Original Article

Performance Evaluation of Deformable Image Registration Systems – SmartAdapt® and Velocity™

M. Anil Kumar, Raghavendra Hajare, Bhakti Dev Nath[1], K. K. Sree Lakshmi, Umesh M. Mahantshetty

Department of Radiation Oncology, Homi Bhabha Cancer Hospital and Research Centre, Visakhapatnam, Andhra Pradesh, [1]Department of Radiation Oncology, Tata Medical Centre, Kolkata, West Bengal, India

Abstract

Aim: To commission and validate commercial deformable image registration (DIR) systems (SmartAdapt® and Velocity™) using task group 132 (TG-132) digital phantom datasets. Additionally, the study compares and verifies the DIR algorithms of the two systems. **Materials and Methods:** TG-132 digital phantoms were obtained from the American Association of Physicists in Medicine website and imported into SmartAdapt® and Velocity™ systems for commissioning and validation. The registration results were compared with known shifts using rigid registrations and deformable registrations. Virtual head and neck phantoms obtained online (DIR Evaluation Project) and some selected clinical data sets from the department were imported into the two DIR systems. For both of these datasets, DIR was carried out between the source and target images, and the contours were then propagated from the source to the target image data set. The dice similarity coefficient (DSC), mean distance to agreement (MDA), and Jacobian determinant measures were utilised to evaluate the registration results. **Results:** The recommended criteria for commissioning and validation of DIR system from TG-132 was error <0.5*voxel dimension (vd). Translation only registration: Both systems met TG-132 recommendations except computed tomography (CT)-positron emission tomography registration in both systems (Velocity ~1.1*vd, SmartAdapt ~1.6*vd). Translational and rotational registration: Both systems failed the criteria for all modalities (For velocity, error ranged from 0.6*vd [CT-CT registration] to 3.4*vd [CT-cone-beam CT (CBCT) registration]. For SmartAdapt® the range was 0.6*vd [CT-CBCT] to 3.6*vd [CT-CT]). Mean ± standard deviation for DSC, MDA and Jacobian metrics were used to compare the DIR results between SmartAdapt® and Velocity™. **Conclusion:** The DIR algorithms of SmartAdapt® and Velocity™ were commissioned and their deformation results were compared. Both systems can be used for clinical purpose. While there were only minimal differences between the two systems, Velocity™ provided lower values for parotids, bladder, rectum, and prostate (soft tissue) compared to SmartAdapt. However, for mandible, spinal cord, and femoral heads (rigid structures), both systems showed nearly identical results.

Keywords: Commissioning, deformable image registration, digital phantom datasets, image registration algorithms, SmartAdapt®, task group 132, validation, velocity™

Received on: 06-12-2023 **Review completed on: 27-02-2024** **Accepted on: 20-03-2024** **Published on: 25-06-2024**

Introduction

Adaptive radiotherapy[1] has seen a rapid rise in the use of deformable image registration (DIR)[2] in recent times. Before clinical usage, every DIR application has to be commissioned and evaluated. Commissioning ideally involves investigating the precision and repeatability of the image registration methods across all sites and imaging modalities (such as computed tomography [CT]-CT, CT-positron emission tomography [CT-PET], CT-magnetic resonance imaging [CT-MRI] etc.). The posttransfer picture data quality, spatial integrity, image orientation, registration and/or deformation correctness, and other system functioning must all be verified during the validation phase of the quality assurance (QA) of the image registration process. Due to a lack of documentation from the commercial systems, commissioning and validation of the DIR systems are a challenge.

The American Association of Physicists in Medicine, Task Group 132 (AAPM TG-132) report[3] recommends the use of

Access this article online

Quick Response Code:

Website: www.jmp.org.in

DOI: 10.4103/jmp.jmp_167_23

Address for correspondence: Mr. M. Anil Kumar, Department of Radiation Oncology, Homi Bhabha Cancer Hospital and Research Centre, Aganampudi, Gajuwaka Mandal, Visakhapatnam - 530 046, Andhra Pradesh, India.
E-mail: murugaya.anil@gmail.com

How to cite this article: Kumar MA, Hajare R, Nath BD, Lakshmi KK, Mahantshetty UM. Performance evaluation of deformable image registration systems – SmartAdapt® and Velocity™. J Med Phys 2024;49:240-9.

physical phantom system end-to-end tests, digital phantom tests, and clinical data tests procedures for commissioning and validation of DIR systems. Physical phantom tests are required to assure proper data representation, image transmission, and integrity verification between imaging equipment, image registration systems, and other radiation systems that use the results of image registration. The accuracy of image registration[4] can be tested under controlled conditions using digital or physical phantoms. Clinical data tests ensure the system's accuracy on samples of the images anticipated for clinical application.

Image registration[5,6] is the process of aligning homologous points, typically represented by image voxels, across multi-temporal, mono-or multi-modal, anatomical, or functional image datasets. This procedure is described by a mathematical transformation that grows in complexity as the dissimilarity between two images increases. Image registrations are two types: rigid registration and deformable registration.

Rigid registration is a global match between image sets that preserves the relative distance between every pair of points from the patient's anatomy. It can include translation or/ and rotation in all direction for which, a 4 × 4 homogeneous transformation matrix can be described that registers a coordinate system A to B, as shown in equation 1.[7]

$$\begin{bmatrix} A_x \\ A_y \\ A_z \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} B_x \\ B_y \\ B_z \\ 1 \end{bmatrix} \quad (1)$$

The coordinates of points in two different frames of reference are represented by vectors A and B, whereas translations are represented by vector T and rotational transformations are represented by matrix R. The final row of ones and zeroes is for uniformity in matrix operations with no physical significance.

The deformable registration, on the other hand, is a process in which a moving image data set (M) is registered or deformed to match a target image data set (T) using a deformation vector field (DVF). The DVF defines the motion of every image voxel from M to T. The registration usually operates using an optimization algorithm which works on a transformation model to increase the similarity measure function of the two image data sets. The transformation models currently available in market include spline and demons, elastic, fluid, finite element model, and free form deformations.

