

The Investigating Image Registration Accuracy and Contour Propagation for Adaptive Radiotherapy Purposes in Line with the Task Group No. 132 Recommendation

Kamonchanok Archawametheekul¹, Chanon Puttanawarut^{2,3}, Sithiphong Suphaphong¹, Chuleeporn Jiarpinitnun¹, Siwaporn Sakulsingharoj¹, Nauljun Stansook¹, Suphalak Khachonkham¹

¹Division of Radiation Oncology, Department of Diagnostic and Therapeutic Radiology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, ²Chakri Naruebodindra Medical Institute, Mahidol University, Samut Prakan, Thailand, ³Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Abstract

Purpose: Image registration is a crucial component of the adaptive radiotherapy workflow. This study investigates the accuracy of the deformable image registration (DIR) and contour propagation features of SmartAdapt, an application in the Eclipse treatment planning system (TPS) version 16.1. **Materials and Methods:** The registration accuracy was validated using the Task Group No. 132 (TG-132) virtual phantom, which features contour evaluation and landmark analysis based on the quantitative criteria recommended in the American Association of Physicists in Medicine TG-132 report. The target registration error, Dice similarity coefficient (DSC), and center of mass displacement were used as quantitative validation metrics. The performance of the contour propagation feature was evaluated using clinical datasets (head and neck, pelvis, and chest) and an additional four-dimensional computed tomography (CT) dataset from TG-132. The primary planning and the second CT images were appropriately registered and deformed. The DSC was used to find the volume overlapping between the deformed contours and the radiation oncologist (RO)-drawn contour. The clinical value of the DIR-generated structure was reviewed and scored by an experienced RO to make a qualitative assessment. **Results:** The registration accuracy fell within the specified tolerances. SmartAdapt exhibited a reasonably propagated contour for the chest and head-and-neck regions, with DSC values of 0.80 for organs at risk. Misregistration is frequently observed in the pelvic region, which is specified as a low-contrast region. However, 78% of structures required no modification or minor modification, demonstrating good agreement between contour comparison and the qualitative analysis. **Conclusions:** SmartAdapt has adequate efficiency for image registration and contour propagation for adaptive purposes in various anatomical sites. However, there should be concern about its performance in regions with low contrast and small volumes.

Keywords: Adaptive radiotherapy, contour propagation, image registration validation

Received on: 07-12-2023

Review completed on: 12-02-2024

Accepted on: 12-02-2024

Published on: 30-03-2024

INTRODUCTION

In radiotherapy, utilization of more than one imaging modality usually involves image registration, which is a valuable method for aligning images and facilitating comparisons of information between them.^[1-4] Image registration plays a crucial role in enabling target localization from multimodal imaging. It is useful for adaptive radiotherapy purposes, where changes to the patient's anatomy compared to the treatment plan produce uncertainty in the delivered radiation dose. Image registration algorithms are very important in implementing a plan adaptation. The contours and the planned dose from the previous treatment can be mapped to the cone-beam computed

tomography (CBCT) image or second planning CT images to optimize the re-planning treatment.

Image registration algorithms have been developed to support these applications. Deformable image registration (DIR) has been increasingly used for quantifying anatomical

Address for correspondence: Dr. Suphalak Khachonkham, 270 Rama 6 Road, Rajataeue, Bangkok 10400, Thailand. E-mail: suphalak.kha@mahidol.ac.th

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Archawametheekul K, Puttanawarut C, Suphaphong S, Jiarpinitnun C, Sakulsingharoj S, Stansook N, *et al.* The investigating image registration accuracy and contour propagation for adaptive radiotherapy purposes in line with the Task Group No. 132 recommendation. *J Med Phys* 2024;49:64-72.

Access this article online

Quick Response Code:



Website:
www.jmp.org.in

DOI:
10.4103/jmp.jmp_168_23

changes and contour propagation in clinical practice. Image registration algorithms (e.g., accelerated demons algorithm and multiresolution B-spline algorithm) have been developed to support these applications which vary the notion of solution.^[3,5-8] However, published studies have shown a variety of deformation predictions from different algorithms. Despite using the same image dataset, different algorithms produce deformable vector fields with different magnitudes and opposite directions.^[6] It is therefore important to understand the limitations of each DIR algorithm and assess the performance of an algorithm before implementing it in clinical practice. The American Association of Physicists in Medicine (AAPM) Task Group No. 132 (TG-132)^[3] has provided a guideline and recommendations for the validation and quality assurance experiments of image registration algorithms used for radiotherapy. The report provides the data of a computational phantom for validating the image registration software and the tolerance recommendation. Virtual phantoms are widely used for a variety of applications. The virtual phantom allows the simulation of both simple shapes and complicated ones that overcome the limitation of the phantom design.^[9,10] It is advised that the algorithm's ability to perform contour propagation should reflect its efficacy in clinical applications.

SmartAdapt (version 16.1) is a DIR method embedded in the Eclipse™ treatment planning system (TPS) (Varian Medical Systems, Palo, Alto, California, USA). It can perform DIR on CT, positron emission tomography, and magnetic resonance imaging (MRI) images and propagate contours between several datasets. SmartAdapt is built on an accelerated demon algorithm^[11] which employs gradients in image intensity. As a result, the possibilities of using SmartAdapt for contour propagation and dose deformation for clinics with restricted sources are appealing to Eclipse™ users. Although studies have been previously conducted to assess the performance of SmartAdapt in various clinical scenarios and anatomical sites,^[11,12] these have only considered contour propagation. There is little information on comprehensively evaluating registration accuracy using the virtual phantom and contour propagation in clinical use, particularly for the new version of Eclipse™ v16.1.

