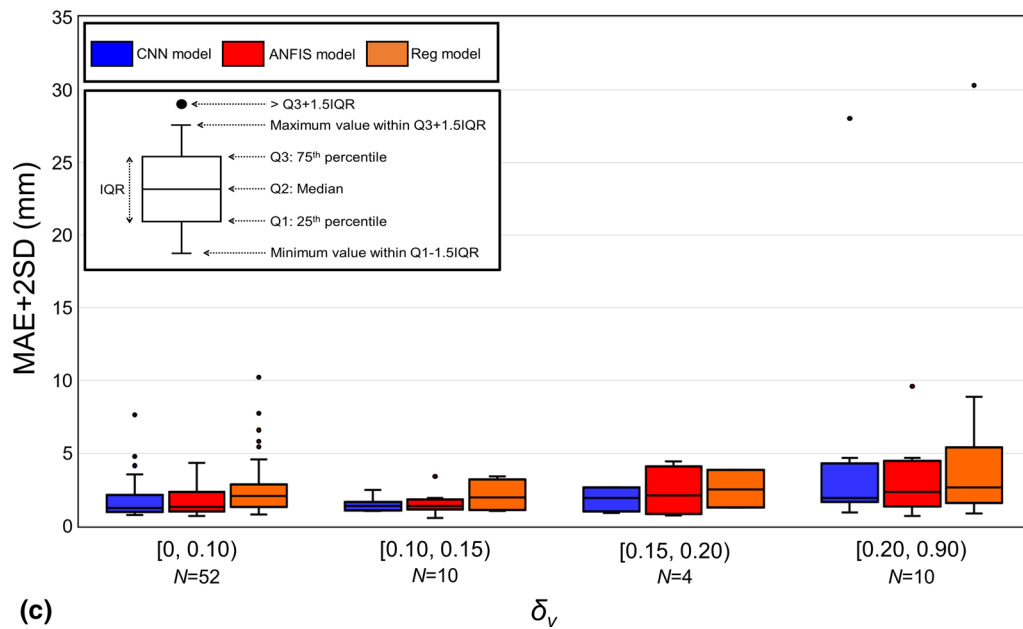


Fig. 4 continued

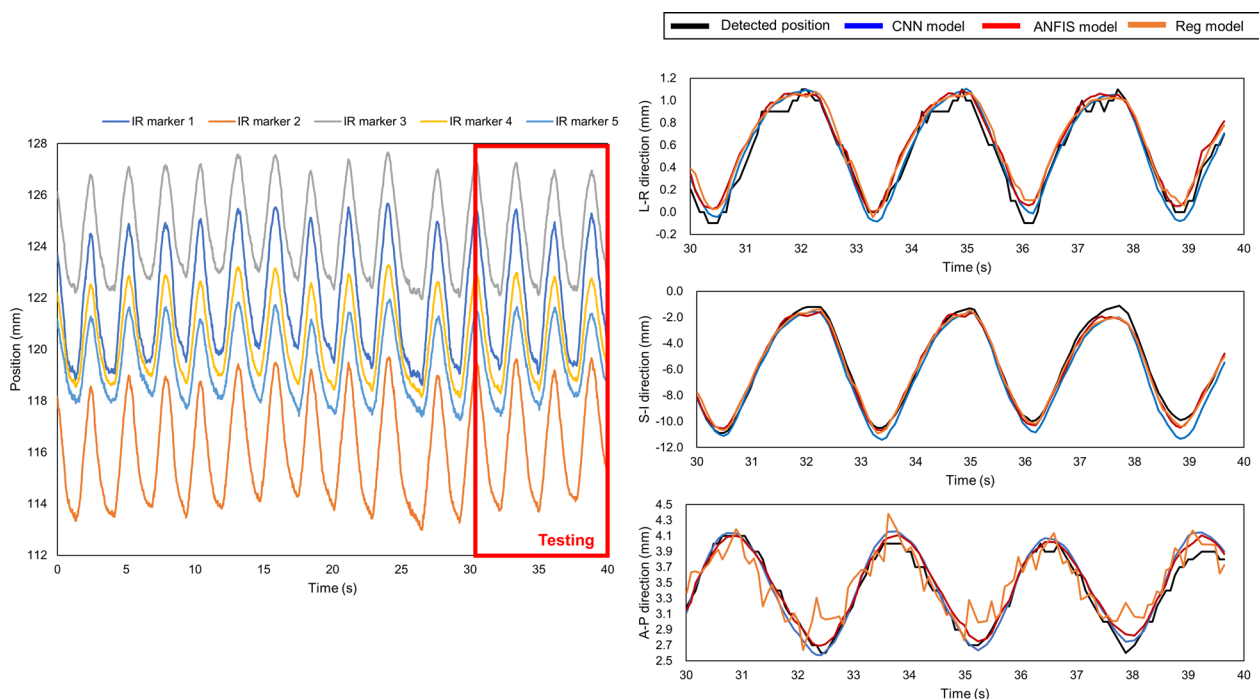


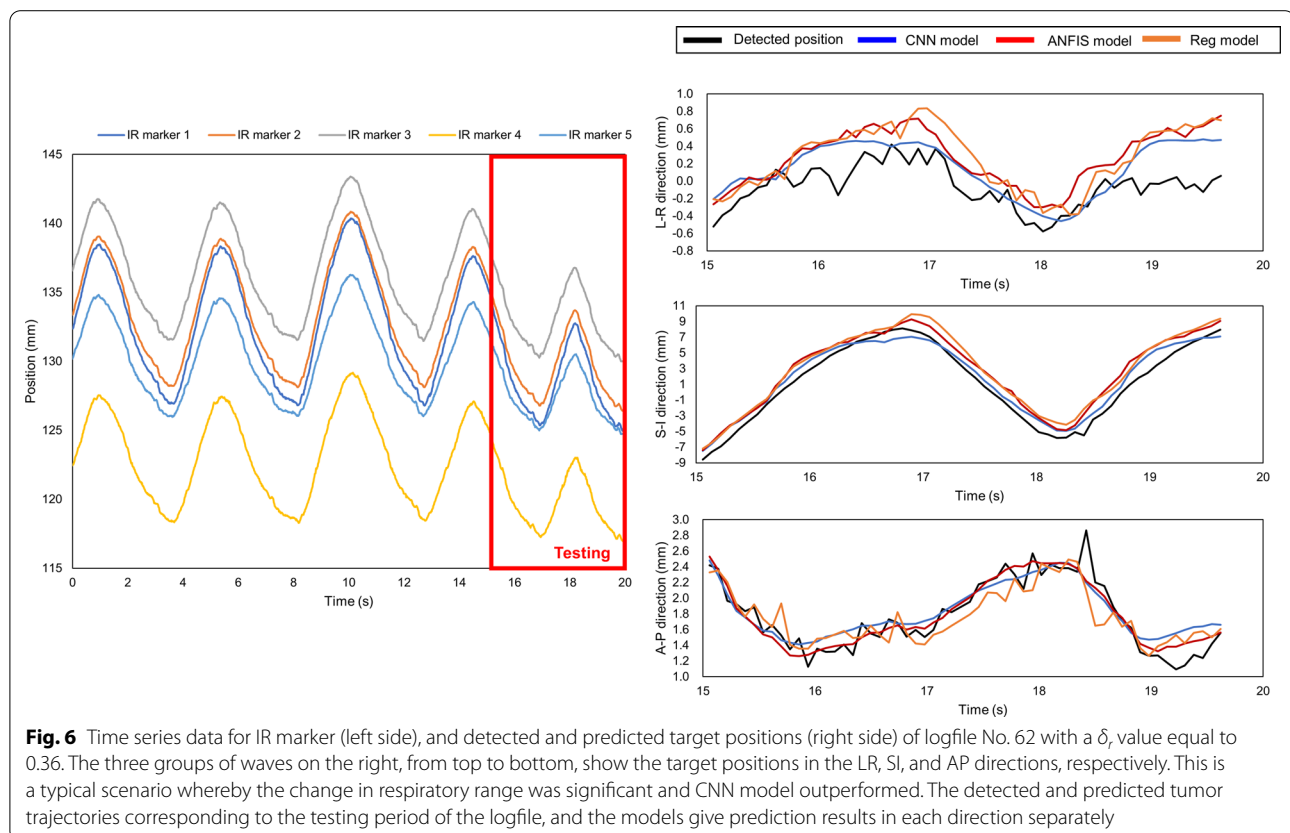
Fig. 5 Time series data for IR marker (left side), detected and predicted target position (right side) of logfile No. 38 with a δ_v of 0.06. The three groups of waves on the right, from top to bottom, show the target positions in the LR, SI, and AP directions, respectively. This is a typical scenario that indicates that the respiratory was stable and the ANFIS model outperformed. The detected and predicted tumor trajectories corresponding to the testing period of the logfile, and the models give prediction results in each direction separately