Jamema *et al*. validated the DIR (SmartAdapt® v13.6) by using physical phantoms and clinical images of various disease sites.[8] Pukala *et al*. constructed a library of 10 virtual head and neck (H and N) phantoms for the quantification of uncertainties in different DIR algorithms.[9] Latifi *et al*. provided a detailed step-by-step description of the quantitative tests for validation of the medical image registration process, which are enumerated in the AAPM TG-132 report.[7]

Our institution has two DIR systems – SmartAdapt® which is available in Varian's Eclipse™ treatment planning system v15.6 (Varian Medical Systems Inc., Palo Alto, CA) and Velocity™ v4.1[10] which is a standalone system, specifically for DIR. Velocity™ is a commercially available DIR software that uses a modified B-Spline deformable with a mutual information-based matching algorithm for DIR SmartAdapt® on the other hand is a DIR application included within the Eclipse treatment planning system which employs demons' method, that utilises gradients in image intensity values to drive registration. Velocity™ is capable of both adaptive re-contouring and deformable dose accumulation whereas SmartAdapt® is only capable of the former.

Since both systems can perform deformation, we have commissioned and compared their performances based on their ability to propagate the structures on to the deformed image set (contour comparison).

Materials and Methods

For commissioning and validation, we followed AAPM TG-132 Report framework and then comparison of both systems was carried out by evaluating contour propagation accuracy using virtual H and N phantoms and clinical dataset from our institute of H and N and gynaecology patients.

Task group 132 digital phantoms

AAPM TG-132 recommends the use of digital phantoms for commissioning and validation which is freely available for download in the AAPM website.[11] These digital phantoms as shown in Figure 1 were generated specifically for the purpose of use in commissioning and QA programs for image registration accuracy tests. The datasets were generated using ImSimQA™ software (Oncology System Limited, UK [OSL, www.imsimqa.com]) in different modalities (i.e. CT, PET, MRI, cone-beam CT [CBCT]) with known translational, rotational shifts, and deformation with added noise. The following registrations were performed on these digital phantoms which were imported into the systems.

Rigid translations

Table 1 shows the known shifts and the corresponding translational values (T values) in two datasets of the digital phantoms. As different modality images (CT, PET, and magnetic resonance [MR]) share the same frame of reference as CT phantom images, these nominal T values within each dataset group remain constant. Any difference between the known T value from Table 1 and those supplied by the registration system is caused by inaccuracies of registration along the cardinal axes (x, y, and z), which can be compared to the relevant image voxel sizes.

Rigid translation and rotation

The computations for transformation (i.e. T [x, y, z] values) are more difficult when rotations are involved, because rotations are applied before translations. As a result, the dataset's known shift values vary not only in their signs but also in their magnitude.

Table 1: Rigid registration tests - translations only. The data are combined from Table 5 and 6 in the TG-132 report

Stationary dataset	Moving dataset	Known shifts	Known - *t* (x, y, z) mm
BPD 2 (CT)	BPD 1 (CT) BPD 1 (PET) BPD 1 (MR1) BPD 1 (MR2) BPD 1 (CBCT)	Dataset 2 is shifted with respect to Dataset 1 by 10 mm to patient lateral, 5 mm anterior, 15 mm superior	−10, 5, −15
BAPD 1 (CT)	BAPD 2 (CT) BAPD 3 (PET) BAPD 4 (MR1) BAPD 5 (MR2)	Datasets 2, 3, 4, 5, 6 shifted with respect to Dataset 1 by 3 mm lateral, 5 mm anterior, 12 mm superior	3, −5, 12

BPD: Basic phantom dataset, CT: Computed tomography, PET: Positron emission tomography, CBCT: Cone-beam CT, MR: Magnetic resonance, BAPD: Basic anatomical phantom dataset

Table 2: Rigid registration tests - translations and rotations

Stationary dataset	Moving dataset	Known shifts	Known rotations	Known - *t* (x, y, z) mm**
BPD3 (CT)	BPD 1 (CT) BPD 1 (PET) BPD1 (MR1) BPD1 (MR2) BPD1 (CBCT)	Dataset 3 is shifted with respect to Dataset 1 by 5 mm to patient lateral, 15 mm anterior, 20 mm superior	−5° around X-axis, 8° Y, 10° Z	−5.07, 17.29, −18.06

**Known *t* values were adopted from Latifi *et al*. BPD: Basic phantom dataset, BAD: Basic anatomical dataset, CT: Computed tomography, PET: Positron emission tomography, CBCT: Cone beam CT, MR: Magnetic resonance

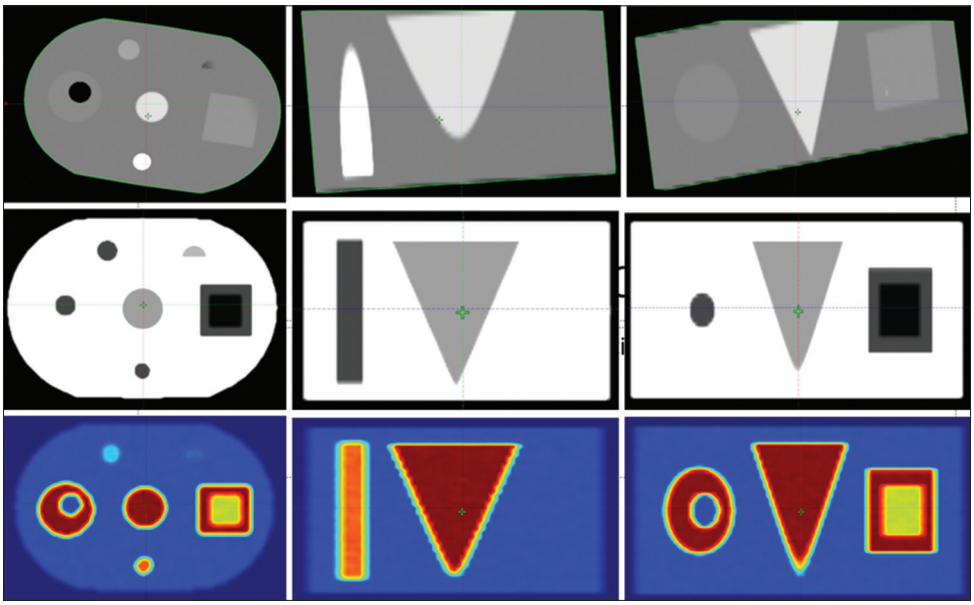

Figure 1: Task group 132 geometric phantoms, axial, sagittal and coronal views. On the top row is the rotated computed tomography phantom, middle row is T2 magnetic resonance and bottom row is positron emission tomography (shown in SmartAdapt). All images have voxel size 0.7 × 0.7 × 3 mm^3

Table 2 shows the updated shifts for rotation followed by translation were taken from Latifi *et al*.[7] Moreover, the signs of the applied translation and rotations varied between SmartAdapt and Velocity as both the systems use a slightly different DICOM coordinate system, as shown in Figure 2.