The objective of this study is to investigate the registration accuracy and assess the contour propagation suitability of the SmartAdapt DIR, an application in the Eclipse TPS, using the virtual phantom obtained from the AAPM TG-132 publication and our clinical datasets. Quantitative verification will be performed using the target registration error (TRE) and Dice similarity coefficient (DSC) as metrics. The acceptable criteria for validation are per the guideline report of the AAPM TG-132 publication. In addition, center of mass (COM) shift, a newly introduced statistic for evaluating registration accuracy that was launched in SmartAdapt version 16.1, will also be utilized in this study.

MATERIALS AND METHODS

Image registration algorithm

The registration was performed using SmartAdapt, a common Varian application framework in Eclipse TPS version 16.1. The default deformable registration algorithm is derived from an accelerated demon algorithm.^[13,14] The accelerated demon registration algorithm is a technique that aligns two images by using differences in gradients within the images.^[14] The intensity-based algorithms use the image voxel data. SmartAdapt requires that a rigid registration be performed before deformable registration to adjust the position as close as possible to the destination.

Dataset selection

The dataset used in the study is divided into two groups: the AAPM TG-132 [Figure 1a-c] datasets and retrospective clinical datasets [Figure 1d]. In this study, the reference CT images without transformation were set as a stationary image, while other datasets with the addition of the translation or rotation, difference modality, and adjusting the position were set as a moving image. The AAPM TG-132 datasets (virtual phantoms generated by the ImSimQA software (Oncology Systems Limited, Shrewsbury, Shropshire, UK) used for rigid registration tests) include modality images with adding translation only and translation with rotation in terms of the geometric phantom (Cases 1–9) and anatomical phantom (Cases 10–14). In geometric datasets, the virtual phantom is defined with a simple shape resembling a cone, square, or semicircle. This differs from anatomical datasets, which are described with complex shapes that more precisely correspond to the anatomy of the human pelvic region.

For deformable registration, the recommended clinical dataset is the four-dimensional CT (4DCT) series (Case 15) from the AAPM TG-132 dataset. This consists of ten-phase images tracking anatomical movement during the respiratory phase. To conform to clinical practice guidelines, we used additional datasets. The retrospective patient images from three lung cancer cases (Cases 16–18), six head-and-neck cancer cases (Cases 19–24), and three pelvic cancer cases (Cases 25–27) treated in our clinic were chosen. All patients required plan adaptation based on pretreatment planning CT scans and the second CT scan which shows the physical change in patient anatomy compared to the initial CT scan. All patients required plan adaptation based on pretreatment planning CT images (moving images) and the second CT scan which shows the physical change in patient anatomy (stationary images) compared to the initial CT scan.

Registration Accuracy

Virtual phantom examinations obtained from AAPM TG-132^[15] are shown in Figure 1. To initiate the registration accuracy assessment, the AAPM TG-132 datasets were segmented using a specific drawing tool, image thresholding in Eclipse™ TPS, to decrease the variation of the structure. This tool generates a structure by isolating the intensity of each image voxel in a

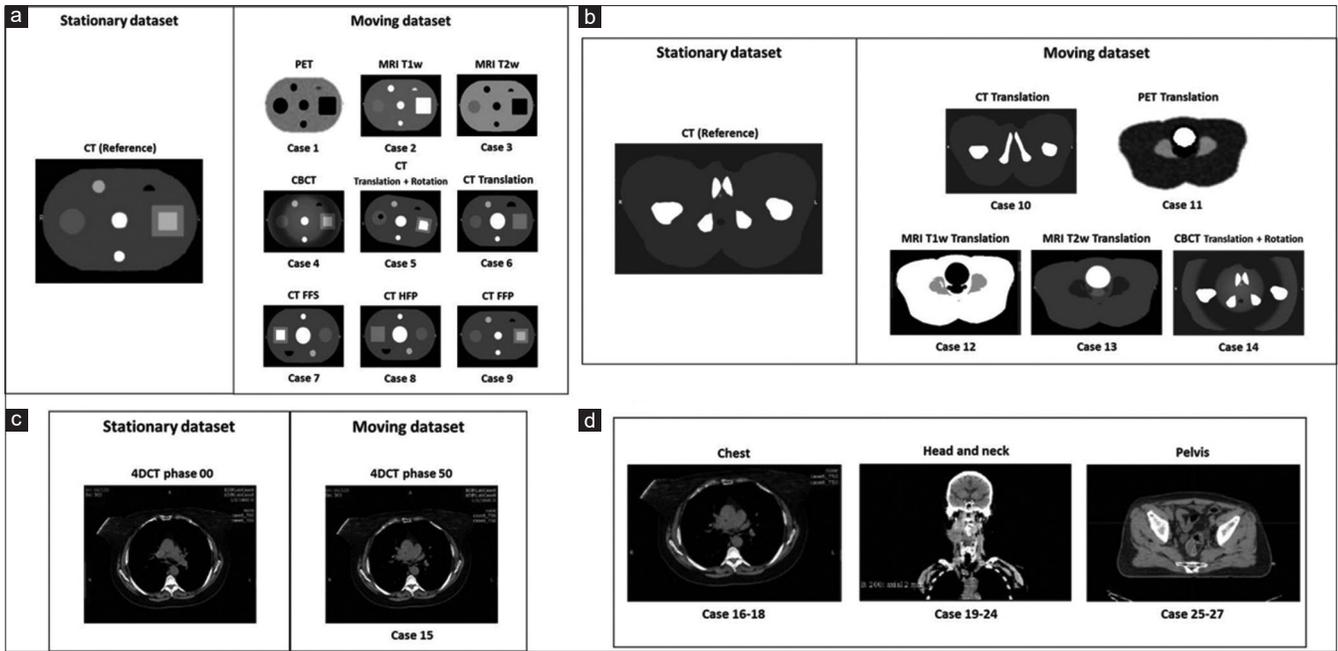


Figure 1: (a and b) Virtual phantom generated by ImSimQA software used for the rigid registration test (Task Group No. 132 [TG-132] datasets), (c) Clinical four-dimensional computed tomography datasets (TG-132 dataset) and (d) clinical datasets included in the contour propagation accuracy test

specific region of interest (ROI). After registering images, the pair of overlapping structural volumes was used to determine the DSC and COM. The DSC calculated the spatial overlap between the contours as equation (1):

