Advances in Radiation Oncology (2024) 9, 101580

advances in radiation oncology

www.advancesradonc.org

Scientific Article

Personalized Deep Learning Model for Clinical Target Volume on Daily Cone Beam Computed Tomography in Breast Cancer Patients

Joonil Hwang, MS,[a,b] Jaehee Chun, PhD,[c,d] Seungryong Cho, PhD,[a,b] Joo-Ho Kim, MS,[e] Min-Seok Cho, MS,[e] Seo Hee Choi, MD,[e,f,*] and Jin Sung Kim, PhD[c,d,f,*]

[a]Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea; [b]Medical Image and Radiotherapy Lab (MIRLAB), Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea; [c]OncoSoft, Seoul, Republic of Korea; [d]Medical Physics and Biomedical Engineering Lab (MPBEL), Yonsei University College of Medicine, Seoul, Republic of Korea; [e]Department of Radiation Oncology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Gyeonggi-do, Republic of Korea; and [f]Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Republic of Korea

Received 26 October 2023; accepted 17 July 2024

Purpose: Herein, we developed a deep learning algorithm to improve the segmentation of the clinical target volume (CTV) on daily cone beam computed tomography (CBCT) scans in breast cancer radiation therapy. By leveraging the Intentional Deep Overfit Learning (IDOL) framework, we aimed to enhance personalized image-guided radiation therapy based on patient-specific learning.

Methods and Materials: We used 240 CBCT scans from 100 breast cancer patients and employed a 2-stage training approach. The first stage involved training a novel general deep learning model (Swin UNETR, UNET, and SegResNET) on 90 patients. The second stage used intentional overfitting on the remaining 10 patients for patient-specific CBCT outputs. Quantitative evaluation was conducted using the Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), mean surface distance (MSD), and independent samples *t* test with expert contours on CBCT scans from the first to 15th fractions.

Results: IDOL integration significantly improved CTV segmentation, particularly with the Swin UNETR model (P values $< .05$). Using patient-specific data, IDOL enhanced the DSC, HD, and MSD metrics. The average DSC for the 15th fraction improved from 0.9611 to 0.9819, the average HD decreased from 4.0118 mm to 1.3935 mm, and the average MSD decreased from 0.8723 to 0.4603. Incorporating CBCT scans from the initial treatments and first to third fractions further improved results, with an average DSC of 0.9850, an average HD of 1.2707 mm, and an average MSD of 0.4076 for the 15th fraction, closely aligning with physician-drawn contours.

Conclusion: Compared with a general model, our patient-specific deep learning-based training algorithm significantly improved CTV segmentation accuracy of CBCT scans in patients with breast cancer. This approach, coupled with continuous deep learning training using daily CBCT scans, demonstrated enhanced CTV delineation accuracy and efficiency. Future studies should explore the adaptability of the IDOL framework to diverse deep learning models, data sets, and cancer sites.

Sources of support: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2020R1A2C2011959), (NRF-2021R1A2C1010900), and (NRF-2022R1A2C2008623) Also was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI23C0730).

Research data are available on request from the corresponding authors (clickby_s@yuhs.ac and jinsung@yuhs.ac). Data will only be available for academic research and not for other objectives (ie, commercial use). A data use agreement and institutional review board approval will be required before the release of data, as appropriate.

*Corresponding authors: Seo Hee Choi, MD and Jin Sung Kim, PhD; Emails: clickby_s@yuhs.ac jinsung@yuhs.ac

https://doi.org/10.1016/j.adro.2024.101580

Introduction

Radiation therapy (RT), including that for whole breast or chest wall and regional nodes, plays a crucial role in breast cancer treatment by reducing the risk of local recurrence after surgery and enhancing survival rates.[1-6] Technical advancements in RT for breast cancer, including the implementation of intensity-modulated RT, simultaneously integrated boost, and partial breast irradiation, have enabled a more precise treatment for the affected area.[7-9] Contemporary radiation techniques used in breast cancer treatment require a precise definition of target areas and organs at risk to devise personalized treatment that efficiently covers the targeted regions while minimizing radiation exposure to organs at risk to mitigate the risk of toxicities.[10,11] These techniques necessitate special attention for RT planning, accurate positioning, and patient monitoring during treatment delivery.

Recently, the importance of adaptive planning using image-guided RT (IGRT) has been demonstrated, and cone beam computed tomography (CBCT) has emerged as a valuable tool for adaptive RT (ART) by enabling accurate and timely guidance.[12,13] Well-implemented ART can minimize treatment-related side effects and avert inadvertent dosage variations in specific regions. The crux of ART process involves aligning and comparing the target volume on the initial planning computed tomography (CT) scan with the corresponding area on the daily CBCT scan. When notable discrepancies between the 2 scans can substantially impact dose distribution, a new ART plan is instituted. However, despite its importance, comprehensive daily comparison between CBCT and planning CT scans can be time-intensive and impractical. Studies have proposed CBCT-guided IGRT for head and neck or prostate treatment.[14,15] Additionally, deep learning (DL)-based methods have been proposed for automatic segmentation of breast masses on ultrasound imaging, breast cancer diagnosis via magnetic resonance imaging (MRI), and detection in mammography.[16-18] DL not only aids in the accurate identification of breast abnormalities but also plays a crucial role in supporting clinical decision-making.[19]

In breast cancer treatment, the area of the patient's body included in the clinical target volume (CTV) can change significantly throughout therapy. To evaluate how these daily changes impact the actual delivered radiation dose and the need for adaptive planning, physicians should manually delineate the CTV on each daily CBCT scan or use deformable registration. However, deformable registration from CT to CBCT is often insufficient and inaccurate, necessitating manual delineation of the CTV on each CBCT scan. This labor-intensive and time-consuming process highlights the convenience of an automatic contouring system. After delineating the CTV on CBCT, it can be compared with the planning CT to determine the dose difference and the necessity for an adaptive plan. Because RT is conducted daily, prompt assessment and implementation of adaptive plans are crucial to ensure accurate radiation delivery and reduce side effects. Automating the CTV contouring process can save time, improve efficiency, and enhance accuracy. However, the application of DL for automatic segmentation of breast CTV on daily CBCT scans remains unexplored.