Deformable image registration

The basic anatomical phantom dataset-1 (CT) and the basic deformable dataset (basic anatomical phantom dataset -1 with added Gaussian noise variation and known global shifts), as shown in Figure 3 were deformed and registered. Contours of bladder, rectum, prostate, and femoral head (both left and right) were done on both datasets by a qualified medical physicist. The images consist of markers placed inside (one marker in each contour) the prostate, rectum, and bladder regions. Evaluation was done by contour comparison using different parameters such as target registration error (TRE),

Table 3: Deformable registration

Stationary dataset	Moving dataset	Evaluation technique
BAD-1 (CT)	BDD-1 (CT)	Virtual fiducials-TRE; contour comparison
Virtual H and N phantoms (SEOT)	Virtual H and N phantoms (SOT)	Contour comparison
Clinical H and N cases (treatment planning CT)	Clinical H and N cases (CBCT)	Contour comparison
Clinical pelvic cases (treatment planning CT)	Clinical pelvic cases (CBCT)	Contour comparison

Contour comparison: DSC, MDA, Jacobian metrics and consistency. BDD: Basic deformation dataset, BAD: Basic anatomical dataset, CT: Computed tomography, CBCT: Cone beam CT, SEOT: Simulated end of treatment, SOT: Start of treatment, TRE: Target registration error, DSC: Dice similarity coefficient, MDA: Mean distance to agreement

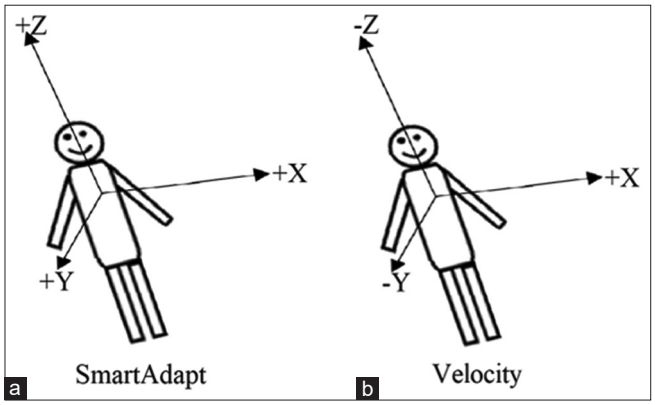

Figure 2: Coordinate system of the Head first supine patient. (a) SmartAdapt (b)Velocity

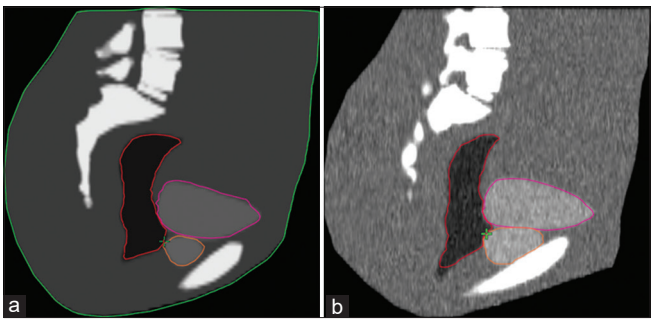

Figure 3: AAPM TG-132 digital basic anatomical phantom dataset. (a) The basic anatomical dataset - 1 (CT), (b) The basic deformable dataset (basic anatomical phantom dataset - 1 with added Gaussian noise variation)

dice similarity coefficient (DSC) and mean distance to agreement (MDA) [Table 3], which are described further below. The consistency of registration was also evaluated.

Virtual head and neck phantoms

For validation of image registration software's, a set of five virtual phantom image sets were downloaded from DIR evaluation project. These phantoms were derived from a pair of CT volumetric image sets of H and N patients acquired prior to the start of treatment (SOT) and near the end of treatment (EOT). Multiple algorithms were used to deform the SOT images to resemble the EOT images. Using the combined DVF, a simulated EOT image (SEOT) was created.

The qualified physicist manually contoured critical organs (such as the parotid L, parotid R, mandible, spinal cord, eye L, and eye R) on both SOT and SEOT image datasets, and the radiation oncologist verified their accuracy. Contours were propagated from SOT to SEOT and vice versa after DIR. These propagated contours were compared with contours that were manually drawn using DSC, Jacobian determinant and MDA parameters.

Clinical dataset

Two clinical sites H and N (n = 10) and gynecology (n = 5), were chosen. For each case, planning CT and CBCT images (acquired on the last day of the treatment) were taken for the study. Organs that were clinically relevant were manually delineated by a qualified medical physicist and then verified by an expert radiation oncologist, which was considered the gold standard. For H and N patients, parotids, mandible, and spinal cord were considered, while for the gynaecological, bladder, rectum, femoral heads, and sigmoid were considered. DIR was performed on these images based on which automatic contour propagation was carried out. These automatic contours generated by DIR were then compared with the contours drawn by a Radiation Oncologist for further evaluation.

Evaluation of deformable image registration

For SmartAdapt®, quantitative evaluation was done using open-source software (3D slicer v5.0.2, https://www.slicer.org/),[10] whereas the required parameters were extractable from Velocity™ itself. The following parameters were evaluated: DSC, MDA, TRE and Jacobian determinant. These indices help in evaluating the accuracy of the segmentation of the contours and the spatial overlap index.