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

When the contours A and B completely overlap, there is a maximum value of 1, and there is a minimum value of 0 if there is no overlap. The AAPM TG-132 recommends the tolerance for acceptable contour variation uncertainty to be a DSC value of 0.8–0.9. The COM metric, on the other hand, is an additional metric from Eclipse that is not yet routinely used to verify image registration and for which the precise tolerance has not been stated by the AAPM TG-132. The COM of an image is determined by calculating the average position of all of the pixels in the image, weighted by their intensity. To quantify the difference in COM between two images, a metric known as the COM metric is used. The COM metric can be calculated using the following equation (2):

$$COM = \left(\frac{\sum_{P_s} I(P_s)P_s}{\sum_{P_s} I(P_s)} - \frac{\sum_{P_m} I(P_m)P_m}{\sum_{P_m} I(P_m)} \right) \quad (2)$$

In equation (2), s is the stationary image and m is the moving image. P is the point of interest in the structure, and $I(P)$ is the intensity of the point P .

We also reported the individual registration error by defining the marker positions between image pairs using the landmark tracking quantitative tools of Eclipse™ TPS. The visible

markers on the phantom dataset were used to estimate the error distance as the TRE. TRE quantifies the magnitude of registration error in three dimensions and can be defined with equation (3):

$$TRE = \sqrt{(x_s - x_m)^2 + (y_s - y_m)^2 + (z_s - z_m)^2} \quad (3)$$

Here, the point (x_s, y_s, z_s) and (x_m, y_m, z_m) is point of the markers on the stationary image and the moving image, respectively.

The value of the evaluation metrics will be used to assess the performance of the registration algorithm. The TREs are recommended to not exceed 2–3 mm, according to the AAPM report.^[3]

Contour propagation accuracy

Quantitative evaluation

Clinical datasets were used to test the performance of the image registration software using real patient CT images, as shown in Figure 1c and d. This study focuses on the most common treatment sites which use DIR: the chest, head and neck, and pelvis.^[15-17] The contour of the gross tumor volume (GTV) and selected organs at risk (OARs) were evaluated. Initial rigid registration is required before performing a DIR for Eclipse image registration. Image registration is performed based on the similarity index and modified demon algorithm. The DIR-generated structures were compared with manually drawn contours from radiation oncologists (ROs) as shown in Figure 2 in terms of their overlapping contours and shifts of the three-dimensional centers of mass. DSC was used to assess the contour volume overlap between the registered images, while the shift in volume centric was determined using COM.

Qualitative evaluation

A qualitative evaluation was performed by an experienced RO to specify the clinical capability of the DIR function for contour propagation. A qualitative score was given to assess the dissimilarity between the deformed contours and the re-contour structures drawn by the RO. The agreement between the two structures was defined as a score from 1 to 3, from highest to lowest. A score of 1 describes propagated contours that do not require modification compared to the re-contour structure, 2 describes propagated contours that require negligible modification, and 3 describes impractical propagated contours that require major modification.^[11,18]

RESULTS

Registration accuracy

A summary of registration accuracy between the reference CT images and multi-modality images for the AAPM TG-132 virtual phantom (Cases 1–14) is shown in Figures 3-5. The mean ± standard deviation (SD) for all

cases and structures was $DSC\ 0.97 \pm 0.02$ (range: 0.91–0.99), $COM\ 0.61 \pm 0.75\ mm$ (range: 0.07–1.97 mm), and $TRE\ 0.29 \pm 0.17\ mm$ (range: 0.21–0.36 mm).

The results of the contour overlaying analysis show a DSC value above the acceptable tolerance in the report ($DSC > 0.8$)^[3] for all structures. The results of registration show the average range of the DSC value to be 0.91–1.00, as shown in Figure 3. In addition, the COM metric was measured using a reference image and a moving image [Figure 4]. The corresponding average COM value was $<1\ mm$ in all directions except in the CBCT image (Case 4), the CT image with added translation and rotation (Case 5), and the T2W MRI (Case 13). The COM values were up to 1.97, 1.20, and 1.36 mm, respectively.

The average TRE was calculated from the point pair of each dataset after the registration process. The results show that the accuracy of registration does not exceed the tolerance recommendation of the report.^[3] The mean TRE is not exceeded 0.4 mm while the maximum TRE is not exceeded 0.8 mm for all structures [Figure 5].

Contour propagation accuracy

Clinical four-dimensional computed tomography (Task Group No. 132)

A summary of contour propagation accuracy between the 4DCT phase 00 and 4DCT phase 50 (Case 15) is shown in Table 1. For this case, the demon algorithm was used to register the two extremely variable respiratory phases of 4DCT images. The mean ± SD for all structures was 0.89 ± 0.07 (range: 0.81–0.97) for DSC and $2.29 \pm 1.21\ mm$ (range: 1.26–3.91 mm) for COM. The DSCs for the esophagus, heart, spinal cord, and left and right lungs presented a satisfactory result of contour propagation. Although the lowest DSC value of the esophagus is due to its ambiguous margins and movability, an acceptable level of DSC score was present ($DSC > 0.8$). Concurrently, the COM shift metric shows quite well scores below 5 mm in all directions.