This study primarily aimed to develop a DL algorithm for accurate CTV segmentation using daily CBCT scans of patients with breast cancer who underwent breast-conserving surgery and adjuvant RT. By leveraging the Intentional Deep Overfit Learning (IDOL) framework, we sought to incorporate prior knowledge, typically available as fractionated ART, to further improve segmentation performance for fraction images.[20] By using the IDOL framework, our study aimed to enable patient-specific learning through individual CBCT data to enhance the accuracy and efficiency of IGRT, which has important implications in CTV segmentation in the context of breast cancer RT and may substantially advance the field.

Material and Method

Patient data

This retrospective study analyzed data from 100 patients with breast cancer who underwent adjuvant RT after breast-conserving surgery between March 2020 and December 2021 at Yongin Severance Hospital were included in this study. The patients underwent whole-breast irradiation using the volumetric modulated arc therapy technique and a simultaneous integrated boost to the tumor bed; RT planning was performed according to the institutional treatment policy.[21] The prescribed doses were 40.05 Gy and 48 Gy in 15 fractions to the whole breast and tumor bed each, respectively.[21-23] All patients were treated on the Elekta Versa HD linear accelerator (Elekta AB) equipped with an ELEKTA XVI CBCT scanner. As part of the IGRT process, daily CBCT images were acquired before each treatment session for all patients using an ELEKTA XVI scanner with a voltage of 100 kVp, voxel size of 1.367 mm × 1.367 mm × 3 mm, and voxel size of 512 × 512 × (range: 122-179).

To train our basic DL model, 1 CBCT scan per patient was obtained from 90 patients, whereas 15 daily CBCT

scans during the entire treatment period were obtained from the remaining 10 patients for validation and personalized model development. Furthermore, for enhanced patient-specific modeling, the CBCT fraction data of the remaining 10 patients were used. This involved leveraging a pretrained model and further training it using all 14 CBCT scans available per patient. In the acquired CBCT scans, the breast CTV on CBCT scans (CTV_{cbct}) was manually delineated using MIM software by a board-certified radiation oncologist with 10 years of experience, in the same manner as on the planning CT. This study was approved by the Institutional Review Board at the Yongin Severance Hospital (approval no. 9-2022-0198), and all procedures were performed in accordance with the relevant guidelines and regulations. The requirement for informed consent was waived because of the retrospective nature of the study and the use of anonymized data.

The median age of the patients was 54 years (range: 21-86 years), and all patients underwent whole-breast RT after breast-conserving surgery for invasive carcinoma or ductal carcinoma in situ. A total of 58 and 42 patients received treatment for left- and right-sided breast cancer, respectively. The baseline clinical characteristics of the patients are summarized in Table 1.

Models

In designing our CTV_{cbct} segmentation approach within the IDOL framework, the selection of Swin

Table 1 Demographic, tumor, and treatment characteristics of patients included in this study

Variables		No. of patients (n = 100)	%
Age (y)		Median 54 (range: 21-86)	
ECOG performance status	0	77	77
	1	23	23
Laterality	Right	42	42
	Left	58	58
Pathology	DCIS	18	18
	IDC	68	68
	ILC	3	3
	Others	11	11
T stage*	Tis	21	21
	T1	67	67
	T2	10	10
	T3	1	1
	Unknown	1	1
N stage*	N0	100	100
M stage*	M0	100	100
Chemotherapy	Performed	28	28
	Not performed	72	72
Neoadjuvant CTx	Performed	4	4
	Not performed	96	96
Adjuvant CTx	Performed	23	23
	Not performed	77	77
Herceptin	Performed	16	16
	Not performed	84	84
Hormone therapy	Not performed	15	15
	Performed (during RT)	80	80
	Performed (after RT)	5	5

Abbreviations: CTx = chemotherapy; DCIS = ductal carcinoma in situ; ECOG = European Cooperative Oncology Group; IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; RT = radiation therapy.
*TNM stage according to the AJCC 8th edition.