Dice similarity coefficient

This metric computes the number of pixels that overlap between the two volumes and normalizes it by the half the sum of the number of nonzero pixels in the two volumes. The result is a value between 0 (no overlap) and 1 (perfect overlap):

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

Where A is the fixed image and B is the segmentation mapped from the deformably registered image. Even though volume overlap is a good indicator of mismatch, but still, it is a poor indicator of shape because does not have the information of the distance. Hence, therefore, DSC alone is not an efficient technique to evaluate the overall accuracy, other metrics are also required. DSC value >0.75 is acceptable, according to Loi *et al*.[12]

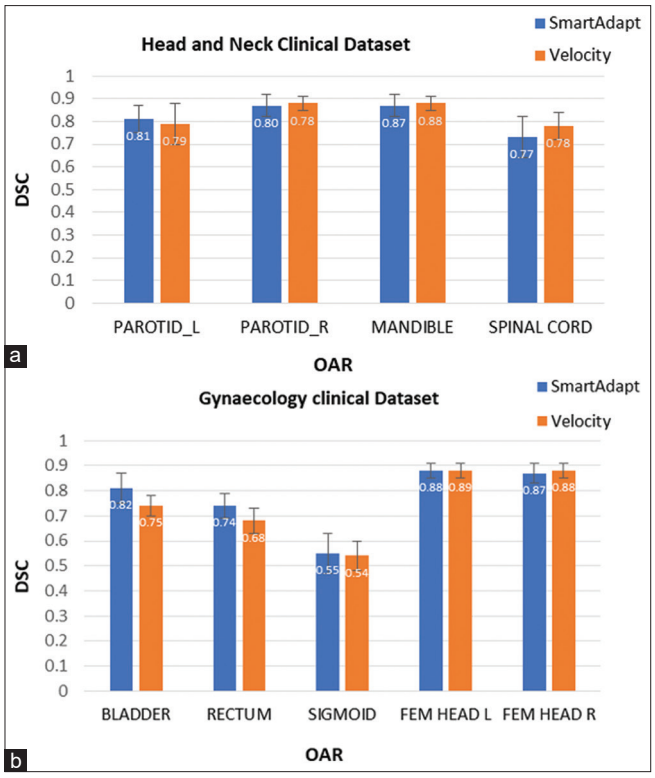

Figure 4: The above figure shows the differences in the dice similarity coefficient of clinical datasets between SmartAdapt® and Velocity™. (a) Head and neck (n = 10); (b) Gynaecology (n = 5). DSC: Dice similarity coefficient, OAR: Organs at risk

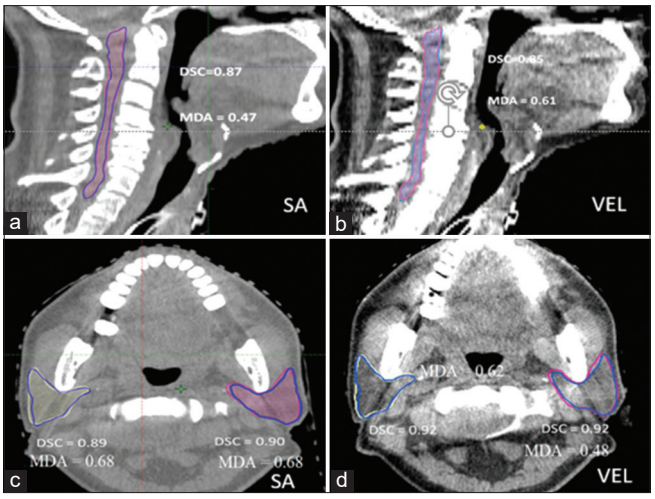

Figure 5: (a-d) Deformable image registration between computed tomography and cone-beam computed tomography for clinical data (head and neck) in both applications (SmartAdapt and velocity). Dice similarity coefficient and mean distance to agreement values are quoted. MDA: Mean distance to agreement, DSC: Dice similarity coefficient, SA: SmartAdapt, VEL: Velocity

Mean distance to agreement

When two image datasets A and B, each having structure set named struct A and struct B are registered, the structure set of image A is mapped onto the structure set of image B, resulting in a deformed structure set named struct A' . The distance between each point on contour B and contour A' would then be calculated to determine the MDA, which would be the average of all distances.

Target registration error

The TRE is a quantitative measure used to evaluate the accuracy of alignment of images. After registration, it is calculated by measuring the distance or error between the transformed positions of anatomical landmarks in one image and their corresponding positions in the other image. In an ideal registration scenario, where the transformation is perfect, the TRE would be zero.

Jacobian determinant

The mathematical function of Jacobian determinant can be calculated from the registration matrix following a deformable registration between two images. It identifies the local volume change as a result of the registration.

J >1 indicates volume expansion, 0 < J < 1 indicates volume reduction, J = 1 indicates no change in volume, J ≤ 0 means nonphysical motion (e.g., regions of the image folded onto itself). A negative value indicates an error in the registration and these areas require careful evaluation for their influence on the results and further applications of the registration.

Table 4: In the TG-132 report, a tolerance table is given for assessing the quantitative metrics of the image registration

Technique	Tolerance
DSC[a]	~ (0.8–0.9)
MDA	Maximum voxel dimension ~ (2–3 mm)
Jacobian	No negative values, values deviating from 1 as expected from clinical scenario 0–1 for structures expected to reduce in volume, greater than 1 for structures expected to expand in volume
Consistency	Maximum voxel dimension ~ (2–3 mm)
TRE	Maximum voxel dimension ~ (2–3 mm)

[a]DSC ≥0.75 can be acceptable as quoted by Loi *et al*.[12] TRE: Target registration error, DSC: Dice similarity coefficient, MDA: Mean distance to agreement

Consistency

This test ensure that the registration is inverse consistence, i.e., registering A to B and B to A. The registration should be consistent in magnitude but in the opposite direction. MDA was used for the comparison between the direct and reverse registrations.

Table 4 shows the summary of technique, the evaluation metric, and their respective tolerances.

Results

Commissioning results

Tables 5 and 6 show the results of rigid registrations (translational only, translational and rotational shifts) on the TG-132 digital phantoms using both Velocity™ and SmartAdapt®.