Clinical thoracic

A summary of contour propagation accuracies between the initial CT images and re-plan CT images for lung cancer (Cases

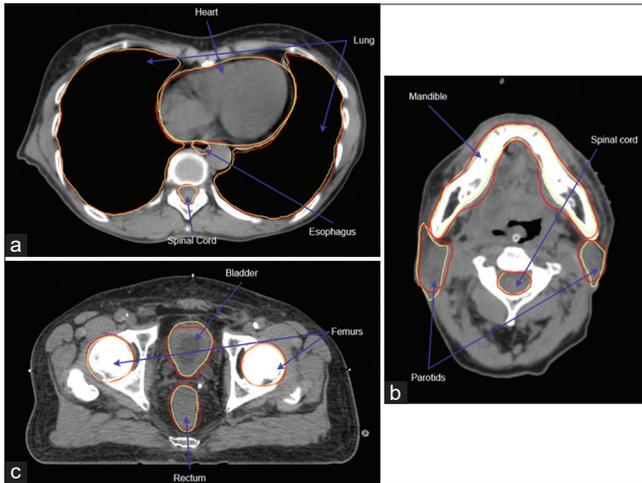


Figure 2: Illustration depicting contour images for each region: Chest (a), Head and Neck (b) and Pelvis(c), showing the comparison between the deformed structure (in red) and the re-contoured structure (in yellow)

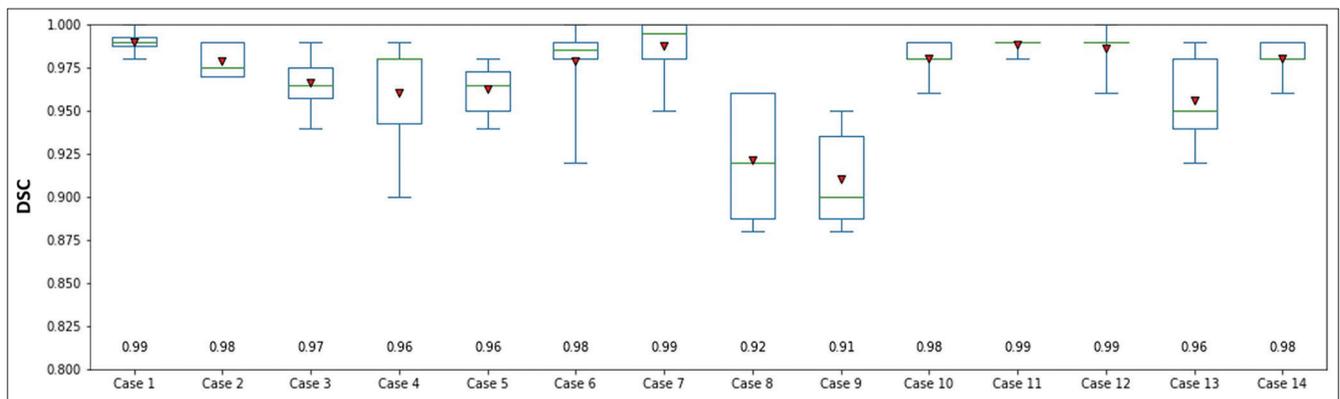


Figure 3: Boxplot with mean (red triangle and label value) of Dice similarity coefficient values of registration results between the reference computed tomography images and multi-modality images of the American Association of Physicists in Medicine Task Group No. 132 virtual phantom. DSC: Dice similarity coefficient

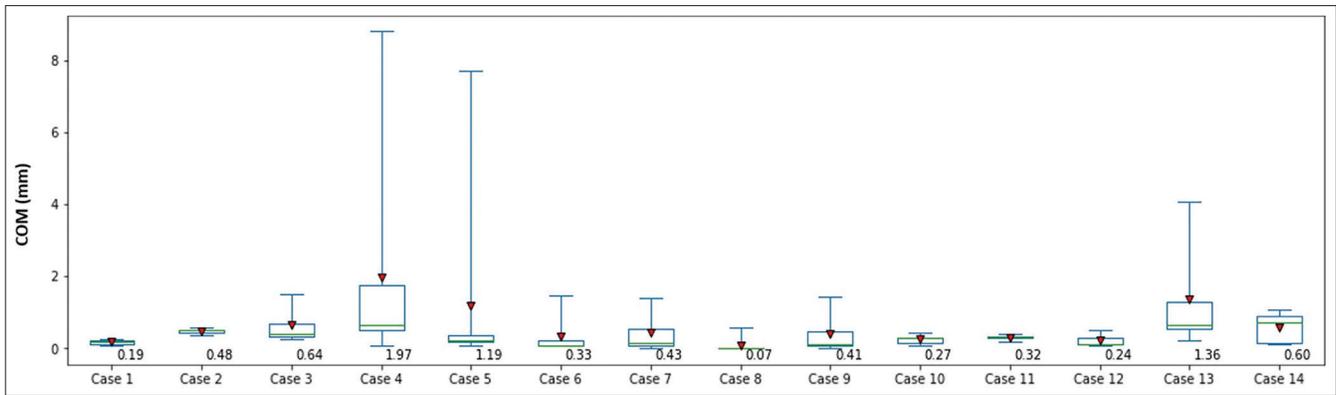


Figure 4: Boxplot with mean (red triangle and label value) of center of mass shift values (mm) of registration results between the reference computed tomography images and multi-modality images of American Association of Physicists in Medicine ask group no. 132 virtual phantom. COM: Center of mass