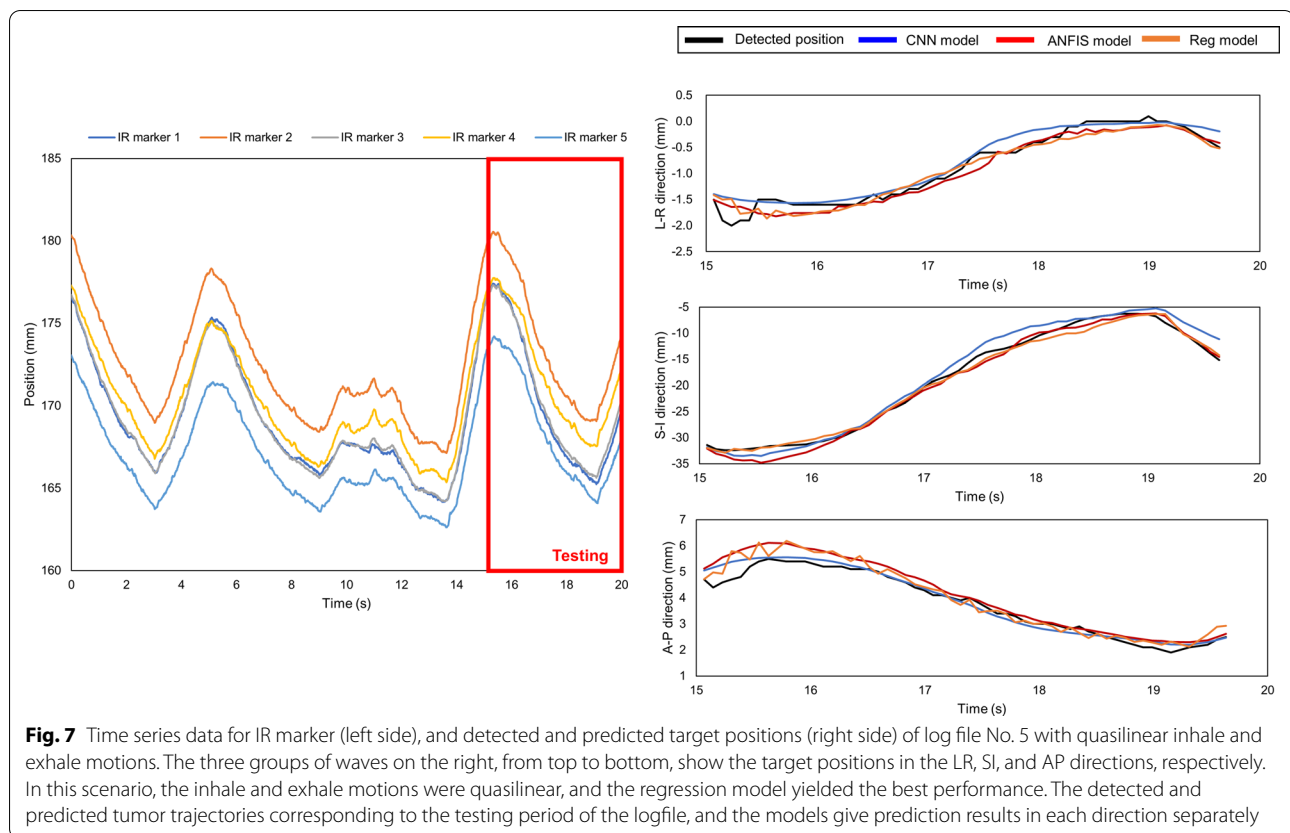
the case of logfile No. 62 with a δ_r value of 0.36 (Fig. 6), the MAE+2SD values of the CNN and ANFIS models were 1.90 and 2.91 mm, respectively. Among the 76 logfiles, the regression model exhibited superior performances in the cases of five logfiles (6.6%), in which the inhale and exhale motions were quasilinear. For log file No. 5 (Fig. 7), the MAE+2SD values of the regression, CNN, and ANFIS model were 1.87, 3.10, and 2.39 mm, respectively. Logfile No. 44 had a δ_r of 2.69, which corresponded to the maximum value among the 76 logfiles. As shown in Fig. 8, the patient inhales deeply during the last quarter of the recording time, leading to an irregular value of δ_r and produced the maximum MAE+2SD for all three prediction models.

Discussion

In this study, the prediction performances of the CNN and ANFIS models were compared to that of a regression model that has been utilized clinically. The CNN model was initially built as a single reference model with the historical dataset, and patient-specific transfer learning was later conducted during the training period. The ANFIS model was driven by ANFIS for each logfile, and a pattern search algorithm was adopted to select the most relevant input data. The test results showed that both AI-driven

prediction models exhibited better overall performance than the regression model tested on the 76 logfiles. For each logfile, the averaged training time of the CNN and ANFIS model was approximately 12 s and 95 s, respectively. Considering the training data acquisition duration, which was 20 s to 40 s, the model construction durations of the CNN and ANFIS models were less than 52 s and 135 s, respectively. This was less than the average model construction duration of the regression model, which was 162 s, as reported by Depuydt et al. [28]. With shorter model construction durations, shorter treatment session durations can be expected if the AI-driven models are applied in clinical practice. The median value of MAE for the 76 logfiles in test dataset was 0.65, 0.66, and 1.02 mm for the CNN, ANFIS, and regression model, respectively. Thus, as the CNN and ANFIS models showed better accuracy and shorter model construction durations, less times and shorter durations of model retraining during a treatment fraction can also be expected. The durations of a treatment session will be further shortened. The p -values between the AI-driven models and regression model were less than 0.05, indicating that the performance of the AI-driven models was significantly better than that of the regression model. Institutionally, the setting of margin for RTTT considers both the errors induced by the