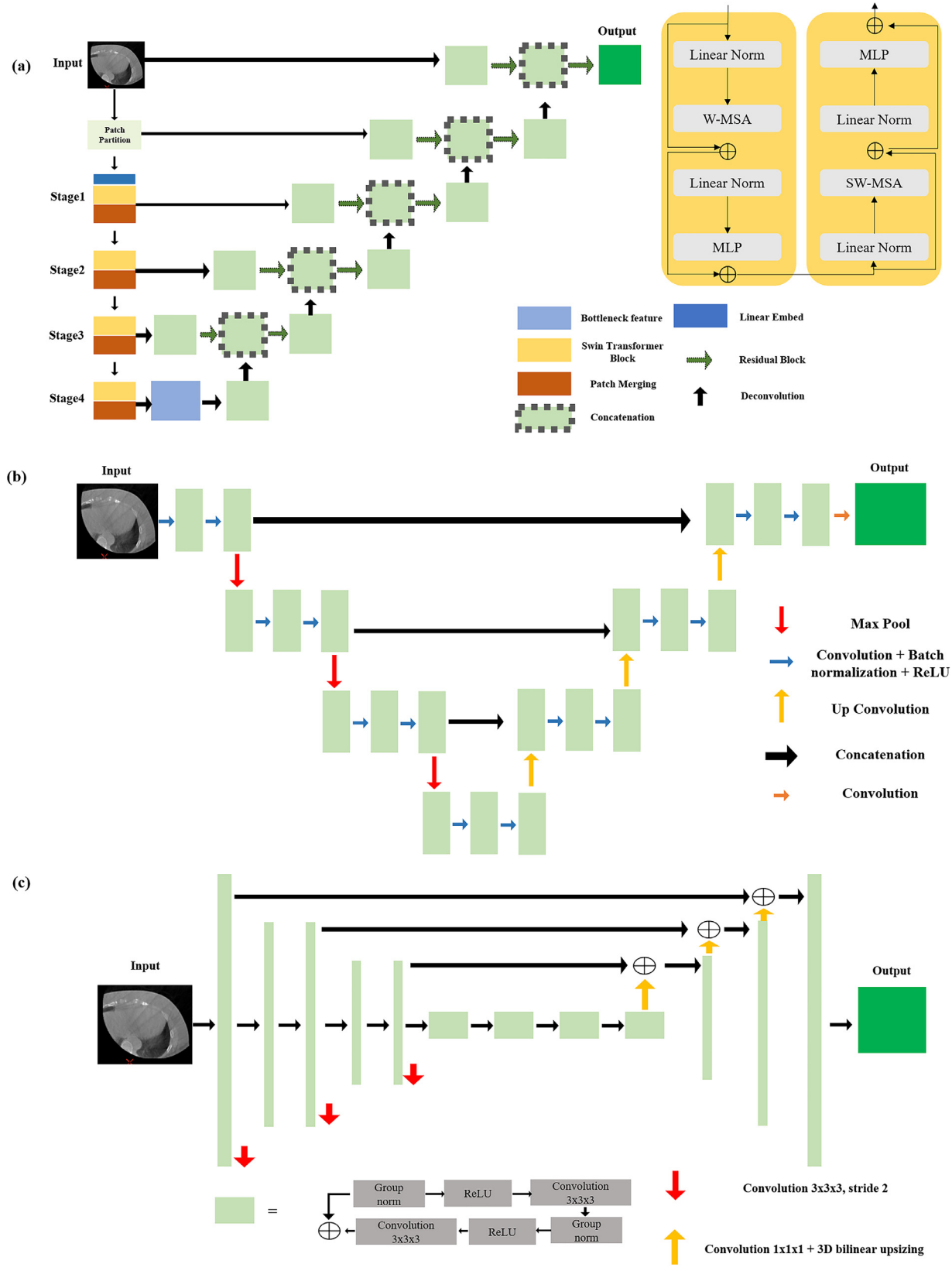

Figure 1 Illustration of the overall 3D network architecture for the IDOL framework. (a) Illustrate Swin UNETR architecture. (b) Illustrate UNET architecture. (c) Illustrate SegResNET architecture.
Abbreviation: IDOL = Intentional Deep Overfit Learning.

UNETR, UNET, and SegResNET was driven by a meticulous evaluation of their individual strengths and their collective synergistic impact on the task at hand.[24-27] The Swin UNETR model, renowned for its successful integration of transformer and convolutional neural network features, demonstrated exceptional performance across diverse medical image analysis tasks. Notably, the Swin UNETR model, proposed in 2022, secured the first position in the "Beyond the Cranial Vault" challenge in 2023. Our implementation (Fig. 1a) leverages the Swin UNETR architecture, featuring an encoder network for robust feature extraction and a decoder network for precise segmentation map generation. This unique blend of transformer and convolutional features proved particularly effective in addressing the intricate challenges associated with CTV_{cbct} segmentation in breast cancer patients.

In addition to Swin UNETR, we incorporated the UNET architecture (Fig. 1b), a well-established model recognized for its proficiency in detection tasks since its proposal in 2019. The inclusion of UNET complements our segmentation strategy, enhancing our ability to capture detailed information crucial for accurate CTV_{cbct} delineation.

Furthermore, the integration of the SegResNET model (Fig. 1c) was motivated by its documented effectiveness in brain tumor segmentation using 3-dimensional (3D) MRI since its introduction in 2018. Although our primary focus is on CTV_{cbct} segmentation in breast cancer patients, we strategically chose SegResNET for its unique architectural characteristics, particularly its capacity to handle complex spatial relationships. We recognize that these attributes can be advantageous in the nuanced task of delineating CTV_{cbct} boundaries.

Our decision to opt for Swin UNETR, UNET, and SegResNET is grounded in their proven performance in medical image analysis, and their integration within the IDOL framework is tailored to address the specific challenges associated with CTV_{cbct} segmentation in the context of breast cancer. This considered selection ensures that our approach leverages the strengths of each 3D model to achieve optimal segmentation outcomes, emphasizing the importance of preserving volumetric information for accurate delineation.