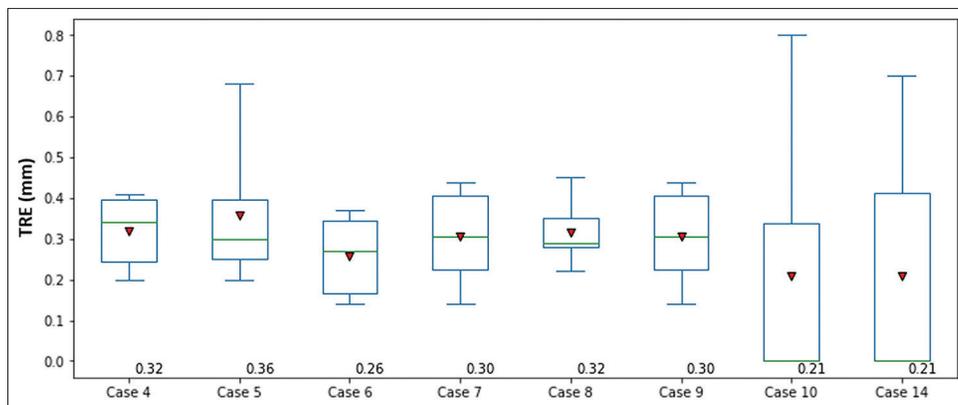


Figure 5: Boxplot with mean (red triangle and label value) of target registration error values (mm) of registration results between the reference computed tomography images and multi-modality images of American Association of Physicists in Medicine Task Group No. 132 virtual phantom. TRE: Target registration error

Table 1: The Dice similarity coefficients and center of mass shift of clinical four-dimensional computed tomography from Task Group No. 132 dataset (Case 15) demonstrates overlapping values between the re-contour structure and the deformed structure of the thoracic region

Structure	DSC	COM (mm)
Esophagus	0.81	1.49
Heart	0.93	2.50
Lungs	0.97	1.26
Spinal cord	0.85	3.91
Mean±SD	0.89±0.07	2.29±1.21

DSC: Dice similarity coefficient, COM: Center of mass, SD: Standard deviation

16–18) is shown in Table 2. The mean ± SD DSC for all data was 0.82 ± 0.28 (range: 0.54–0.96), and the COM was 4.43 ± 4.16 mm (range: 0.79–13.56 mm). SmartAdapt demonstrated good performance for the lung and heart, receiving average DSC scores of 0.93 ± 0.03 and 0.89 ± 0.09 , respectively [Table 2]. Although some structures had a DSC value below 0.8, the DSC value is not <0.75 except for in one case, the structure of the GTV_N in Case 17. This low DSC

value is due to the impact of a small GTV volume (<5 cm³) and the low contrast of the region. In the thoracic region, the supplementary COM measuring was on average >4 mm for the esophagus and heart and might reach 6 mm for the lungs. Furthermore, the maximum displacement of the heart and lung structures gave a remarkably high COM value of up to 12–13 mm. This is due to the respiratory cycle effect.

Clinical head and neck

A summary of the contour propagation accuracy between the initial CT images and the re-plan CT images for head-and-neck cancers (Cases 19–24) is shown in Table 3. The mean ± SD for all data was 0.83 ± 0.25 (range: 0.40–0.93) for DSC and 2.45 ± 1.91 mm (range: 0.54–8.42 mm) for COM. Table 3 shows that the detailed DSC values in each case are within the range of 0.70–0.93, except for the GTV in Case 24 (DSC = 0.4). The low performance for GTV may be explained by the volume dependence of the DSC parameter. A small volume of organs of interest may be impacted more than the large one.^[19] The COM metric for the head and neck had an average value of <2 mm for the brainstem and eye. Displacement ranges of 2–3 mm for the mandible and parotids, while the spinal cord and GTV have displacement ranges of 3–4 mm. There were no

Table 2: The Dice similarity coefficients of three patients demonstrate overlapping values between the re-contour structure and the deformed structure of the thoracic region

Structure	DSC				COM (mm)			
	Case 16	Case 17	Case 18	Mean±SD	Case 16	Case 17	Case 18	Mean±SD
Esophagus	0.75	0.83	0.76	0.78±0.04	2.52	2.32	9.86	4.90±4.30
Heart	0.79	0.92	0.95	0.89±0.09	13.56	1.80	0.99	5.45±7.03
Lungs	0.90	0.92	0.96	0.93±0.03	12.53	4.20	1.42	6.05±5.78
Spinal cord	0.85	0.86	0.76	0.82±0.06	2.55	1.90	5.24	3.13±1.84
GTV_P	0.78	-	0.84	0.81±0.04	7.38	-	2.33	4.86±3.57
GTV_N	-	0.54	0.78	0.66±0.17	-	1.88	0.79	1.33±0.77
Mean±SD	0.81±0.06	0.81±0.16	0.84±0.09		7.64±5.35	2.42±1.01	3.44±3.55	

DSC: Dice similarity coefficient, COM: Center of mass, SD: Standard deviation, GTV: Gross tumor volume

significant variations in the COM values between the OARs of the head-and-neck area including target organs, as indicated by the low SD.

Clinical pelvic

A summary of the contour propagation accuracy between the initial CT images and re-plan CT images for pelvic cancer (Cases 25–27) is shown in Table 4. The mean ± SD for all data was 0.81 ± 0.22 (range: 0.67–0.94) for DSC and 4.80 ± 4.36 mm (range: 0.88–16.41 mm) for COM. The low contrast area is a major challenge in aligning the position of organs in the pelvis region, which is a limitation of demon's force driving based on the intensity difference between the two images. However, the DSC results show good outcomes for the average DSC in each organ. Although the DSC value is below the report recommendation (DSC >0.8–0.9), the average value for all structures exceeds 0.7 [Table 4]. The findings demonstrate that the femur bones were consistently well matched and the COM was <2 mm. The pelvic target organs were revealed to have a COM of 3.07 mm, whereas the bladder, rectum, and small bowel all had COMs of above 4 mm. The most variable results were obtained from the rectum and small bowel, as shown by the comparatively high SD for the centric misalignment measurements of up to 7.28 mm and 4.84 mm, respectively.

Qualitative evaluation

The DSC and the qualitative score of the deformed structures and the re-contour structures evaluated by an experienced RO are shown in Figure 6. A score of 1 (40%) indicates an acceptable ROI without modification, and 2 (38%) indicates an acceptable ROI with negligible modifications. Only 25 of 112 ROIs (22%) had a score of 3, indicating that they require major modifications. Most of the OARs were scored as 1 (51%), with some of the OARs scored as 2 (25%) or 3 (24%), whereas only 23% of the target ROIs scored 1.