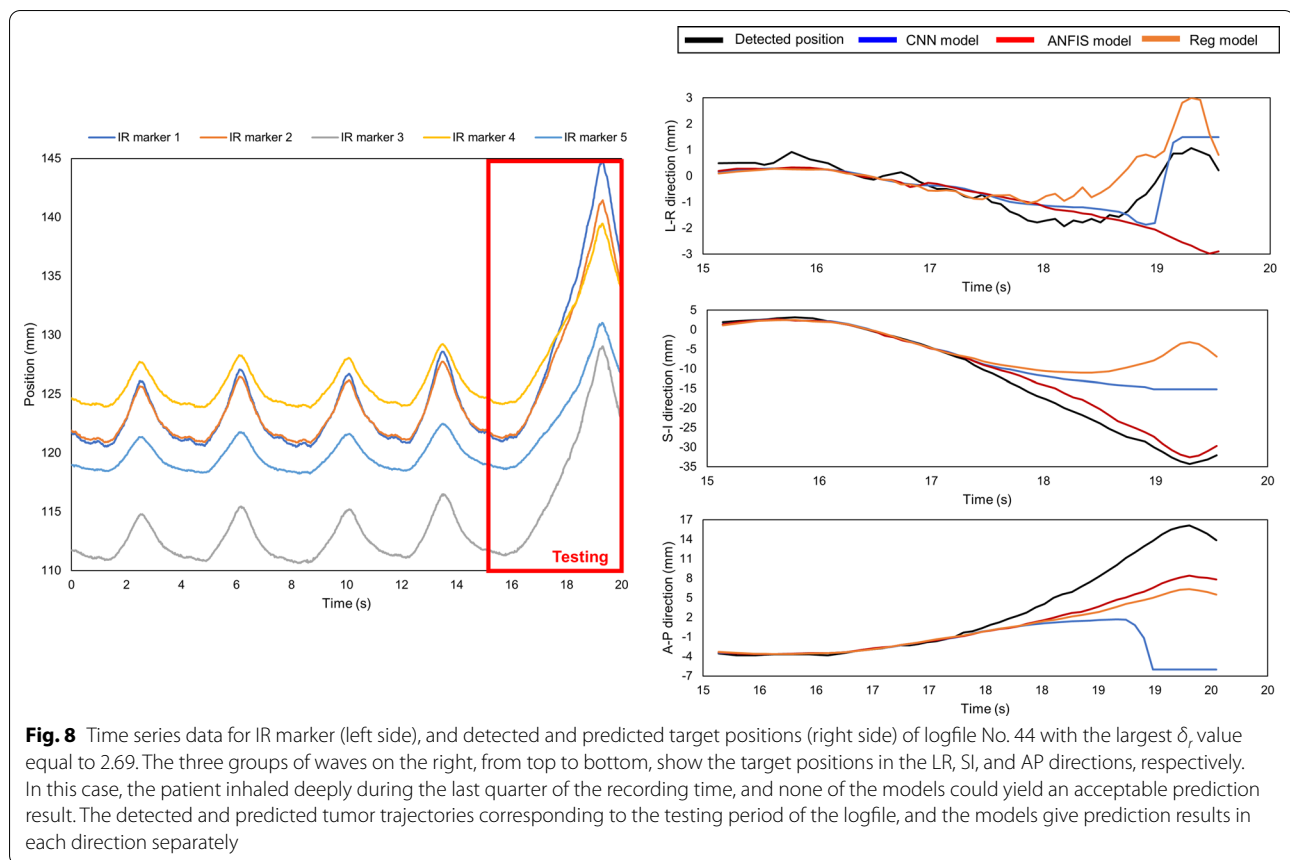


internal markers [29] and the accuracy of the prediction model [10]. The implementation of AI-driven prediction models in clinical practice is expected to reduce the margin derived by the positional error of prediction models and benefit the patient in the future.

The CyberKnife (CK) system can perform RTTT using a prediction model other than Vero4DRT [9, 30]. The conventional [31] and ANFIS approach [20, 21] to construct a prediction model for a CK system was reported in previous studies. According to Poels et al. [31], the prediction accuracy of the regression model was comparable with that of conventional CK models. The results of this study demonstrated that the proposed model notably outperformed the regression model. Considering this statement, it can be concluded that the CNN and ANFIS models would outperform the conventional CK models. The work conducted by Torshabi and Ghorbanzadeh et al. adapted ANFIS to predict target motion with external marker motions for a CK system [20, 21]. In their work, the patients were divided into a control group whose tumor tracking was carried out smoothly, and a worst group that was the opposite. For their research, the average 3D root mean square error for the control group was 1.1 mm. In the present work, taking into consideration all the logfiles in the test dataset, the median value

of MAE was 0.65 and 0.66 mm for the CNN and ANFIS models, respectively. This indicates that the present AI-driven models show better performances as compared to those discussed in previous works.

Compared to the Gaussian process regression model [17], which uses a rubber hot-water bottle to simulate respiratory motion, the proposed models were trained and tested with actual clinical data. The application prospects of such models in clinical practice may be more promising. Moreover, the predictions obtained using the support vector regression [16] and neural networks [18] pertained to a small patient cohort (7 and 3 patients, respectively). The present research was performed based on considerations of 76 logfiles, which corresponded to more reliable results. Comparing ours to the work done by Isaksson et al. [18], the performance of their neural network model decreased notably within 5 s; thus, the model needs to be updated within every 5 s. For the AI-driven models presented in this work, the testing period ranged from 5 to 10 s and the performance was stable, as demonstrated in Figs. 5, 6 and 7. The subsequent model accuracy will depend on δ_r , δ_p , and δ_v , as shown in Fig. 4. The model presented by Teo et al. [19] required the detected target position with EPID at a frequency of 7.5 Hz and provided the



prediction results only in the superior–inferior direction during the treatment beam delivery. In contrast, the target position was predicted in 3D with IR markers at 60 Hz without information of the internal target position. If the orthogonal kV X-ray imaging subsystem works at a higher frequency, the patient may receive additional dosage. Considering both our situation and the trade-off between the dosage and prediction accuracy, the models presented in this work may be more suitable for us.

Although the proposed prediction models can notably outperform the regression model, certain limitations remain. The CNN model exhibited a high performance when the scenarios were similar to those of the logfile learnt from the historical data based on the CNN. Furthermore, the ANFIS model was trained and tested solely based on the logfile and benefited from the pattern search algorithm. When the associated respiratory motion was stable, or when training was performed for a larger number of respiratory cycles, the performance of the ANFIS model would be comparable to that of the CNN model. According to this finding, the classification of the respiratory motion followed by the selection of appropriate models is expected to lead to higher prediction accuracy.

In the unique logfile No. 44 (Fig. 8a), the CNN model could not produce an accurate prediction result. This may have been caused by the imbalance in the historical dataset because the irregular respiratory patterns, for example that in logfile No. 44, were seldom included in the historical data. If additional logfiles similar to file No. 44 were to be included during the construction of the reference model, or if the irregular IR motion was included during the training period for fine-tuning, the performance of the CNN model could be improved in a similar situation. The δ_r value of logfile No. 44 was 2.69. It was significantly large for the ANFIS model to provide accurate results. In the case of the regression model, the velocity of the IR markers changed drastically, and the coefficients of the regression model were not suitable for this scenario; this resulted in inferior performance. Owing to the advantages of pattern recognition ability and robustness of the historical dataset, the ANFIS model corresponded to a lower MAE + 2SD (9.61 mm) in this case. In contrast, the MAE + 2SD for logfile No. 44 was 28.02 mm when the CNN model was used. Nevertheless, none of the considered prediction models could provide an acceptable prediction result (Fig. 8b). For such cases, our current clinical protocol already has a fail-safe

approach. When the prediction error is larger than a pre-defined threshold (e.g., 3 mm, approximately half of the margin), the treatment beam will be automatically turned off [32]. When a systematic deviation is observed, the prediction model will be rebuilt and updated.

Among the 76 logfiles, the CNN, ANFIS, and regression models exhibited superior performances with minimum MAE + 2SD in 56.6%, 36.8%, and 6.6%, respectively. Even though the CNN and ANFIS models outperformed the regression model, all the possible scenarios in clinical practice cannot be covered. Specifically, for scenarios wherein the respiratory range changes considerably, as that shown in Fig. 4a whereby δ_r was greater than 0.3, the performances of the prediction models decreased. Based on the current study, the performance of the prediction model will decrease when δ_r increases. Currently, the input of the AI-driven models was the 1D IR marker position. Correspondingly, whether the relationship between the IR marker and internal target positions was stable may have a dominant influence on the performance of AI-driven prediction models. This implies that regardless of how the velocity and period changed, if the relationship of the internal and external position was stable, the performance of the prediction model was stable. However, if the respiratory range changes significantly during the testing period compared to that during the training period, the AI-driven models cannot learn the position relationship during the training period. This change would cause a negative influence on the prediction accuracy (Fig. 4a). To address such situations, the automatic beam-off function can be implemented, in which the MV beam delivery is automatically turned off if the detected 3D target position is beyond a predefined threshold [32]. In addition, the use of high-dose-rate, flattening, filter-free beams could significantly reduce the radiation delivery time, potentially contributing toward stabilization of the prediction accuracy.