Table 5: Rigid registration tests - translations only

Stationary data set	Moving data set	Expected '*T*' (mm)	Velocity™			SmartAdapt®		
			Measured '*T*' (mm)	Difference/VD	Result	Measured *T* (mm)	Difference/VD	Result
BPD-2 (CT)	BPD-1 (CT)	10	9.67	0.5	Pass	10	0	Pass
		−5	4.69	0.4	Pass	−5.1	0.1	Pass
		15	−15.05	0.0	Pass	15	0	Pass
	BPD-1 (PET)	10	10.75	1.1	Fail	10.9	1.3	Fail
		−5	5.78	1.1	Fail	−6.1	1.6	Fail
		15	−14.93	0.0	Pass	15	0	Pass
	BPD-1 (MR1)	10	9.97	0.0	Pass	10.3	0.4	Pass
		−5	5.03	0.0	Pass	−5.3	0.4	Pass
		15	−15.04	0.0	Pass	15.2	0.1	Pass
	BPD-1 (MR2)	10	9.9	0.1	Pass	10.3	0.4	Pass
		−5	4.97	0.0	Pass	−5.3	0.4	Pass
		15	−15.1	0.0	Pass	15.4	0.1	Pass
	BPD-1 (CBCT)	10	10.01	0.0	Pass	9.8	0.3	Pass
		−5	5	0.0	Pass	−5.1	0.1	Pass
		30	−31.43	0.5	Pass	31.4	0.5	Pass
BAD-1 (CT)	BAD-2 (CT)	−3	−3.01	0.0	Pass	−3	0	Pass
		5	−5.46	0.5	Pass	5	0	Pass
		−12	11.996	0.0	Pass	−11.9	0	Pass
BAD-1 (CT)	BAD-3 (PET)	−3	−2.05	1.0	Fail	−1.6	1.5	Fail
		5	−4.05	1.0	Fail	3.5	1.6	Fail
		−12	11.84	0.1	Pass	−11.3	0.2	Pass
BAD-1 (CT)	BAD-4 (MR T1)	−3	−3.01	0.0	Pass	−2.6	0.2	Pass
		5	−5.45	0.2	Pass	5.4	0.2	Pass
		−12	12	0.0	Pass	−12.1	0.0	Pass
BAD-1 (CT)	BAD-5 (MR T2)	−3	−3.46	0.3	Pass	−2.5	0.3	Pass
		5	−5.46	0.3	Pass	5.5	0.3	Pass
		−12	12.01	0	Pass	−12	0.2	Pass

VD: Voxel dimension, BAD: Basic anatomical dataset, CT: Computed tomography, PET: Positron emission tomography, CBCT: Cone beam CT, MR: Magnetic resonance, BPD: Basic phantom dataset

Translation only

For rigid translational registration, for both Velocity and SmartAdapt, except the CT-PET registration, all others were well within the recommended tolerance of 0.5*voxel dimension (vd). In velocity, CT-PET registration of both geometric and anatomical phantom, showed deviation from the tolerance value in x and y directions, with a maximum deviation of 1.1*vd, whereas in SmartAdapt the maximum deviation was 1.6*vd.

Translation and rotation

As shown in Table 6, in Velocity, for CT-CT and CT-PET registrations, all directions fail with a maximum and minimum of 3.4*vd and 0.6*vd, respectively. For SmartAdapt, the interval is between 0.6*vd to 3.6*vd. Only x-direction passed for CT-MR image registrations, and CT-CBCT in both Velocity and SmartAdapt.

With SmartAdapt, the rotational axis angular misalignment varies from 0.1° (CT to MR registration) to 0.2° (CT to CBCT registration). In velocity, it is between 0.3° (CT to CT registration) and 1.5° (CT to PET registration).

Deformable registration

As shown in Table 7, with regard to TRE, for SmartAdapt and velocity, the maximum inaccuracy was 1.7 and 2.2 mm, respectively. SmartAdapt has good agreement in MDA values ranging from 0.66 to 1.71 mm. DSC values ranged from 0.88 to 0.96. For velocity, MDA values range from 1.10 to 2 mm and DSC values are between 0.87 and 0.93. Consistency test using MDA of the deformable registrations showed a maximum difference of 0.3 mm for rectum in SmartAdapt and 0.63 mm for bladder in velocity.

Deformable image registration results

Virtual head and neck phantoms

Table 8 shows the DIR results (mean ± standard deviation) of virtual H and N phantoms that were compared in SmartAdapt and Velocity using the DSC, MDA, Jacobian metrics, and consistency.

In SmartAdapt and velocity, parotid_L's average DSC score was 0.84 ± 0.04 and 0.80 ± 0.03, respectively, whereas parotid_R's average DSC score was 0.81 ± 0.06 and 0.79 ± 0.08, respectively. Spinal cord and mandible was 0.82 ± 0.03 and 0.84 ± 0.05 in both systems. SmartAdapt and

Table 6: Rigid registration tests - translational and rotational

Stationary data set	Moving data set	Velocity™							SmartAdapt®				
		Expected rotation (°)	Measured rotation (°)	Difference	Expected *T* (mm)	Measured *T* (mm)	Difference/VD	Result	Measured rotation (°)	Difference	Measured *T* (mm)	Difference/VD	Result
BPD-3 (CT)	BPD-1 (CT)	−5	−3.6	1.4	5.07	−4.68	0.6	Fail	−5	0	5	0.1	Pass
		8	8.3	0.3	−17.29	−14.91	3.4	Fail	−8	0	−14.8	3.6	Fail
		10	−9.4	0.6	18.06	19.94	0.6	Fail	10	0	20.4	0.8	Fail
BPD-3 (CT)	BPD-1 (PET)	−5	−3.5	1.5	5.07	−5.57	0.7	Fail	−5.1	0.1	5.9	1.2	Fail
		8	8.7	0.7	−17.29	−15.73	2.2	Fail	−8.2	0.2	−18.4	1.6	Fail
		10	−9.3	0.7	18.06	20	0.6	Fail	9.9	0.1	18.0	0	Pass
BPD-3 (CT)	BPD-1 (MR1)	−5	−3.6	1.4	5.07	−5.23	0.2	Pass	−5	0	5.3	0.3	Pass
		8	8.7	0.7	−17.29	−15.18	2.6	Fail	−8.1	0.1	−15.2	3.0	Fail
		10	−9.4	0.6	18.06	20.14	0.7	Fail	10	0	20.8	0.7	Fail
BPD-3 (CT)	BPD-1 (MR2)	−5	−3.6	1.4	5.07	−4.77	0.4	Pass	−5	0	5.3	0.3	Pass
		8	8.7	0.7	−17.29	−14.96	3.3	Fail	−8.1	0.1	−15.2	3.0	Fail
		10	−9.4	0.6	18.06	20.15	0.7	Fail	9.9	0.1	20.1	0.7	Fail
BPD-3 (CT)	BPD-1 (CBCT)	−5	−3.6	1.4	5.07	−4.75	0.5	Pass	−5	0	5.1	0.0	Pass
		8	8.3	0.3	−17.29	−14.92	3.4	Fail	−8.2	0.2	−15	3.3	Fail
		10	−9.4	0.6	18.06	20.02	0.7	Fail	10	0	19.8	0.6	Fail