DISCUSSION

The AAPM TG-132 report details the fundamental concepts of validating image registration, a list of acceptable tolerances, and a comprehensive guide on constructing a virtual phantom dataset. Furthermore, the report advocates for the inclusion

of clinical datasets in the validation process, emphasizing flexibility to cater to individual department needs, without specifying particular data or endorsing a gold standard method. In line with this report, our study assesses the practicality of image registration solutions within the Eclipse TPS, utilizing the TG-132 virtual phantom available for download and SmartAdapt's contour propagation capabilities on common anatomical sites. To understand the limitations of these features and accommodate for appropriate uncertainties in most regions for the task, the user should perform the validation before implementing it in clinical practice.

The average quantitative measurements for the AAPM TG-132 virtual phantom across all cases and structures were DSC 0.97, COM 0.61 mm, and TRE 0.29. The AAPM TG-132 considers DSC values between 0.80 and 0.90 and a TRE within the range of 2–3 mm as acceptable.^[3] Kadoya *et al.*^[20] have introduced a physical geometric phantom for the assessment of commercial DIR software, and Wu *et al.*^[21] developed a user-friendly physical phantom for testing both rigid and DIR accuracy. These studies reported the mean DSC values that exceeded 0.80 for DIR accuracy, which is consistent with our own study results. The low DSC value is also observed in Case 8 and Case 9. This is attributed to variations in positioning between the supine and prone datasets. These differences can result in variations in structure and location between the two positions, leading to a reduced overlap in structural volume. In addition, the Eclipse registration package provides a quantitative parameter COM to define the centric structure that deviates from the reference image. In the virtual phantom study, structures with partial volume loss were observed in certain datasets, such as CBCT (Case 4) and CT with translation and rotation (Case 5). In these cases, the corresponding center of mass (COM) value appears to be higher in the superior-inferior (SI) axis, exceeding 1 mm in all directions. Additionally, in Case 13, significant displacement is often noted in the SI direction, attributable to variations in pixel volume among modalities.

In the contour propagation accuracy test, SmartAdapt gave satisfactory results for the head-and-neck and thoracic regions and performed fairly for the pelvic area with noticeable erroneous registration in some image slices [Figure 7a]. In the

Table 3: The Dice similarity coefficients of six patients demonstrate overlapping values between the re-contour structure and the deformed structure of the head-and-neck region

Structure	DSC						COM (mm)							
	Case 19	Case 20	Case 21	Case 22	Case 23	Case 24	Mean±SD	Case 19	Case 20	Case 21	Case 22	Case 23	Case 24	Mean±SD
Brainstem	0.87	0.88	0.92	0.88	-	0.89	0.89±0.02	0.54	0.85	0.64	1.24	-	1.49	0.95±0.40
Spinal cord	0.81	0.82	0.83	0.81	0.91	0.87	0.84±0.04	2.54	0.81	8.42	3.93	2.34	2.00	3.34±2.68
Eyes	0.86	0.90	0.85	0.89	0.93	0.76	0.84±0.06	2.10	0.69	0.83	0.72	0.84	3.69	1.48±1.21
Mandible	0.80	0.84	0.86	0.80	-	0.80	0.82±0.03	1.63	0.79	1.00	5.68	-	3.66	2.55±2.08
Parotids	0.85	0.87	0.81	0.76	0.91	0.82	0.84±0.05	2.46	1.57	2.58	4.11	1.19	3.93	2.64±1.19
GTV	0.70	0.90	0.89	0.78	-	0.40	0.73±0.20	6.21	1.26	3.54	3.52	-	4.19	3.74±1.77
Mean±SD	0.82±0.06	0.87±0.03	0.86±0.04	0.82±0.05	0.92±0.01	0.76±0.18		2.58±1.92	1.00±0.34	2.83±2.96	3.20±1.88	1.46±0.79	3.16±1.12	

DSC: Dice similarity coefficient, COM: Center of mass, SD: Standard deviation, GTV: Gross tumor volume

pelvic region, deformation errors tend to occur in establishes with a low tissue intensity variation area, such as the bladder, rectum, or bowel. However, this issue is typically not detected in bony structures. This could be explained by the limitations of the algorithm implemented in SmartAdapt software. The demon's force is based on the gradient of the image, which decreases the performance in low-contrast conditions.^[14,22,23] Compared to other studies, the average DSC of 0.75 would be acceptable according to a study by Loi *et al.*^[24] of mostly failed pelvic sites. We also found that the relatively low DSC values (DSC: 0.85) and the relatively high COM of the spinal cord, this associated with deformed structures exhibiting alterations on the superior side adjacent to the organ (brain) characterized by a low image gradient. These limitations are inherent to the accelerated demons algorithm. For the thoracic region, the DSC score for the heart and lungs complies with the TG-132 recommendation (mean DSC: 0.89 and 0.93, respectively) and is slightly lower for the movable structure, esophagus (DSC: 0.75-0.83, mean DSC of 0.78). Most structures in the HN region performed well in quantitative evaluations, with DSC values exceeding 0.80. However, the varied range of DSC values for some structures of HN case was present and slightly lower than expected. Some structures with very low DSC values, such as the GTV in HN (Case 24) shown in Table 3, could be supported by an idea of DSC calculation which depends on the volume of the structure. Structures with a very small volume ($\leq 3 \text{ cm}^3$) generally show a lower DSC value than a large volume structure, and uncertainty is increased due to interpolation.^[19,25] In clinical datasets, the analysis of COM values is a challenging task because multiple factors affect the values. Moreover, this COM parameter is not typically reported with established detailed tolerances, hence these data are provided for observation purposes. Compared to previous research, Saroudis *et al.*^[23] applied COM and DSC parameters to examine the impact of deformation outcomes on changes in organ displacement. The result showed that larger shifts lead to greater failure of the deformation result. According to our findings, we observed a COM shift of more than 10 mm resulting in a slight decrease in DSC values. However, since the result of each deformation depends on an extensive number of factors, we are unable to reach definitive conclusions using a single element alone.