Overall, this study was associated with three notable limitations. Firstly, point-by-point predictions were only performed during the last quarter of the logfiles. This corresponded to approximately 5 to 10 s of the target motion. As reported by Poels et al. [33], if a patient's breathing motion is not stable and the prediction accuracy becomes unacceptable during treatment, the prediction model must be updated. To overcome this limitation, sequential prediction model updates can be implemented during beam delivery, or the technique presented by Teo et al. can be adapted to reduce the tracking drift in position [34]. Secondly, only the data from Vero4DRT were adapted to train and test the prediction models. The performances of the models on other systems must be examined in the future at different sampling rates, such as the CK system. Thirdly, only δ_r , δ_p and δ_v , which represented

respiratory pattern changes, were considered in this study; however, there may be other factors that may need to be used to reduce the tracking accuracy. Lastly, this was a retrospective study. Currently, the models were developed, trained, and tested on previously acquired logfiles. In the future, more well-conceived experiments will be considered. The future experiments may contain longer recording durations and more irregular respiratory patterns to further improve the AI-driven models.

Conclusions

The overall performance of the proposed CNN and ANFIS models were considerably better than that of the currently employed regression model. The CNN model performed slightly better than the ANFIS model based on tests conducted with the 76 randomly selected logfiles. Changes in the model performances were examined at different patient scenarios. In the case of considerable changes in the respiration range, the CNN model may exhibit the optimal performance. In contrast, in the case of stable respiratory ranges, the ANFIS model may achieve high prediction accuracy. Additional work can be performed to expand the application scenarios of the AI-driven models and conduct parameter optimization.

Abbreviations

1D: One-dimensional; 3D: Three-dimensional; 4D: Four-dimensional; AI: Artificial intelligence; ANFIS: Adaptive neuro-fuzzy inference system; CK: CyberKnife; CNN: Convolutional neural network; EPID: Electronic portal imaging device; FIS: Fuzzy inference system; IR: Infrared reflective; IR Tracking: Infrared reflective marker-based hybrid real-time tumor tracking; ISO: International Organization of Standardization; MAE: Mean absolute error; P_{detect} : Detected 3D internal position; P_{IR} : 1D infrared reflective marker position; P_{predict} : 3D predicted target position; R : Peak-to-peak motion range of infrared reflective marker; RTTT: Real-time tumor tracking; SD: Standard deviation; T : Breathing period; v_{90} : 90Th percentile of the respiratory velocity.

Acknowledgements

We sincerely appreciate the technical support and guidance from all the staff members of the Medical Physics Laboratory of Kyoto University Graduate School of Medicine (<http://medicalphysics.hs.med.kyoto-u.ac.jp/>) provided throughout the study.

Authors' contributions

DZ and MN planned the study, performed the statistical analysis, and drafted the manuscript. NM, HT, YI, MY, MK, YM, and TM conceived the study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Funding

This research was partly supported by AMED [Grant Number JP20he2302001].

Availability of data and materials

Authors are not able to share data.

Declarations

Ethical approval and consent to participate

This study followed all the dictates of the Declaration of Helsinki and the Ethics Review Board of Kyoto University Hospital, and the Faculty of Medicine

approved the research. Written consent to participate was obtained from the patients.

Consent for publication

Written consent was obtained from the patients for publication of this report and any accompanying images.

Competing interests

We have no competing interest to disclose.

Author details

¹Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, 53 Kawahara-Cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. ²Department of Radiation Oncology and Image-Applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto, Japan. ³Department of Radiological Technology, Kobe City Medical Center General Hospital, Hyogo, Japan. ⁴Department of Radiation Oncology, Kobe City Medical Center General Hospital, Hyogo, Japan.

Received: 20 October 2021 Accepted: 14 February 2022

Published online: 23 February 2022

References

- Matsuo Y, Onishi H, Nakagawa K, et al. Guidelines for respiratory motion management in radiation therapy. *J Radiat Res.* 2013;54(3):561–8. <https://doi.org/10.1093/jrr/rrs122>.
- International Commission on Radiation Units and Measurements ICRU Report 62. Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50), ICRU, Bethesda, MD (1999)
- Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys.* 2006;33(10):3874–900. <https://doi.org/10.1118/1.2349696>.
- Schweikard A, Shiomi H, Adler J. Respiration tracking in radiosurgery. *Med Phys.* 2004;31(10):2738–41. <https://doi.org/10.1118/1.1774132>.
- Hiraoka M, Mizowaki T, Matsuo Y, et al. The gimbaled-head radiotherapy system: rise and downfall of a dedicated system for dynamic tumor tracking with real-time monitoring and dynamic WaveArc. *Radiother Oncol.* 2020. <https://doi.org/10.1016/j.radonc.2020.07.002>.
- Matsuo Y, Ueki N, Takayama K, et al. Evaluation of dynamic tumour tracking radiotherapy with real-time monitoring for lung tumours using a gimbal mounted linac. *Radiother Oncol.* 2014;112(3):360–4. <https://doi.org/10.1016/j.radonc.2014.08.003>.
- Iizuka Y, Matsuo Y, Ishihara Y, et al. Dynamic tumor-tracking radiotherapy with real-time monitoring for liver tumors using a gimbal mounted linac. *Radiother Oncol.* 2015;117(3):496–500. <https://doi.org/10.1016/j.radonc.2015.08.033>.
- Nakamura A, Hiraoka M, Itasaka S, et al. Evaluation of dynamic tumor-tracking intensity-modulated radiotherapy for locally advanced pancreatic cancer. *Sci Rep.* 2018;8(1):1–10. <https://doi.org/10.1038/s41598-018-35402-7>.
- Dieterich S, Cleary K, D'Souza W, et al. Locating and targeting moving tumors with radiation beams. *Med Phys.* 2008;35(12):5684–94. <https://doi.org/10.1118/1.3020593>.
- Akimoto M, Nakamura M, Mukumoto N, et al. Predictive uncertainty in infrared marker-based dynamic tumor tracking with Vero4DRT. *Med Phys.* 2013;40(9):091705. <https://doi.org/10.1118/1.4817236>.
- Mukumoto N, Nakamura M, Sawada A, et al. Accuracy verification of infrared marker-based dynamic tumor-tracking irradiation using the gimbaled x-ray head of the Vero4DRT (MHI-TM2000). *Med Phys.* 2013;40(4):041706. <https://doi.org/10.1118/1.4794506>.
- Mukumoto N, Nakamura M, Yamada M, et al. Intrafractional tracking accuracy in infrared marker-based hybrid dynamic tumour-tracking irradiation with a gimbaled linac. *Radiother Oncol.* 2014;111(2):301–5. <https://doi.org/10.1016/j.radonc.2014.02.018>.
- Akimoto M, Nakamura M, Mukumoto N, et al. Baseline correction of a correlation model for improving the prediction accuracy of infrared marker-based dynamic tumor tracking. *J Appl Clin Med Phys.* 2015;16(2):14–22. <https://doi.org/10.1120/jacmp.v16i2.4896>.
- El Naqa I, Ruan D, Valdes G, et al. Machine learning and modeling: data, validation, communication challenges. *Med Phys.* 2018;45(10):e834–40. <https://doi.org/10.1002/mp.12811>.
- Ernst F, Martens V, Schlichting S, et al. Correlating chest surface motion to motion of the liver using e-SVR: a porcine study. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, 2009; 356–364. https://doi.org/10.1007/978-3-642-04271-3_44.
- Ernst F, Bruder R, Schlaefer A, et al. Correlation between external and internal respiratory motion: a validation study. *Int J Comput Assist Radiol Surg.* 2012;7(3):483–92. <https://doi.org/10.1007/s11548-011-0653-6>.
- Özbek Y, Bárdosi Z, Freysinger W. respiTrack: patient-specific real-time respiratory tumor motion prediction using magnetic tracking. *Int J Comput Assist Radiol Surg.* 2020;15(6):953–62. <https://doi.org/10.1007/s11548-020-02174-3>.
- Isaksson M, Jalden J, Murphy MJ. On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Med Phys.* 2005;32(12):3801–9. <https://doi.org/10.1118/1.2134958>.
- Teo P, Ahmed B, Kawalec P, et al. Feasibility of predicting tumor motion using online data acquired during treatment and a generalized neural network optimized with offline patient tumor trajectories. *Med Phys.* 2018;45(2):830–45. <https://doi.org/10.1002/mp.12731>.
- Torshabi AE, Pella A, Riboldi M, Baroni G. Targeting accuracy in real-time tumor tracking via external surrogates: a comparative study. *Technol Cancer Res Treat.* 2010;9(6):551–62. <https://doi.org/10.1177/153303461000900603>.
- Ghorbanzadeh L, Torshabi AE, Nabipour JS, et al. Development of a synthetic adaptive neuro-fuzzy prediction model for tumor motion tracking in external radiotherapy by evaluating various data clustering algorithms. *Technol Cancer Res Treat.* 2016;15(2):334–47. <https://doi.org/10.1177/1533034615571153>.
- Teo P, Crow R, Van Nest S, et al. Tracking lung tumour motion using a dynamically weighted optical flow algorithm and electronic portal imaging device. *Meas Sci Technol.* 2013;24(7):074012. <https://doi.org/10.1088/0957-0233/24/7/074012>.
- Negoro Y, Nagata Y, Aoki T, et al. The effectiveness of an immobilization device in conformal radiotherapy for lung tumor: reduction of respiratory tumor movement and evaluation of the daily setup accuracy. *Int J Radiat Oncol Biol Phys.* 2001;50(4):889–98. [https://doi.org/10.1016/s0360-3016\(01\)01516-4](https://doi.org/10.1016/s0360-3016(01)01516-4).
- Dhont J, Vandemeulebroucke J, Burghelée M, et al. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. *Radiother Oncol.* 2018;126(2):339–46. <https://doi.org/10.1016/j.radonc.2017.09.001>.
- Babuška R, Verbruggen H. Neuro-fuzzy methods for nonlinear system identification. *Annu Rev Control.* 2003;27(1):73–85. [https://doi.org/10.1016/S1367-5788\(03\)00009-9](https://doi.org/10.1016/S1367-5788(03)00009-9).
- Jang J-SR. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern.* 1993;23(3):665–85. <https://doi.org/10.1109/21.256541>.
- International Organization for Standardization. Accuracy (Trueness and Precision) of Measurement Methods and Results – DIN ISO 5725-2; 1994.
- Depuydt T, Poels K, Verellen D, et al. Treating patients with real-time tumor tracking using the Vero gimbaled linac system: implementation and first review. *Radiother Oncol.* 2014;112(3):343–51. <https://doi.org/10.1016/j.radonc.2014.05.017>.
- Ueki N, Matsuo Y, Nakamura M, et al. Intra- and interfractional variations in geometric arrangement between lung tumours and implanted markers. *Radiother Oncol.* 2014;110(3):523–8. <https://doi.org/10.1016/j.radonc.2014.01.014>.
- Hoogeman M, Prévost JB, Nuytens J, et al. Clinical accuracy of the respiratory tumor tracking system of the CyberKnife: assessment by analysis of log files. *Int J Radiat Oncol Biol Phys.* 2009;74(1):297–303. <https://doi.org/10.1016/j.ijrobp.2008.12.041>.
- Poels K, Dhont J, Verellen D, et al. A comparison of two clinical correlation models used for real-time tumor tracking of semi-periodic motion: a focus on geometrical accuracy in lung and liver cancer patients. *Radiother Oncol.* 2015;115(3):419–24. <https://doi.org/10.1016/j.radonc.2015.05.004>.