VD: Voxel dimension, CT: Computed tomography, CBCT: Cone-beam CT, MR: Magnetic resonance, BPD: Basic phantom dataset

velocity values for eye R were 0.87 ± 0.02 and 0.89 ± 0.03, respectively, whereas eye L was 0.88 ± 0.02 and 0.90 ± 0.02, respectively.

The average MDA value for all structures were < 2 mm in both the systems. The Jacobian determinant was also evaluated. For Parotid_L and Parotid_R, the voxel size has been increased in SmartAdapt, while in velocity, it reduced. The spinal cord in both systems has the same voxel size. However, in both systems, the mandible voxel size has shrunk. Whereas in SmartAdapt, the size of the voxels in the regions of the Eye_L and Eye_R has decreased, while the velocity has not changed, it was almost equivalent to 1 ± 0.01.

For virtual H and N phantom, to evaluate consistency, MDA parameters were assessed after SEOT to SOT and SOT to SEOT registration. The mean value for five patients is shown in Table 8. The maximum registrational error recorded for MDA was 0.97 mm for spinal cord in SmartAdapt and 1.02 mm for Eye_R in Velocity, respectively.

Clinical data set

Site: Head and neck

In SmartAdapt, as shown in Figure 4a the DSC score for the spinal cord was 0.77 ± 0.06, the parotids, and the mandible were all >0.8, while in Velocity only the mandible had DSC >0.8, the remaining other structures were just above 0.75, as shown in Table 9. MDA for both the systems was <2 mm. In SmartAdapt, the Jacobian determinant was >1, except for the mandible (J ~1), whereas in velocity, it was <1, except for the spinal cord.

Site: Gynaecology

The DIR results of gynaecological clinical dataset are shown in Table 10. For both systems, as shown Figure 4b DSC of the femoral heads scored >0.8 and the rectum and sigmoid <0.75. Bladder scored 0.82 ± 0.06 and 0.75 ± 0.04 in SmartAdapt and velocity, respectively. MDA showed high values in both Velocity and SmartAdapt while MDA for the femoral heads were <3 mm. Jacobian determinant is roughly equal to 1 for both systems, with the exception of the bladder 1.15 ± 0.12 and rectum 0.82 ± 0.25 in SmartAdapt. In testing for consistency, the maximum difference in magnitude of the registrational error for MDA is 1.43 mm for femoral head L and 1.33 mm for femoral head R in SmartAdapt and Velocity, respectively.

Discussion

A number of conceptual approaches can be used to validate commercial DIR algorithms, including: (a) comparing the deformation vector field (DVF)[13,14] with the ground truth value; (b) examining the spread of anatomical landmarks (points) to determine TRE; (c) examining the overlap of the deformably propagated contours with the known segmentation results and (d) physical phantom evaluations.

The comparison of the deformation vector field (DVF) with the ground truth value is the most effective way among the four

Table 7: Deformable image registration comparison between SmartAdapt® versus Velocity™

Structure	DSC		MDA (mm)		TRE (mm)		Consistency (MDA difference in mm)	
	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity
Bladder	0.96	0.89	0.66	2.00	1.1	1.9	0.01	0.63
Rectum	0.91	0.90	1.27	1.42	1.5	2.2	0.30	0.3
Prostate	0.90	0.87	0.98	1.10	1.7	1.6	0.01	0.06
Femoral head neck_Left	0.88	0.93	1.71	1.19	-	-	0.02	0.19
Femoral head neck_Right	0.88	0.92	1.67	1.74	-	-	0.00	0.59

DSC: Dice similarity coefficient, MDA: Mean distance to agreement, TRE: Target registration error

Table 8: Virtual head and neck phantoms (n=5)

OAR	DSC, mean±SD		MDA (mm), mean±SD		Jacobian, mean±SD		Consistency (MDA difference in mm), mean±SD	
	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity
Parotid left	0.84±0.04	0.80±0.03	1.10±0.51	1.51±0.28	1.23±0.18	0.85±0.10	0.25±0.05	0.35±0.19
Parotid right	0.81±0.06	0.79±0.08	1.35±0.51	1.56±0.59	1.14±0.15	0.85±0.07	0.32±0.20	0.32±0.34
Mandible	0.84±0.05	0.84±0.05	0.95±0.20	0.90±0.14	0.91±0.23	0.92±0.04	0.31±0.35	0.20±0.17
Spinal cord	0.82±0.05	0.82±0.03	0.61±0.18	0.87±0.20	1.00±0.20	1.00±0.04	0.24±0.41	0.11±0.11
Eye left	0.88±0.02	0.90±0.02	0.82±0.28	0.61±0.28	0.91±0.06	0.99±0.01	0.36±0.26	0.18±0.15
Eye right	0.87±0.02	0.89±0.03	0.95±0.35	0.76±0.33	0.90±0.06	1.00±0.01	0.08±0.30	0.52±0.45

DSC: Dice similarity coefficient, MDA: Mean distance to agreement, SD: Standard deviation, OAR: Organs at risk

techniques, however creating the ground truth value is difficult task and necessitates the use of different software tools. It is a useful method for confirming the DIR since it determines TRE[15] by using a lot of anatomical markers (points). As there are not enough data points available for our analysis of TRE, putting the data points together requires a lot of care.