Although quantitative metrics such as TRE and DSC provide exact numerical data, these metrics may not provide sufficient insight into whether they are clinically acceptable. The qualitative scoring system was introduced by Hardcastle *et al.*^[18] The score is the clinical utility as determined by an expert physician, as described in the methods. In this study, the qualitative scores are inconsistently correlated with expert RO assessments and DSC values, as shown in Figure 6. Some structures with high DSC values (DSC above 0.80) support the clinician's assessment that the deformation is effective without modification, while others require minor or major modification. On the other hand, several structures that specialists have identified as effective deformers

Table 4: The Dice similarity coefficients of three patients demonstrate overlapping values between the re-contour structure and the deformed structure of the pelvic region

Structure	DSC				COM (mm)			
	Case 25	Case 26	Case 27	Mean±SD	Case 25	Case 26	Case 27	Mean±SD
Bladder	0.85	0.67	0.78	0.77±0.09	6.38	3.42	5.07	4.96±1.49
Femurs	0.94	0.94	0.94	0.94±0.00	1.18	0.88	1.47	1.18±0.30
Rectum	-	0.72	0.74	0.73±0.01	-	16.41	6.11	11.26±7.28
Small bowel	0.78	0.80	0.80	0.79±0.01	11.08	1.69	4.34	5.70±4.84
GTV	0.79	0.87	0.77	0.81±0.05	3.59	1.19	4.43	3.07±1.68
Mean±SD	0.84±0.84	0.80±0.11	0.81±0.08		5.56±4.25	4.72±6.61	4.29±1.72	

DSC: Dice similarity coefficient, COM: Center of mass, SD: Standard deviation, GTV: Gross tumor volume

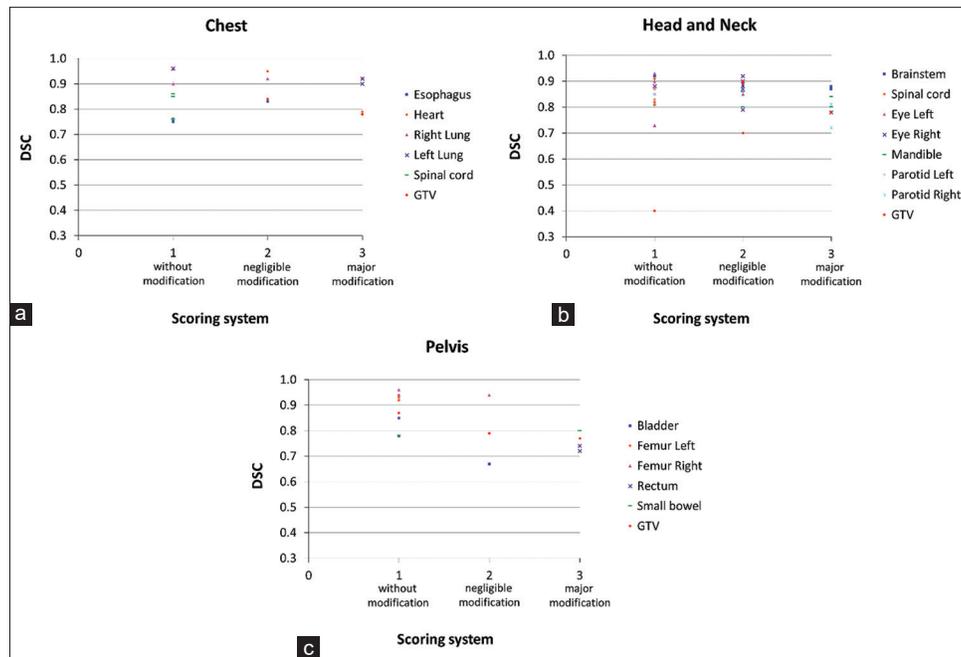


Figure 6: Scatter plot of relationship between the Dice similarity coefficient values of the re-contour structure and the deformed structure related to qualitative score results for each region: Chest (a), Head and Neck (b) and Pelvis (c). DSC: Dice similarity coefficient

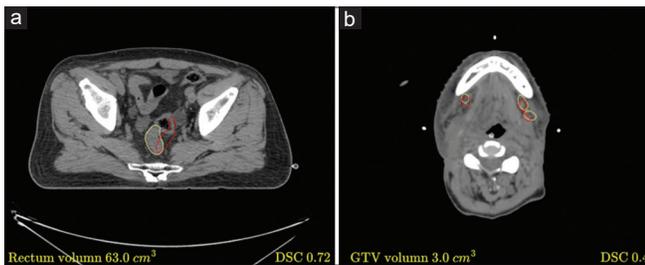


Figure 7: (a) The result of image registration in a pelvic patient (Case 26) presents the rectum mismatch between the deformed structure (red) and re-contour structure (yellow) in the low contrast region. (b) The overlap of the deformed structure and the re-contour structure of small volume of gross tumor volume in a head-and-neck patient (Case 24) represents a very low Dice similarity coefficient value. DSC: Dice similarity coefficient

have low DSC values. This finding demonstrates that the relationship between the two parameters is complex and influenced by various circumstances. The DSC values shown are relatively high in contrast to the qualitative scores and

appear to require major modification, while some structures with very low DSC values were assessed as acceptable without modification (score 1). Although the two analysis methods were incompatible, they can be used to support explanations. The results of the qualitative analysis show that 78% of all structures were either acceptable without modification or required insignificant adaptation. Most of the structures that needed to be modified, such as GTV, were small or distorted in shape, as shown in Figure 7b. The expert RO also noted that corrections are often made for the 2–3 superior and inferior slices of the CT image. These RO scores represent the assessment of the clinical utility and reliability of automatically propagated contour for adaptive planning workflow. However, cautious expert inspection of the structures generated by DIR remains important, particularly in cases of significant deformation variation or poor image contrast. In future work, advanced validation should explore the impact of DIR uncertainty on dosimetric accuracy.