32. Garibaldi C, Russo S, Ciardo D, et al. Geometric and dosimetric accuracy and imaging dose of the real-time tumour tracking system of a gimbal mounted linac. *Phys Med*. 2015;31(5):501–9. <https://doi.org/10.1016/j.ejmp.2015.04.001>.
33. Poels K, Depuydt T, Verellen D, et al. Improving the intra-fraction update efficiency of a correlation model used for internal motion estimation during real-time tumor tracking for SBRT patients: fast update or no update? *Radiother Oncol*. 2014;112(3):352–9. <https://doi.org/10.1016/j.radonc.2014.09.007>.
34. Teo P, Guo K, Fontaine G, et al. Reducing the tracking drift of an uncon-toured tumor for a portal-image-based dynamically adapted conformal radiotherapy treatment. *Med Biol Eng Comput*. 2019;57(8):1657–72. <https://doi.org/10.1007/s11517-019-01981-4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Development of independent dose verification plugin using Eclipse scripting API for brachytherapy

Dejun Zhou¹, Mitsuhiro Nakamura^{1,2,*}, Yohei Sawada³, Tomohiro Ono², Hideaki Hirashima², Hiraku Iramina², Takanori Adachi², Takahiro Fujimoto³ and Takashi Mizowaki²

¹Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

²Department of Radiation Oncology and Image-Applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

³Division of Clinical Radiology Service, Kyoto University Hospital, Kyoto 606-8507, Japan

*Corresponding author. Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, 53 Kawaharacho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. Tel: +81-75-751-4176; E-mail: m_nkmr@kuhp.kyoto-u.ac.jp

(Received 9 June 2022; revised 14 August 2022; editorial decision 26 August 2022)

ABSTRACT

In this study, an independent dose verification plugin (DVP) using the Eclipse Scripting Application Programming Interface (ESAPI) for brachytherapy was developed. The DVP was based on the general 2D formalism reported in AAPM-TG43U1. The coordinate and orientation of each source position were extracted from the translation matrix acquired from the treatment planning system (TPS), and the distance between the source and verification point (r) was calculated. Moreover, the angles subtended by the center-tip and tip-tip of the hypothetical line source with respect to the verification point (θ and β) were calculated. With r , θ , β and the active length of the source acquired from the TPS, the geometry function was calculated. As the TPS calculated the radial dose function, $g(r)$, and 2D anisotropy function, $F(r, \theta)$, by interpolating and extrapolating the corresponding table stored in the TPS, the DVP calculated $g(r)$ and $F(r, \theta)$ independently from equations fitted with the Monte Carlo data. The relative deviation of the fitted $g(r)$ and $F(r, \theta)$ for the GammaMed Plus HDR ¹⁹²Ir source was 0.5% and 0.9%, respectively. The acceptance range of the relative dose difference was set to $\pm 1.03\%$ based on the relative deviation between the fitted functions and Monte Carlo data, and the linear error propagation law. For 64 verification points from sixteen plans, the mean of absolute values of the relative dose difference was 0.19%. The standard deviation (SD) of the relative dose difference was 0.17%. The DVP maximizes efficiency and minimizes human error for the brachytherapy plan check.

Keywords: brachytherapy; independent dose verification; eclipse scripting

INTRODUCTION

Brachytherapy delivers a high dose to the target volume while realizing a steep dose falling away from the target. Compared with external beam radiotherapy (EBRT), brachytherapy can protect normal tissue while increasing the dose to the target volume. Brachytherapy treatments are conducted with a small fraction and large dose per fraction. Once there is a difference between the planned and delivered doses, it is hard to compensate for the negative effect caused by the difference [1]. Multiple works have recommended performing independent dose verifications to assess the dose calculated by the treatment planning system (TPS) before the brachytherapy treatment [2–4]. Currently the calculation-based verification is the realistic way to check

the brachytherapy plan before dose delivery rather than measurement-based verification.