IDOL framework

We employed the innovative IDOL framework to develop a patient-specific DL model through a 2-stage training methodology. Initially, we trained novel DL models using CBCT scans from 90 patients, with a focus on the first fraction, and evaluated performance on the 15th fraction of the remaining 10 patients. The second stage involved patient-specific learning, achieved by intentional overfitting of the pretrained network with CBCT scans from the first to 14th fractions and testing on the remaining fraction. A schematic representation of the IDOL framework, highlighting the optimal model architecture (Fig. 2a), is presented in Fig. 2.

During the first stage, the IDOL framework may exhibit a higher computational loss in the validation set and a generalization error, E_{gen}, comparable with traditional DL methods. Challenges such as saturation in the validation error, E_{valid}, were observed initially, attributed to factors like data set size, generalization limitations, and model architecture constraints. In the second stage, intentional overfitting of the pretrained network occurred, leveraging task- and patient-specific prior information. This deliberate overfitting aimed to create a personalized IDOL model with a customized learned hypersurface, leading to a significant reduction in E_{valid} and an improvement in the IDOL framework error, E_{IDOL} (detailed in Appendix E1).[20]

Although sharing elements with transfer learning, the IDOL framework surpasses conventional practices. In the first stage, similarities to traditional DL methods may emerge, including challenges like E_{gen}. However, intentional overfitting in the second stage distinguishes IDOL from typical transfer learning approaches. Notably, IDOL transitions from generalization to personalization, whereas transfer learning moves from one generalization to another. This deliberate overfitting incorporates patient-specific information, enabling the creation of a highly customized model that excels in patient-specific performance.

The IDOL framework's distinctive advantage lies in its ability to adapt to patient-specific nuances. Through intentional overfitting based on individualized data, the IDOL framework tailors the DL model to the specific patient and task, resulting in enhanced generalization performance and reduced errors. This patient-centric approach positions the IDOL framework as a powerful tool for optimizing DL models in the realm of personalized medical applications.

Training setting

We used several networks and automated segmentation methods to contour the breast CTV and further enhanced it using various preprocessing techniques to improve the segmentation accuracy. These techniques included pixel value clipping, resolution standardization, random cropping, flipping, rotation, and intensity shifting. The Houndsfield Units (HU) of the CBCT scans were normalized by clipping values outside the range of −450 HU to 250 HU, and the resolution was standardized to 1.5, 1.5, and 2.0 mm for all scans. In addition, we augmented the data set by randomly cropping the volume images to 96, 96, and 96 sizes and randomly flipping and rotating them with a probability of 0.1 for all 3 axes. To further improve the intensity information, we randomly shifted the intensity range with a probability of 0.5 and an offset of 0.1. The DL model was optimized using a combination of dice and cross-entropy losses as the loss function, with the AdamW optimizer and GradScaler for