CONCLUSIONS

DIR is offered under SmartAdapt, integrated in the recent release of Eclipse TPS version 16.1. Based on our validation using virtual phantoms and clinical datasets, SmartAdapt has achieved registration accuracy and adequate contour propagation for adaptive purposes in the chest, head-and-neck, and pelvic regions. However, except for some contours that are likely attributed to low-contrast regions and contain small volumes, a careful review of the contour propagation by an expert physician is recommended.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Paganelli C, Meschini G, Molinelli S, Riboldi M, Baroni G. "Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats". *Med Phys* 2018;45:e908-22.
- Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, *et al.* Deformable image registration for radiation therapy: Principle, methods, applications and evaluation. *Acta Oncol* 2019;58:1225-37.
- Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM radiation therapy committee task group no. 132. *Med Phys* 2017;44:e43-76.
- Oh S, Kim S. Deformable image registration in radiation therapy. *Radiat Oncol J* 2017;35:101-11.
- Lawson JD, Schreiber E, Jani AB, Fox T. Quantitative evaluation of a cone-beam computed tomography-planning computed tomography deformable image registration method for adaptive radiation therapy. *J Appl Clin Med Phys* 2007;8:96-113.
- Pukala J, Johnson PB, Shah AP, Langen KM, Bova FJ, Staton RJ, *et al.* Benchmarking of five commercial deformable image registration algorithms for head and neck patients. *J Appl Clin Med Phys* 2016;17:25-40.
- Castadot P, Lee JA, Parraga A, Geets X, Macq B, Grégoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. *Radiat Oncol* 2008;89:1-12.
- Kadoya N, Miyasaka Y, Yamamoto T, Kuroda Y, Ito K, Chiba M, *et al.* Evaluation of rectum and bladder dose accumulation from external beam radiotherapy and brachytherapy for cervical cancer using two different deformable image registration techniques. *J Radiat Res* 2017;58:720-8.
- Singh K, Kirby N, Pouliot J. A three-dimensional head-and-neck phantom for validation of multimodality deformable image registration for adaptive radiotherapy. *Med Phys* 2014;41:121709.
- Pukala J, Meeks SL, Staton RJ, Bova FJ, Mañon RR, Langen KM. A virtual phantom library for the quantification of deformable image registration uncertainties in patients with cancers of the head and neck. *Med Phys* 2013;40:111703.
- Ramadaan IS, Peick K, Hamilton DA, Evans J, Iupati D, Nicholson A, *et al.* Validation of Varian's Smartadapt® deformable image registration algorithm for clinical application. *Radiat Oncol* 2015;10:73.
- Vickress J, Rangel Baltazar MA, Afsharpour H. Evaluation of Varian's Smartadapt for clinical use in radiation therapy for patients with thoracic lesions. *J Appl Clin Med Phys* 2021;22:150-6.
- Thirion JP. Image matching as a diffusion process: An analogy with Maxwell's demons. *Med Image Anal* 1998;2:243-60.
- Wang H, Dong L, O'Daniel J, Mohan R, Garden AS, Ang KK, *et al.* Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol* 2005;50:2887-905.
- Hussein M, Akintonde A, McClelland J, Speight R, Clark CH. Clinical use, challenges, and barriers to implementation of deformable image registration in radiotherapy – The need for guidance and QA tools. *Br J Radiol* 2021;94:20210001.
- Kadoya N, Kito S, Kurooka M, Saito M, Takemura A, Tohyama N, *et al.* Factual survey of the clinical use of deformable image registration software for radiotherapy in Japan. *J Radiat Res* 2019;60:546-53.
- Yuen J, Barber J, Ralston A, Gray A, Walker A, Hardcastle N, *et al.* An international survey on the clinical use of rigid and deformable image registration in radiotherapy. *J Appl Clin Med Phys* 2020;21:10-24.
- Hardcastle N, van Elmpst W, De Ruysscher D, Bzdusek K, Tomé WA. Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy. *Radiat Oncol* 2013;8:243.
- Kumarasiri A, Siddiqui F, Liu C, Yechieli R, Shah M, Pradhan D, *et al.* Deformable image registration based automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting. *Med Phys* 2014;41:121712.
- Kadoya N, Sakulsinhroj S, Kron T, Yao A, Hardcastle N, Bergman A, *et al.* Development of a physical geometric phantom for deformable image registration credentialing of radiotherapy centers for a clinical trial. *J Appl Clin Med Phys* 2021;22:255-65.
- Wu RY, Liu AY, Wisdom P, Zhu XR, Frank SJ, Fuller CD, *et al.* Characterization of a new physical phantom for testing rigid and deformable image registration. *J Appl Clin Med Phys* 2019;20:145-53.
- Alvarez L, Weickert J, Sánchez J. Reliable estimation of dense optical flow fields with large displacements. *Int J Comput Vis* 2000;39:41-56.
- Sarudis S, Karlsson A, Bibac D, Nyman J, Bäck A. Evaluation of deformable image registration accuracy for CT images of the thorax region. *Phys Med* 2019;57:191-9.
- Loi G, Fusella M, Lanzi E, Cagni E, Garibaldi C, Iacoviello G, *et al.* Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: A multi-institutional study. *Med Phys* 2018;45:748-57.
- Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, *et al.* Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: A multi-expert study. *Phys Med Biol* 2011;56:4557-77.