The whole process of brachytherapy treatment in our hospital is shown in Fig. 1. First, the patient underwent computed tomography (CT) simulation with applicators. The brachytherapy treatment plan is then made based on the planning CT images. We do not change applicator setup during treatment planning. The independent dose verification is conducted after brachytherapy treatment planning. If the relative dose difference is within the tolerance, the brachytherapy will be conducted and dose will be delivered. If not, the treatment will be re-planned until the verification pass. Currently, in our institution, a Microsoft Excel-based independent dose verification is under clinical

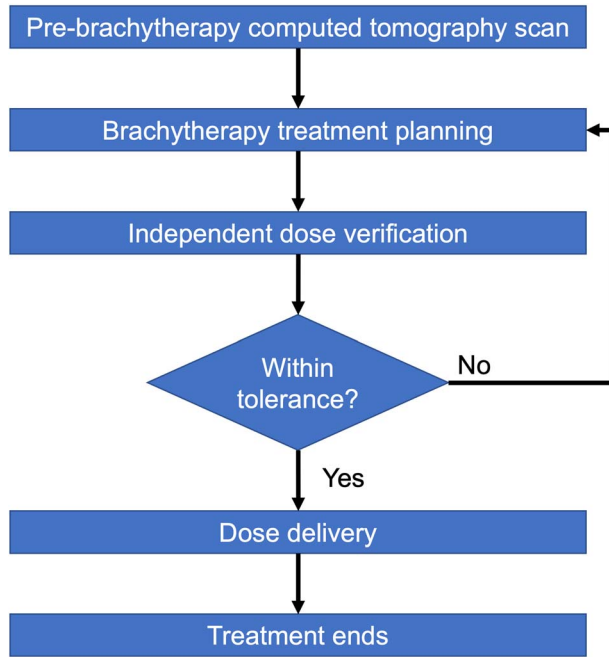


Fig. 1. The flowchart of brachytherapy treatment.

practice. Shortcomings of current verification method were revealed. Software other than the TPS was not allowed to be installed on clinical treatment planning machines. To perform the dose verification, the plan information was transferred from the machine with the TPS to another machine with the Microsoft Excel application and pasted to the template file; then, the verification dose was calculated. This procedure is not efficient and may induce human error. Moreover, as the Microsoft Excel-based independent dose verification does not adapt the orientation of each source position in the calculation, the current acceptance range of the relative dose difference of a selected dose verification point is a $\pm 5\%$ setting between the calculation results of the TPS and verification results. This is a relatively large range compared to the one used in the study reported by Carmona *et al.* [5], where the relative dose difference was within $\pm 2\%$. A potentially large acceptance range may give a false positive verification judgement.

To overcome the shortcomings, an independent dose verification plugin (DVP) using Eclipse Scripting Application Programming Interface (ESAPI; Varian Medical Systems, Palo Alto, CA, USA) for brachytherapy was developed [6]. The purpose of this study was to simplify the dose verification procedure for brachytherapy and improve the accuracy of the verification dose calculation.

MATERIALS AND METHODS

Description of dose verification plugin

In our institution, we use a Bravos unit (Varian Medical Systems) with the GammaMed Plus HDR ^{192}Ir active source and BrachyVision V16.1 (Varian Medical Systems) for brachytherapy treatment planning. For this reason, the DVP was dedicated to the Varian TPS and could be integrated with the TPS interface. The DVP was written in C# using ESAPI. With ESAPI, the DVP can access the plan information directly from the

TPS and show the dose verification report on the clinical machine with few clicks. There is no need to export, copy and paste data between computers and files. The plan details, especially the orientation of each source position, were adapted in the dose verification calculation in the DVP. In this way, the DVP will show more accurate calculation results, and the acceptance range for the dose verification will be narrowed. The source file of the DVP was one single file. It is easy to update the plugin and distribute it between institutions.

Dose calculation

The DVP calculates the dose at a verification point (P_{ver}) independently and compares the TPS results at the same point to verify the dose calculation. The dose calculation algorithm of the DVP was the general 2D formalism reported in the Association of Physicists in Medicine's update, Task Group 43 (AAPM-TG43U1) [7].

$$\dot{D}(r, \theta) = S_k \cdot \Lambda \cdot \frac{G_L(r, \theta)}{G_L(r_0, \theta_0)} \cdot g_L(r) \cdot F(r, \theta), \quad (1)$$

where r is the distance from the center of the active source to P_{ver} , θ is the angle subtended by the central axis of the active source and the line connecting the center of the active source and P_{ver} . r_0 and θ_0 are specified to 1 cm and 90° , respectively, according to AAPM-TG43U1. S_k is the air-kerma strength on the treatment day. Λ is the dose rate constant of the active source. G_L is the geometry function. g_L is the radial dose function, and F_L is the 2D anisotropy function.

The DVP accesses the calibration S_k and half-life of the active source as well as the calibration and treatment dates at 12 a.m. directly from the TPS with ESAPI. Subsequently, the value of S_k at the day of treatment was calculated using the information above and the law of radioactive decay. The DVP also directly acquires Λ from the TPS with ESAPI.

The active source in our institution is a line source. The DVP calculates the G_L based on the line-source model reported in AAPM-TG43U1 [7].

$$G_L(r, \theta) = \begin{cases} \frac{\beta}{L r \sin \theta}, & \text{if } \theta \neq 0^\circ \\ \left(r^2 - \frac{L^2}{4}\right)^{-1}, & \text{if } \theta = 0^\circ \end{cases}, \quad (2)$$

where L is the active length of the source. β is the angle subtended by the tips of the hypothetical line source with respect to the P_{ver} .

The first step of calculating G_L was to obtain the 3D coordinates of the active source center, both active source tips, and P_{ver} . The active length of the source was stored in the TPS, and the DVP acquired it directly. The TPS stored the position and orientation of the source with a transform matrix. The third column is the source orientation, and the fourth column is the center position. The positions of both tips of the source were calculated by the center position plus/minus the results of half of the active length, multiplied by the source orientation. The position of P_{ver} was defined in the TPS and directly acquired by the DVP. With the coordinate's information above, r , θ , β and G_L were calculated.

In the TPS, g_L and F_L were calculated by interpolating and extrapolating the corresponding data table stored in the radioactive source

Table 1. Fitted parameters of the anisotropy function for GammaMed Plus HDR ^{192}Ir . Zero values are represented by dashes

i	k_i	a_i	b_i	e_i
1	-2.30569	-	4.97×10^{-1}	-1.7×10^{-3}
2	-1.98×10^{-2}	-	-1.46	-2.96
3	2.847×10^{-2}	-3.25×10^{-1}	5.2×10^{-1}	-
4	2.27378	11.5962	24.586	1.469
i		a'_i	b'_i	e'_i
1		-	-14.54	-1.14×10^{-1}
2		-	-1.5588×10^{-1}	-1.057
3		-6.3265×10^{-1}	-4.47×10^{-1}	-1.81×10^{-2}
4		17.0192	39.889	1.2924

model. In the DVP, these two functions were calculated independently with fitted dosimetric parameters and equations reported by Lliso *et al.* [8]. The function for g_L was:

$$g_L(r) = \frac{hr^i}{1+hr^k}, \quad (3)$$

where for the GammaMed Plus HDR ^{192}Ir source, h , i , j and k were equals to 1.001, 7.69×10^{-3} , 2.1×10^{-4} and 2.63, respectively.

The general functional forms of F_L were

$$F(r, \theta) = k(r) + \frac{a(r)\left(\frac{\theta}{180^\circ}\right)^{e(r)}}{1+b(r)\left(\frac{\theta}{180^\circ}\right)^{e(r)}} + \frac{a'(r)\left(1-\frac{\theta}{180^\circ}\right)^{e'(r)}}{1+b'(r)\left(1-\frac{\theta}{180^\circ}\right)^{e'(r)}}, \quad (4)$$

where

$$k(r) = k_1 r^{k_2} + k_3 r + k_4,$$

$$a(r) = a_1 r^{a_2} + a_3 r + a_4, a'(r) = a'_1 r^{a'_2} + a'_3 r + a'_4$$

$$b(r) = b_1 r^{b_2} + b_3 r + b_4, b'(r) = b'_1 r^{b'_2} + b'_3 r + b'_4$$

$$e(r) = e_1 r^{e_2} + e_3 r + e_4, e'(r) = e'_1 r^{e'_2} + e'_3 r + e'_4$$

The fitted parameters of F_L for the GammaMed Plus HDR ^{192}Ir are summarized in Table 1.