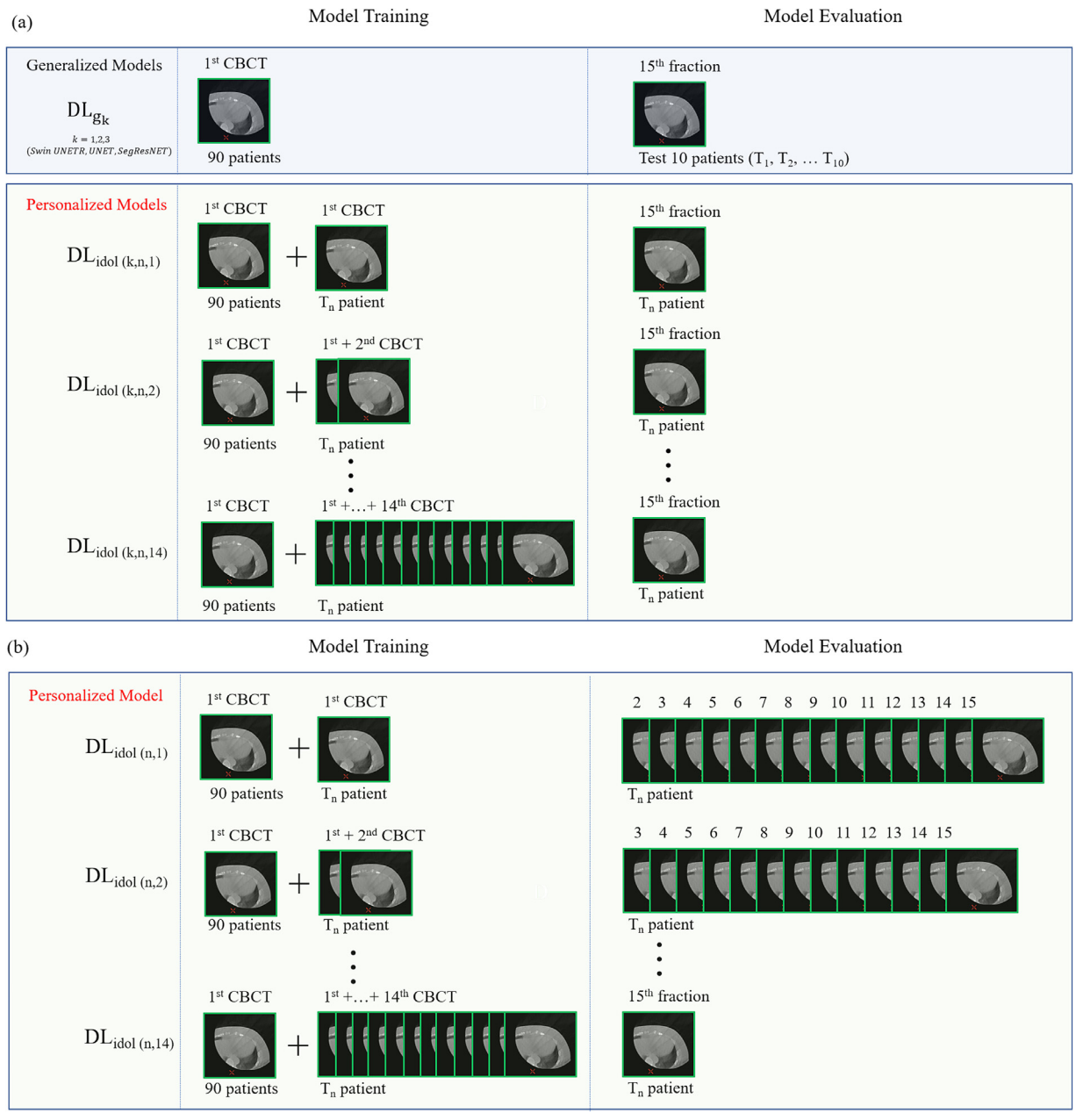

Figure 2 (a) The workflow demonstrating the efficacy of the IDOL framework using multiple networks, enhancing the segmentation accuracy of breast CTV for the 15th fraction. Conversely, (b) illustrates a workflow that emphasizes the optimal juncture of the IDOL effect. This considers additional training involving varying numbers of patient fractions and employs the best model architecture used in (a).
Abbreviations: CTV = clinical target volume; IDOL = Intentional Deep Overfit Learning.

scaling. We employed a learning rate of 1.00E-04 and a weight decay of 1.00E-05. In the first stage of the IDOL framework, we pretrained the network using 30,000 iterations with CBCT scans from 90 patients. In the second stage, we loaded the pretrained network and trained it for 50 epochs using the first to 14th fractions of CBCT scans of the patients, followed by an evaluation of the patient-specific network on the remaining fractions. (See Appendix E2 for further details.)

Evaluation

To evaluate network performance, we used the Dice Similarity Coefficient (DSC), 95 percentile Hausdorff distance (HD), and mean surface distance (MSD). The DSC is a statistical measure commonly used to assess the similarity between 2 sets or regions in data analysis, particularly in image segmentation tasks. The DSC equation is represented by equation (1):

$$\mathbf{DSC} = \frac{2 \times |\boldsymbol{X} \cap \boldsymbol{Y}|}{|\boldsymbol{X}| + |\boldsymbol{Y}|} \quad (1)$$

X represents the network output volume, and Y represents the corresponding ground truth of the reference label volume.

The HD is a mathematical measure that quantifies the extent of dissimilarity between 2 subsets of a metric space. Regarding image analysis or point cloud data, HD measures how far 1 set can be from another while remaining within a certain distance of the closest point on the other set. It provides a robust method to compare the similarity or dissimilarity between 2 sets by considering both the closeness and spread of points in the sets. HD is particularly useful for evaluating the discrepancy between segmented regions in medical imaging or any other scenario where sets of points or shapes need to be compared.[28] The HD equation is represented in equation (2):

$$\mathrm{H(X,\ Y)} = \max\left(sup_{x \in X} inf_{y \in Y} d(x,y), sup_{y \in Y} inf_{x \in X} d(x,y)\right) \quad (2)$$

$d(x,y)$ represents the distance between points x and y in metric space. sup denotes supremum, which is the lowest upper bound. inf denotes the infimum, which is the highest lower bound.

MSD is a metric used to quantify the average distance between corresponding points on the surfaces of 2 objects or regions. It is commonly employed to evaluate the accuracy of segmentation results, registration methods, or other spatial mapping tasks in medical image analysis and computer vision. The MSD equation is represented by equation (3):

$$\mathrm{MSD} = \frac{1}{N}\sum_{i=1}^{N} d(x_i, y_i) \quad (3)$$

N is the total number of corresponding point pairs, and $\boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{y_i})$ represents the distance between corresponding points $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ on the surfaces of the 2 objects or regions being compared.

To ascertain the statistical significance between the 2 data sets, we employed the differences in the means for DSC, HD, and MSD as metrics. Assuming a normal distribution for the 2 data sets, an independent samples t test was conducted to compare the means of 2 distinct samples derived from separate populations. By analyzing the average DSC, HD, and MSD differences between the 2 data sets, we decided whether to accept or reject the null hypothesis. The null hypothesis postulated that the average DSC, HD, and MSD values between the 2 data sets are identical, and the significance level chosen is below 0.05.

Our analysis involved comparing the data sets for 2 scenarios: first, between the generalized DL models and the deformed planning CT, where the planning CT was matched to the CBCT using a deformable image registration technique, and the resulting vector field was applied to the planning CT's CTV for evaluation; second, between the generalized DL model and IDOL model trained with multiple fractions. Additionally, we compared the data sets for the IDOL model trained with subsequent fractions. All statistical evaluations were conducted using the R programming language.