The next easy technique is called "contour comparison," where "fiducial points" are replaced by "organ(s) surface contours" and "point-to-point TRE values" are replaced by "contour overlap" or "closeness" metrics, which are less precise. On the plus side, this kind of test may be created by almost any facility as all that is needed is a pair of clinical datasets that have been correctly segmented. Physical phantoms are not a practical option in the majority of radiation facilities, due to limited application in assessing DIR algorithms as they cannot completely evaluate effect of anatomical differences on the algorithm's effectiveness. In addition, the phantoms' deformations may not be always anatomically realistic. Hence, we have not used any physical phantoms in this study's DIR evaluations.

Since velocity is a new system in our center and SmartAdapt is a constituent of Eclipse software, we chose to commission both systems and compare them to determine which one is operating as intended. In our investigation, commissioning and validation of the DIR algorithm were performed on both systems utilizing TG-132 digital phantom datasets, and both of them nearly produced identical findings. In velocity, all modalities of TG-132 digital phantoms from AAPM library could be imported directly, but in Eclipse, these image collections had to be made anonymous before being imported.

When commissioning and validating DIR software, the ground truth values – that is, known translational and rotational values – that are downloaded with these image dataset collections are very helpful, as shown in Figure 2, the variations in the DICOM[16] coordinates between the SmartAdapt® Velocity™ systems account for the difference in the signs of the measured values. Thus, in order to prevent confusion, the measured and expected rotational values [Tables 5 and 6] as well as the measured and expected T values have not been taken into consideration. In addition, TG-132 phantoms include deformable anatomical phantoms with fiducial implants in the prostate, rectum, and bladder. TRE measurements have been made by registering the two picture sets. For femoral heads, TRE was not possible because of the lack of fiducial implants.

In our investigation, we observed that the difference between actual and measured rotational measurements is more pronounced in Velocity compared to SmartAdapt. This is because both systems use a process called rigid registration to optimize[15] the images. Wherein the optimization of three-dimensional (3D)-3D images is a complex process involving critical modules such as similarity measure, optimizer, preprocessor, and interpolator. It's noteworthy that these modules are proprietary software components. Both Velocity and SmartAdapt systems utilize these essential parameters for optimizing 3D images, although the specific details regarding the parameters employed for this optimization are not very clear.

For comparing the DIR of SmartAdapt® and Velocity™, we registered CT and CBCT images of actual patient to create a clinical dataset, as shown in Figures 5 and 6. Last week's

Table 9: Head and neck clinical data set (*n*=10)

OAR	DSC (mean)		MDA (mm) (mean)		Jacobian (mean)	
	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity
Parotid left	0.81±0.06	0.79±0.09	1.08±0.32	1.27±0.44	1.27±0.15	0.88±0.09
Parotid right	0.80±0.07	0.78±0.09	1.17±0.49	1.32±0.59	1.28±0.17	0.91±0.07
Mandible	0.87±0.05	0.88±0.03	0.81±0.36	0.70±0.20	1.03±0.10	0.99±0.04
Spinal cord	0.77±0.06	0.78±0.06	0.94±0.31	1.02±0.34	1.36±0.13	1.01±0.02

DSC: Dice similarity coefficient, MDA: Mean distance to agreement, OAR: Organs at risk

Table 10: Gynecology clinical data set (*n*=5)

OAR	DSC (mean)		MDA (mm) (mean)		Jacobian (mean)		Consistency (MDA difference in mm) (mean)	
	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity	SmartAd	Velocity
Bladder	0.82±0.06	0.75±0.04	3.44±1.38	5.21±1.09	1.15±0.12	0.98±0.08	0.48±0.49	0.58±0.21
Rectum	0.74±0.05	0.68±0.04	2.90±1.08	3.25±0.93	0.82±0.25	1.06±0.13	0.12±0.08	0.56±0.38
Sigmoid	0.55±0.08	0.54±0.06	4.24±0.92	4.52±0.89	1.08±0.13	0.98±0.03	0.38±0.22	0.52±0.40
Femoral head left	0.88±0.03	0.89±0.03	2.11±0.97	1.97±0.83	1.01±0.14	1.05±0.02	0.60±0.55	0.53±0.32
Femoral hear right	0.87±0.04	0.88±0.03	2.19±1.19	1.94±0.92	0.98±0.05	1.04±0.03	0.65±0.53	0.64±0.62

DSC: Dice similarity coefficient, MDA: Mean distance to agreement, OAR: Organs at risk

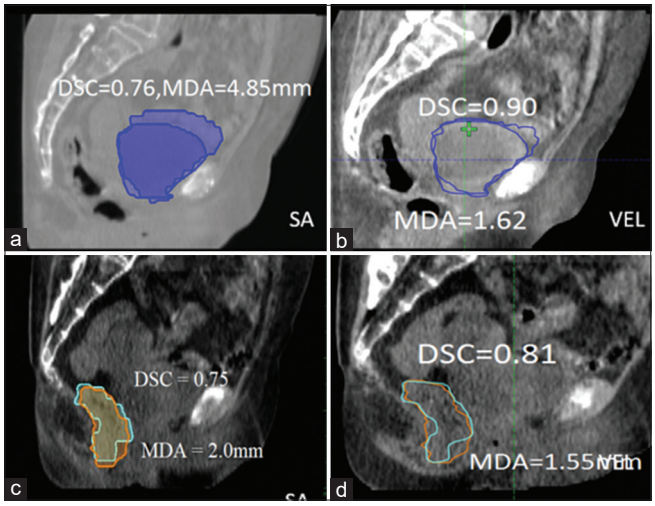

Figure 6: (a-d) Deformable image registration between computed tomography and cone-beam computed tomography for clinical data (Gyne) in both applications (SmartAdapt and velocity). Dice similarity coefficient and mean distance to agreement values are quoted. MDA: Mean distance to agreement, DSC: Dice similarity coefficient, SA: SmartAdapt, VEL: Velocity