Subsequently, the dose rate of each source position at P_{ver} was calculated using equation 1, the DVP calculated the dose at P_{ver} using the following equation:

$$D_{DVP} = \sum_{i=1}^N \dot{D}_i(r, \theta) \bullet t_i, \quad (5)$$

where i is the index of the active source in the plan, and t is the dwell time.

Dose verification

The P_{ver} dose calculated by the DVP was compared with the dose calculated by the TPS. The relative dose difference ($D_{diff\%}$) was calculated using the following equation:

$$D_{diff\%} = 100\% \times \frac{D_{DVP} - D_{TPS}}{D_{TPS}}, \quad (6)$$

where D_{TPS} was the dose at P_{ver} calculated by the TPS.

According to Lliso *et al.* [8], the average absolute value of the relative deviation between the anisotropy function and Monte Carlo data was 0.9% for the GammaMed Plus HDR ^{192}Ir source. For the radial dose function, the value was 0.5%. Based on the linear error propagation theory,

$$\Delta(g_L \bullet F_L) = \pm \sqrt{(\Delta g_L)^2 + (\Delta F_L)^2}, \quad (7)$$

where Δg_L and ΔF_L were the average relative deviations of g_L and F_L , respectively. $\Delta(g_L \bullet F_L)$ was the average relative deviation of g_L multiplied by that of F_L . We assumed that, other than g_L and F_L , there were no deviations in the other components of the dose calculation. In this case, we set $\pm 1.03\%$ as the acceptance range for $D_{diff\%}$.

Patient characteristics

This study was approved by the Institutional Review Board of Kyoto University Hospital (approval number: R1446). Our institution started treating patients with the Bravos system from April 2022. Three patients who underwent brachytherapy treatment were included in this study. Two patients were treated with tandem-ovoid applicators, and one was treated with a tandem cylinder. The fractional dose was 6 Gy. At the time of writing, two patients underwent 3 fractions, and one patient underwent 1 fraction. The patients underwent a computer tomography scan and were prepared for each fraction. For each plan, four verification points located at the bladder, rectum, point A at the left side and point A at the right side of the tandem, were determined by the on-site medical physicists or radiation technician, according to the International Commission on Radiation Units and Measurements Report 38 recommendations [9].

Clinical workflow

The user interface of the DVP is shown in Fig. 2. Before running the DVP, at least one reference point with a location should be selected as the dose verification point and stored in the reference point list of the current plan. The first window of the DVP was the dose verification-point selection window (Fig. 2a). The list of the combo box contained all non-abstract reference points of the plan. After selection and clicking

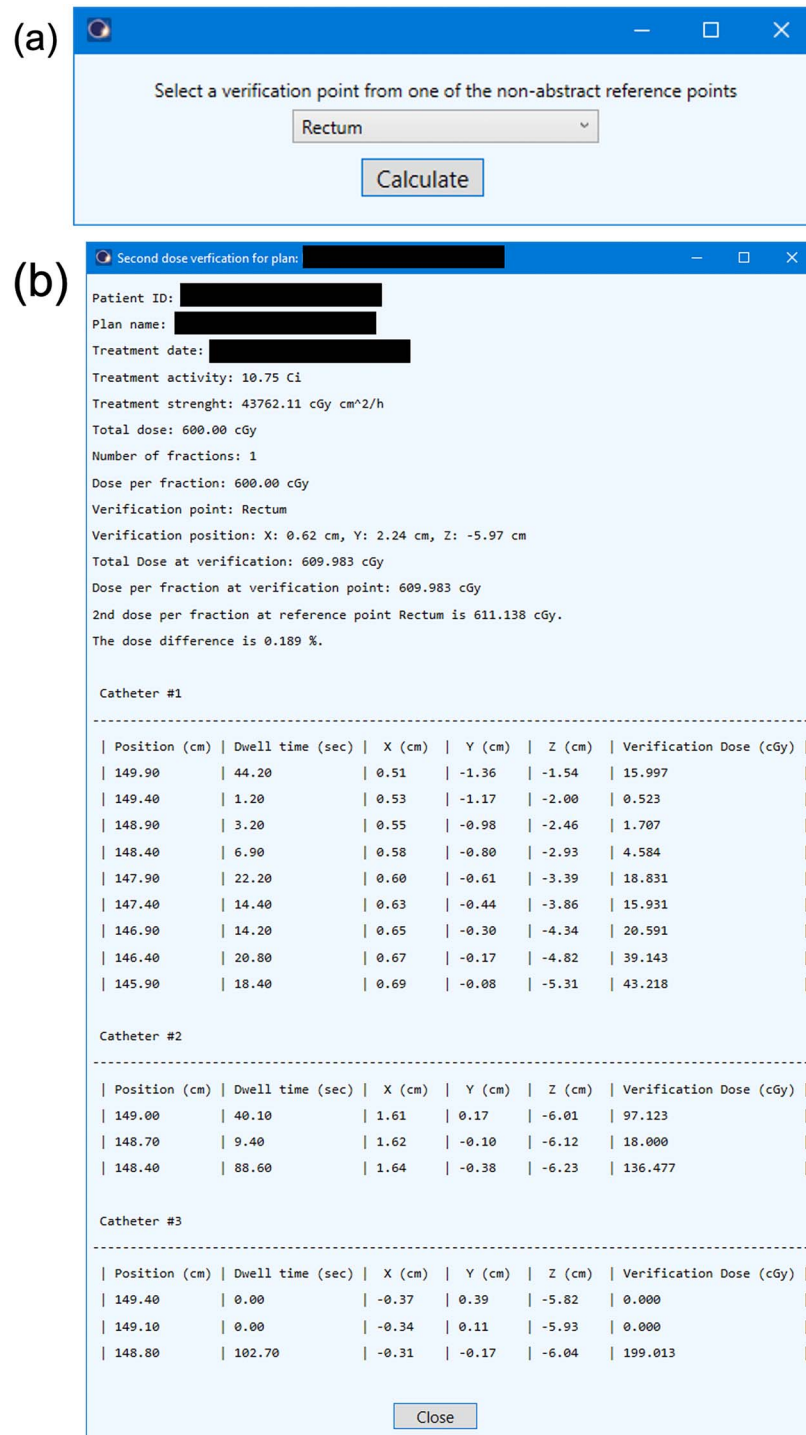


Fig. 2. User interface of the DVP. (a) The dose verification point selection window. (b) The dose verification report windows.

the 'calculate' button, the calculation report was shown in the next window (Fig. 2b). The calculation report contains important treatment information, D_{DVP} , D_{TPS} , $D_{diff\%}$ and the dose calculated by the DVP at each source position.

RESULTS AND DISCUSSION

For 16 plans, 64 verification points were selected and included in the statistic results. The mean of the absolute $D_{diff\%}$ was 0.19%. The standard deviation (SD) of $D_{diff\%}$ was 0.17%.