Results

Generalized Models with IDOL result

The accuracy of the breast CTV contours generated by the DL models was assessed using DSC. For the 15th fraction of CBCT of the patients, the average DSC values were 0.9611, 0.9646, 0.9634, and 9322 for the Swin UNETR, UNET, SegResNET, and deformed planning CT, respectively. Additionally, the training process was evaluated through the examination of loss graphs. The loss graphs of the networks demonstrate a notable convergence, indicating effective learning during the training phase. Specifically, both the network train loss and the validation mean DSC exhibit pronounced convergence, signifying the robust performance and generalization capabilities of the models. Further details, including the loss graph trajectories, can be found in Appendix E3. Importantly, considering the computational cost aspect, the training durations for each model were 210 minutes for Swin UNETR, 135 minutes for UNET, and 141 minutes for SegResNET, using the NVIDIA TITAN RTX. The inference time for processing a single patient averaged around 120 seconds. These computational costs highlight the efficiency and practical feasibility of implementing the Swin UNETR, UNET, and SegResNET models in our study. Notably, when incorporating the IDOL-enhanced first fraction training results, the average DSC values for the 15th fraction substantially improved to 0.9819, 0.9806, and 0.9788 for Swin UNETR, UNET, and SegResNET, respectively. Table 2 illustrates the average DSC plot for the DL models with the IDOL outcomes of Fig. 2a.

The average HD values for Swin UNETR, UNET, SegResNET, and deformed planning CT were 4.0118, 3.5295, 3.6574, and 5.9585 mm, respectively. Notably, when incorporating the IDOL-enhanced first fraction training results, the average HD values for the 15th fraction decreased substantially to 1.3935, 1.5039, and 1.6675 mm for Swin UNETR, UNET, and SegResNET, respectively. Table 2 illustrates the average HD plot for the DL models with the IDOL outcomes of Fig. 2a.

The average MSD values were 0.8723, 0.8530, 0.8772, and 2.053 for Swin UNETR, UNET, SegResNET, and deformed planning CT, respectively. Notably, when incorporating the IDOL-enhanced first fraction training results, the average MSD values for the 15th fraction

Table 2 The average DSC, HD, and MSD of the IDOL framework using several networks with standard deviation, as illustrated in Fig. 2a

Evaluation metric	Network	Train step (pretrain + fractions)														
		Pre	Pre+1	Pre+2	Pre+3	Pre+4	Pre+5	Pre+6	Pre+7	Pre+8	Pre+9	Pre+10	Pre+11	Pre+12	Pre+13	Pre+14
DSC	Swin UNETR	0.9612 ± 1.138E-2	0.9819 ± 2.570E-3	0.9835 ± 2.573E-3	0.9843 ± 2.628E-3	0.9847 ± 2.619E-3	0.9853 ± 1.918E-3	0.9856 ± 2.146E-3	0.9861 ± 1.913E-3	0.9861 ± 2.551E-3	0.9860 ± 1.939E-3	0.9865 ± 1.547E-3	0.9868 ± 1.837E-3	0.9865 ± 1.595E-3	0.9869 ± 1.951E-3	0.9872 ± 1.848E-3
	UNET	0.9646 ± 1.204E-2	0.9806 ± 2.842E-3	0.9826 ± 2.213E-3	0.9829 ± 2.342E-3	0.9834 ± 2.179E-3	0.9840 ± 2.190E-3	0.9845 ± 2.333E-3	0.9845 ± 2.109E-3	0.9847 ± 1.735E-3	0.9846 ± 2.049E-3	0.9845 ± 2.360E-3	0.9854 ± 2.490E-3	0.9855 ± 1.960E-3	0.9848 ± 2.213E-3	0.9856 ± 2.192E-3
	SegResNET	0.9634 ± 1.089E-2	0.9788 ± 2.574E-3	0.9809 ± 2.761E-3	0.9824 ± 2.119E-3	0.9820 ± 2.541E-3	0.9830 ± 1.923E-3	0.9829 ± 1.938E-3	0.9831 ± 2.481E-3	0.9834 ± 2.264E-3	0.9843 ± 2.028E-3	0.9837 ± 2.230E-3	0.9834 ± 1.841E-3	0.9843 ± 2.025E-3	0.9842 ± 2.674E-3	0.9846 ± 2.120E-3
HD	Swin UNETR	4.012 ± 1.797E0	1.393 ± 4.193E-1	1.271 ± 3.080E-1	1.124 ± 2.001E-1	1.166 ± 2.139E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1	1.041 ± 1.310E-1	1.083 ± 1.746E-1	1.000 ± 0.000E0	1.000 ± 0.000E0
	UNET	3.530 ± 1.869E0	1.504 ± 3.919E-1	1.224 ± 3.353E-1	1.239 ± 2.691E-1	1.166 ± 2.139E-1	1.166 ± 2.139E-1	1.124 ± 2.001E-1	1.207 ± 2.183E-1	1.124 ± 2.001E-1	1.183 ± 3.348E-1	1.207 ± 2.183E-1	1.124 ± 2.001E-1	1.083 ± 1.746E-1	1.207 ± 2.183E-1	1.083 ± 1.746E-1
	SegResNET	3.657 ± 1.950E0	1.668 ± 3.644E-1	1.522 ± 3.061E-1	1.266 ± 3.300E-1	1.363 ± 2.155E-1	1.166 ± 2.139E-1	1.239 ± 2.691E-1	1.197 ± 2.709E-1	1.207 ± 2.183E-1	1.083 ± 1.746E-1	1.207 ± 2.183E-1	1.166 ± 2.139E-1	1.124 ± 2.001E-1	1.083 ± 1.746E-1	1.083 ± 1.746E-1
MSD	Swin UNETR	0.8723 ± 2.434E-1	0.4603 ± 6.925E-2	0.4264 ± 6.620E-2	0.4076 ± 5.164E-2	0.3980 ± 7.058E-2	0.3871 ± 4.760E-2	0.3754 ± 4.103E-2	0.3607 ± 5.162E-2	0.3634 ± 5.121E-2	0.3673 ± 5.530E-2	0.3563 ± 4.727E-2	0.3449 ± 4.763E-2	0.3559 ± 4.662E-2	0.3394 ± 3.942E-2	0.3354 ± 3.870E-2
	UNET	0.8530 ± 2.648E-1	0.5027 ± 1.004E-1	0.4525 ± 7.873E-2	0.4480 ± 6.757E-2	0.4335 ± 6.556E-2	0.4205 ± 5.856E-2	0.4084 ± 6.017E-2	0.4081 ± 6.878E-2	0.3988 ± 5.496E-2	0.4044 ± 7.103E-2	0.4085 ± 6.336E-2	0.3829 ± 5.724E-2	0.3811 ± 5.916E-2	0.3971 ± 5.162E-2	0.3767 ± 5.973E-2
	SegResNET	0.8772 ± 2.349E-1	0.5466 ± 8.870E-2	0.4885 ± 1.032E-1	0.4617 ± 6.703E-2	0.4661 ± 6.147E-2	0.4410 ± 3.815E-2	0.4432 ± 6.166E-2	0.4417 ± 7.569E-2	0.4359 ± 5.948E-2	0.4098 ± 4.737E-2	0.4314 ± 6.816E-2	0.4360 ± 4.835E-2	0.4117 ± 5.671E-2	0.4166 ± 5.460E-2	0.4054 ± 5.368E-2