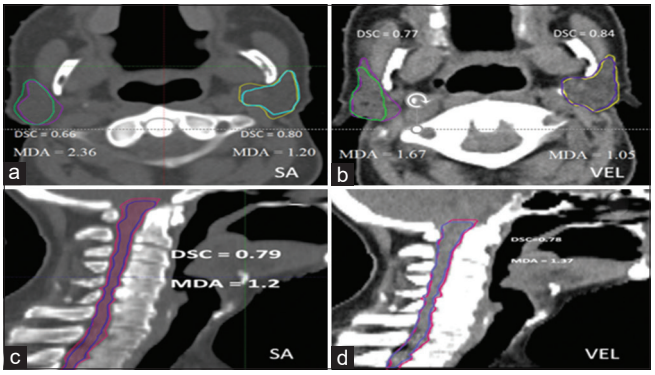

Figure 7: (a-d) Deformable image registration between virtual phantom dataset (start of treatment vs. simulated end of treatment) in both applications (SmartAdapt and velocity). Dice similarity coefficient and mean distance to agreement values are quoted. MDA: Mean distance to agreement, DSC: Dice similarity coefficient, SA: SmartAdapt, VEL: Velocity

CBCT taken so that maximum deformation can be tested. In addition to using a clinical dataset, we incorporated digitally altered images. The alteration aimed to simulate the maximum deformation seen in the original patient from the start of the treatment to the end of the treatment. We compared contours for both the digitally altered virtual phantom (SOT vs. SEOT), as shown in Figure 7 and the clinical dataset (CT vs. CBCT). However, we did not perform TRE analysis due to the absence of markers in the images.

Finally, according to the AAPM TG-132 study, the validation procedure can be roughly categorized into three ways: Utilizing real phantoms with known deformations, landmark tracking on digitally produced phantoms, and by contour comparison using real clinical images. Each approach has its own restrictions and range of applicability and the DIR validation procedure is still being refined.

Although rigid registration of translation with rotation fails the stringent criteria laid down by TG-132, it can be seen that even the maximum deviations are a maximum of 2–3 mm for all directions for the digital phantoms in CT-MR cases and this is similar to the results that various authors like Latifi *et al*.,[7] Jamema *et al*.[8] and have reported for different systems/algorithms of DIR. Hence, we have also developed a departmental clinical protocol (as recommended by AAPM TG-132) to use different image modalities for use in appropriate scenarios with caution.

Based on results, a departmental protocol has been developed where image registration by either system is used for:

1. Localization with optimum margins, target, and organs at risk (OAR) definition for brain site with CT-MR rigid registration
2. Target and OAR definition for H and N site adaptive cases with CT-CT rigid registration, general location of nodes/primary target in nasopharynx cases with CT-PET rigid registration
3. General localization of target for thoracic cases using CT-PET rigid registration
4. General localization of target for gynaecologic and prostate cases with rigid registration of CT-MR images
5. Target delineation for liver stereotactic body radiotherapy cases with rigid registration using soft-tissue landmarks (liver to liver) deformable maybe used for complementary information
6. Dose compositing is to be explored for adaptive plans in future scenarios.

Both systems have shortcomings, and SmartAdapt does not provide the appropriate evaluation criteria for DIR (example: DSC, MDA), so additional third-party software is needed to assess the deformation. On SmartAdapt, Jacobian values were recorded for each voxel that was encircled by a contour on a slice with a 5-slice gap. While the nomenclature for velocity is identical for various registrations despite being given the parameters for direct evaluation, this leads to confusion.

Conclusion

Commissioning tests were successfully carried out according to procedures laid out in AAPM TG-132 on both velocity and SmartAdapt software. Both of them were deemed to be usable for treatment planning, treatment delivery and treatment monitoring, and adaptation contouring related work as discussed.

On comparison of deformation results of the H and N (both clinical and digital phantom data) and gynecology sites, this study suggests no significant difference between SmartAdapt® and Velocity™. There were only very slight differences between SmartAdapt and Velocity, i.e. lower values given by velocity for soft tissues (such as bladder, rectum, prostate, and parotids) and nearly identical values for both systems for rigid structures (such as the mandible, spinal cord, and femoral heads).

Ethical consideration

For this paper, an exemption from formal review and waiver of consent has been granted by the Institutional ethical committee of Homi Bhabha Cancer Hospital and Research Center, Visakhapatnam, India, affirming adherence to ethical standards.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. Phys Med Biol 1997;42:123-32.
2. 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.
3. 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.
4. Chen M, Tustison NJ, Jena R, Gee JC. Image registration: Fundamentals and recent advances based on deep learning. In: Colliot O, editors. Machine Learning for Brain Disorders, Neuromethods. Vol. 197. New York, NY: Humana; 2023.
5. Kim H, Park SB, Monroe JI, Traughber BJ, Zheng Y, Lo SS, *et al.* Quantitative analysis tools and digital phantoms for deformable image registration quality assurance. Technol Cancer Res Treat 2015;14:428-39.
6. Rong Y, Rosu-Bubulac M, Benedict SH, Cui Y, Ruo R, Connell T, *et al.* Rigid and deformable image registration for radiation therapy: A self-study evaluation guide for NRG oncology clinical trial participation. Pract Radiat Oncol 2021;11:282-98.
7. Latifi K, Caudell J, Zhang G, Hunt D, Moros EG, Feygelman V. Practical quantification of image registration accuracy following the AAPM TG-132 report framework. J Appl Clin Med Phys 2018;19:125-33.
8. Jamema SV, Phurailatpam R, Paul SN, Joshi K, Deshpande DD. Commissioning and validation of commercial deformable image registration software for adaptive contouring. Phys Med 2018;47:1-8.
9. 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.
10. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, *et al.* 3D slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012;30:1323-41.
11. Deformable Image Registration Evaluation Project. Available from: https://www.sites.google.com/site/dirphantoms/home. [Last accessed on 2023 Apr 23].
12. 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.
13. 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.
14. Dang J, Luo O, Gu X, Wang J. Deformation vector fields (DVF)-driven image reconstruction for 4D-CBCT. J Xray Sci Technol 2015;23:11-23.
15. Al-Jaberi F, Fachet M, Matthias M, Skalej M, Hoeschen C. Optimization techniques for semi-automated 3D rigid registration in multimodal image-guided deep brain stimulation. Curr Dir Biomed Eng 2023;9:355-8.
16. Bidgood WD Jr., Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. J Am Med Inform Assoc 1997;4:199-212.