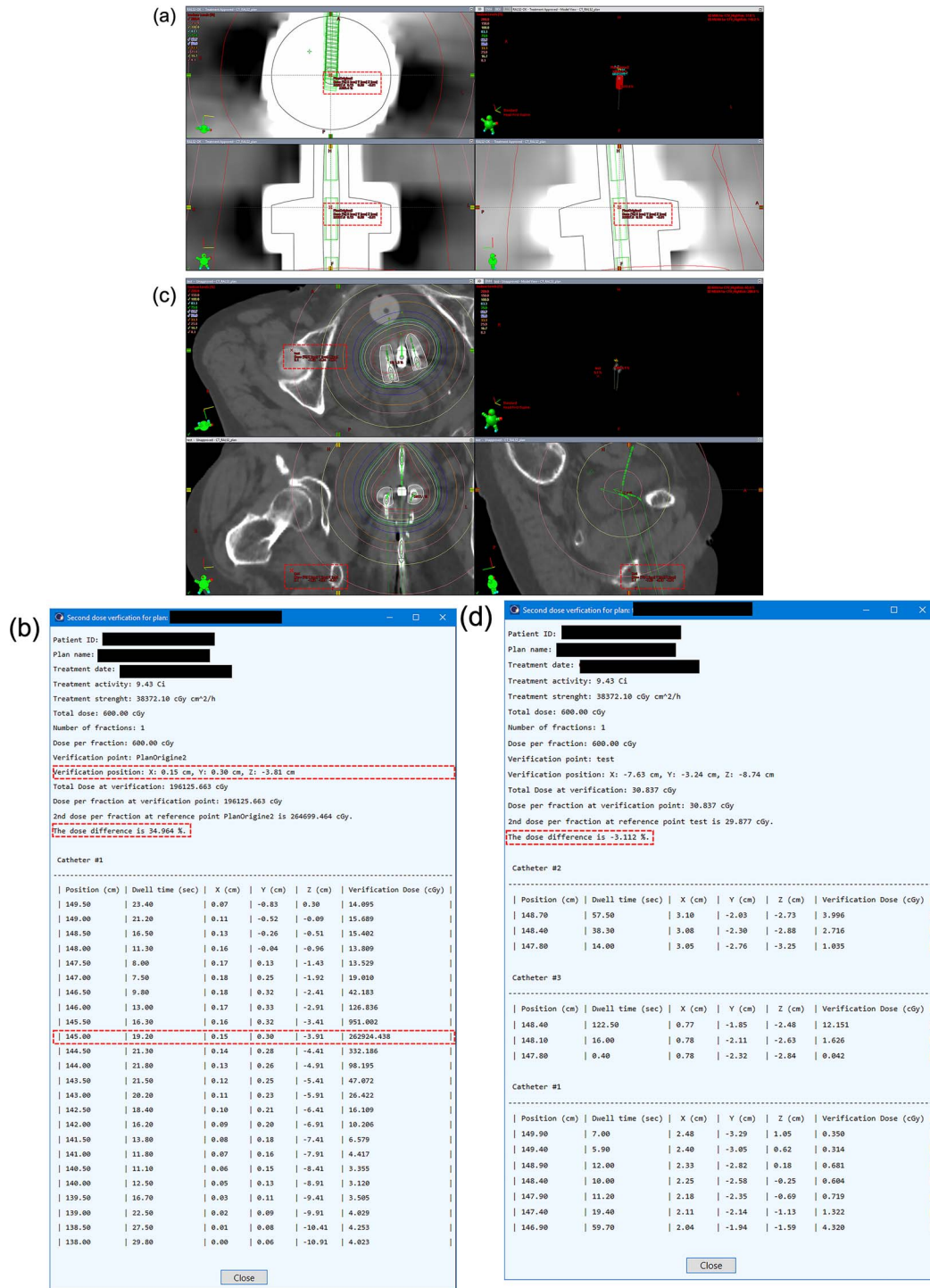


Fig. 3. Demonstration of the failed dose verification when the dose verification point was too close to one of the source positions ([a] and [b]), and too far from the sources ([c] and [d]).

Among all the 64 results, the calculation result of the DVP was smaller than that of the TPS for 83% of verification points (S3 verification points). This was caused by the curve of the fitted ΔG_L , and ΔF_L was always below the curve of the Monte Carlo simulated data, as reported by Lliso *et al.* [8].

In the two situations, the absolute $D_{diff\%}$ may be out of the acceptance range. First, if the verification point was inside one of the source positions, a large dose difference was observed. As shown in Fig. 3a–b, the verification point was inside one of the source positions. $D_{diff\%}$ of this verification point was 34.96%.

Second, when the verification point was intentionally positioned far from the sources, $D_{diff\%}$ was out of the acceptance range, as shown in Fig. 3c–d. $D_{diff\%}$ of this verification point was -3.11% . Both situations were caused by the dosimetric parameters used to calculate g_L and F_L in the TPS, and the DVP was not accurate when the verification point was too close or far from the sources. Once $D_{diff\%}$ is greater than the tolerance, the location of the verification point needs to be checked for appropriateness. If the verification point is confirmed as appropriate, the plan may need optimization. In this way, the purpose of the dose verification for brachytherapy is achieved.

This work presented the DVP with GammaMed Plus HDR ^{192}Ir source. However, the DVP can be extended to verify brachytherapy with other type of active sources easily. All need to do is to change the fitted parameters of g_L and F_L to the parameters of the corresponding active source according to previous work [8].

CONCLUSION

An independent DVP dedicated to Eclipse TPS for brachytherapy was developed. For the GammaMed Plus HDR ^{192}Ir source, the acceptance range of the relative dose difference between the TPS and plugin was $\pm 1.03\%$. For 64 verification points, the mean of the absolute values of the relative dose difference was 0.19% . The SD of the relative dose difference was 0.17% . The entire clinical workflow of the plugin contained a few clicks. Once the plugin is under clinical practice, it will maximize efficiency and minimize human error for the brachytherapy plan check before treatment. The code of the DVP will be shared upon reasonable request.

CONFLICT OF INTEREST

The authors declare they have no conflicts of interest.

ACKNOWLEDGMENTS

We sincerely appreciate all the staff and members of the Medical Physics Laboratory of Kyoto University Graduate School of Medicine,

Department of Information Technology and Medical Engineering, and Human Health Sciences (<http://medicalphysics.hs.med.kyoto-u.ac.jp/>) for their excellent technical support and valuable comments on this study.

REFERENCES

1. Tanderup K, Menard C, Polgar C et al. Advancements in brachytherapy. *Adv Drug Deliv Rev* 2017;109:15–25.
2. Erickson BA, Demanes DJ, Ibbott GS et al. American Society for Radiation Oncology (ASTRO) and American College of Radiology (ACR) practice guideline for the performance of high-dose-rate brachytherapy. *Int J Radiat Oncol Biol Phys* 2011;79:641–9.
3. Nath R, Anderson LL, Meli JA et al. Code of practice for brachytherapy physics: report of the AAPM Radiation Therapy Committee Task Group No. 56. American Association of Physicists in Medicine. *Med Phys* 1997;24:1557–98.
4. Takahashi Y, Koizumi M, Sumida I et al. The usefulness of an independent patient-specific treatment planning verification method using a benchmark plan in high-dose-rate intracavitary brachytherapy for carcinoma of the uterine cervix. *J Radiat Res* 2012;53:936–44.
5. Carmona V, Perez-Calatayud J, Lliso F et al. A program for the independent verification of brachytherapy planning system calculations. *J Contemp Brachytherapy* 2010;2:129–33.
6. Pyry J, Keranen W. Varian APIs: a handbook for programming in the varian oncology software system. 2018. (1 April 2022, date last accessed).
7. Rivard MJ, Coursey BM, DeWerd LA et al. Update of AAPM Task Group No. 43 Report: A revised AAPM protocol for brachytherapy dose calculations. *Med Phys* 2004;31:633–74.
8. Lliso F, Perez-Calatayud J, Carmona V et al. Technical note: fitted dosimetric parameters of high dose-rate ^{192}Ir sources according to the AAPM TG43 formalism. *Med Phys* 2003;30:651–4.
9. International Commission on Radiation Units and Measurements (ICRU) Report 38. *Dose and Volume Specification for Reporting Intracavitary Brachytherapy in Gynecology*. Bethesda, MD: International Commission on Radiation Units and Measurements, 1985.