Abbreviations: DSC = Dice Similarity Coefficient; HD = Hausdorff distance; MSD = mean surface distance.

substantially decreased to 0.4603, 0.5027, and 0.5466 for Swin UNETR, UNET, and SegResNET, respectively. Table 2 illustrates the average MSD plot for the DL models with the IDOL outcomes of Fig. 2a, revealing the notable accuracy of the DL models in generating breast CTV contours.

We selected Swin UNETR because of its capacity to achieve high-accuracy contours, particularly when combined with the IDOL framework. Statistical analysis indicated significant differences in the average DSC for a patient's 15th fraction between Swin UNETR and UNET and between Swin UNETR and SegResNET (P values $<$ .05). Similarly, for average DSC data sets using IDOL results from the first fraction training, significant differences were observed between Swin UNETR and UNET and between Swin UNETR and SegResNET (P values $<$ .05). However, the analyses for the average HD and MSD data sets yielded no significant differences for Swin UNETR compared with UNET and SegResNET.

IDOL framework of finding the optimal point

By implementing the IDOL framework using the CBCT scans acquired during the initial treatment, noteworthy improvements were observed in the average DSC for the 15th fraction, which increased from 0.9611 to 0.9819. Across all patients and fractions spanning the second to 15th, the average DSC increased from 0.9622 to 0.9817. Notably, further enhancements were achieved by incorporating CBCT scans from the first to third fractions for training, yielding an average DSC of 0.9850 for the 15th fraction and 0.9843 for all patients and fractions from the 4th to 15th. Table 3 shows the average DSCs of the IDOL framework using Swin UNETR, which is a result of Fig. 2b.

In addition, a preliminary statistical analysis was conducted to compare the performance of the generalized DL model (Swin UNETR) and IDOL model trained with varying patient fractions. The results showed a significant improvement in the segmentation accuracy with the IDOL model (P value $<$ .05) (see more in Appendix E4). Subsequently, a more detailed statistical assessment was conducted to compare the performance of the IDOL model trained with subsequent fractions. A comparison between the IDOL model trained using the first and second fractions and that trained with the first to third fractions demonstrated a significant enhancement in segmentation accuracy (P value $<$ .05). However, for subsequent comparisons, the differences were not significant (P values $>$ .05) (see more on Appendix E5).

Furthermore, the implementation of the IDOL framework using the CBCT scans from the initial treatment led to noteworthy improvements in the average HD for the 15th fraction, reducing it from 4.0118 mm to 1.3935 mm. Similarly, for all patients and fractions spanning the second to 15th, the average HD decreased from 4.6879 mm to 1.5649 mm. Similar to the DSC analysis, the incorporation of CBCT scans from the first to third fractions for training further amplified these improvements, which resulted in an average HD of 1.2707 mm for the 15th fraction and 1.3616 mm for all patients and fractions from the 4th to 15th. Table 3 presents the average HDs of the IDOL framework using the Swin UNETR.

Regarding the MSD, the implementation of the IDOL framework using the CBCT scan from the patient's initial treatment led to noteworthy improvements in the average MSD for the 15th fraction, reducing it from 0.8723 to 0.4603. Similarly, for all patients and fractions spanning the second to 15th, the average MSD decreased from 0.8530 to 0.5027. Similar to the DSC analysis, the incorporation of CBCT scans from the first to third fractions for training further amplified these improvements, which resulted in an average MSD of 0.4076 for the 15th fraction and 0.4480 for all patients and fractions from the 4th to 15th. Table 3 presents the average MSD results of the IDOL framework using the Swin UNETR.

Overall, these findings emphasize the substantial improvements achieved through the IDOL framework in terms of both DSC and HD metrics and highlight the significance of these enhancements. The IDOL training process required an average of 107s for 1 CBCT scan and 331s for 3 CBCT scans.

IDOL result comparison with image

Figure 3a-c presents a visual comparison of a reference contour manually drawn by a physician, the CTV output of deformed planning CT output, CTV output of the pretrained network (Swin UNETR), and CTV output of the additional training using the first fraction. The visualizations are provided in the horizontal, sagittal, and coronal planes. The CTV output obtained from additional training with the first fraction demonstrated a significant improvement in contouring accuracy compared with the pretrained network output or the deformed planning CT. This improvement aligns with the higher DSC scores and lower HD and MSD values observed, indicating the enhanced performance of the model. The reference contour served as the gold standard for comparison, and additional training incorporating patient-specific information resulted in a contour that closely matched the contour drawn by the physician (Fig. 3d-f) (see more in Appendix E6). The red and blue colors represent the difference between the 3D contour drawn by the physician as seen in Fig. 3d-f.

Discussion

In recent years, considerable progress has been made in the field of IGRT. Notably, studies have investigated

Table 3 The average DSC, HD, and MSD of the IDOL framework using Swin UNETR with standard deviation, as illustrated in Fig. 2b

Evaluation metric	Train step (pretrain + fractions)														
	Pre	Pre+1	Pre+2	Pre+3	Pre+4	Pre+5	Pre+6	Pre+7	Pre+8	Pre+9	Pre+10	Pre+11	Pre+12	Pre+13	Pre+14
DSC	0.9622 ± 1.058E-2	0.9810 ± 2.523E-3	0.9836 ± 1.704E-3	0.9844 ± 1.804E-3	0.9848 ± 1.907E-3	0.9852 ± 1.596E-3	0.9858 ± 1.945E-3	0.9858 ± 1.500E-3	0.9857 ± 2.732E-3	0.9858 ± 3.294E-3	0.9864 ± 2.430E-3	0.9863 ± 1.643E-3	0.9867 ± 1.902E-3	0.9869 ± 2.386E-3	0.9876 ± 1.848E-3
HD	4.688 ± 4.369E0	1.565 ± 3.281E-1	1.362 ± 1.565E-1	1.225 ± 1.607E-1	1.213 ± 2.027E-1	1.162 ± 1.236E-1	1.155 ± 1.416E-1	1.121 ± 1.080E-1	1.130 ± 1.298E-1	1.130 ± 1.737E-1	1.159 ± 1.734E-1	1.111 ± 1.273E-1	1.130 ± 1.299E-1	1.043 ± 8.647E-2	1.000 ± 0.000E0
MSD	1.107 ± 7.772E-1	0.4615 ± 4.735E-2	0.4247 ± 5.233E-2	0.4064 ± 4.765E-2	0.3971 ± 5.849E-2	0.3947 ± 4.567E-2	0.3755 ± 3.935E-2	0.3685 ± 4.306E-2	0.3712 ± 4.915E-2	0.3740 ± 6.374E-2	0.3617 ± 4.675E-2	0.3532 ± 4.034E-2	0.3643 ± 4.711E-2	0.3481 ± 4.125E-2	0.3354 ± 3.970E-2

Abbreviations: DSC = Dice Similarity Coefficient; HD = Hausdorff distance; MSD = mean surface distance.

the application of CBCT for precise IGRT in head, neck, and prostate treatments.[14,15] These endeavors have contributed substantially to the development of IGRT protocols, ultimately enhancing the accuracy of RT delivery. Furthermore, the integration of DL methodologies into the field of medical imaging has marked a transformative phase. Recent investigations have effectively used DL to automate the segmentation of breast masses in ultrasound and diagnose breast cancer through MRI.[16,17] These advancements underscore the burgeoning role of artificial intelligence in augmenting breast cancer diagnosis and management.

Despite the advances in the application of DL in medical image processing, the use of DL for segmentation of daily CBCT scans has not been attempted. We employed DL techniques for the automated segmentation of the CTV of the breast using daily CBCT scans, which is important to address a fundamental need in breast cancer RT. Herein, the challenge was to ensure a continuous and dynamic segmentation of the breast CTV, accounting for variations in patient anatomy and positioning throughout the treatment course. Unlike previous studies, our research was focused exclusively on this task, aiming to establish a robust solution for daily CBCT-based breast CTV delineation.[14-17]

The introduction of the IDOL framework is central to this study. Our approach is novel due to its extensive use of real clinical ART patient data for testing, which enabled the construction of patient-specific DL networks tailored to the evolving anatomy and treatment progress of patients. The impact of the IDOL framework was profound, yielding remarkable improvements across key metrics, such as DSC, HD, and MSD. Compared with other studies, we achieved the highest DSC score, exceeding 0.98, whereas other studies typically reported scores below 0.9.[16,17] Similarly, our approach yielded an HD value <2 mm, a critical achievement for breast CTV segmentation, as other studies have mostly reported values >2 mm when employing the IDOL framework.[14,15] This is primarily because of the unique challenges posed by breast anatomy.

These outcomes underscore the immense potential of the IDOL framework for enhancing the precision and efficiency of breast CTV segmentation using CBCT scans. Through rigorous experimentation, we identified the optimal points for implementing the IDOL framework to ensure its practical efficacy. Most notably, our patient-specific model not only streamlined the labor-intensive manual segmentation process but also reduced the time demands. Crucially, it significantly enhanced the segmentation accuracy, establishing our approach as a compelling and versatile solution for a range of ART applications.

Our study further demonstrated that physicians could confidently perform CTV contouring up to the third fraction, which led to substantial improvements in DSC scores, reduced HD and MSD values, and reduced labor time. This study underscores the effectiveness of the

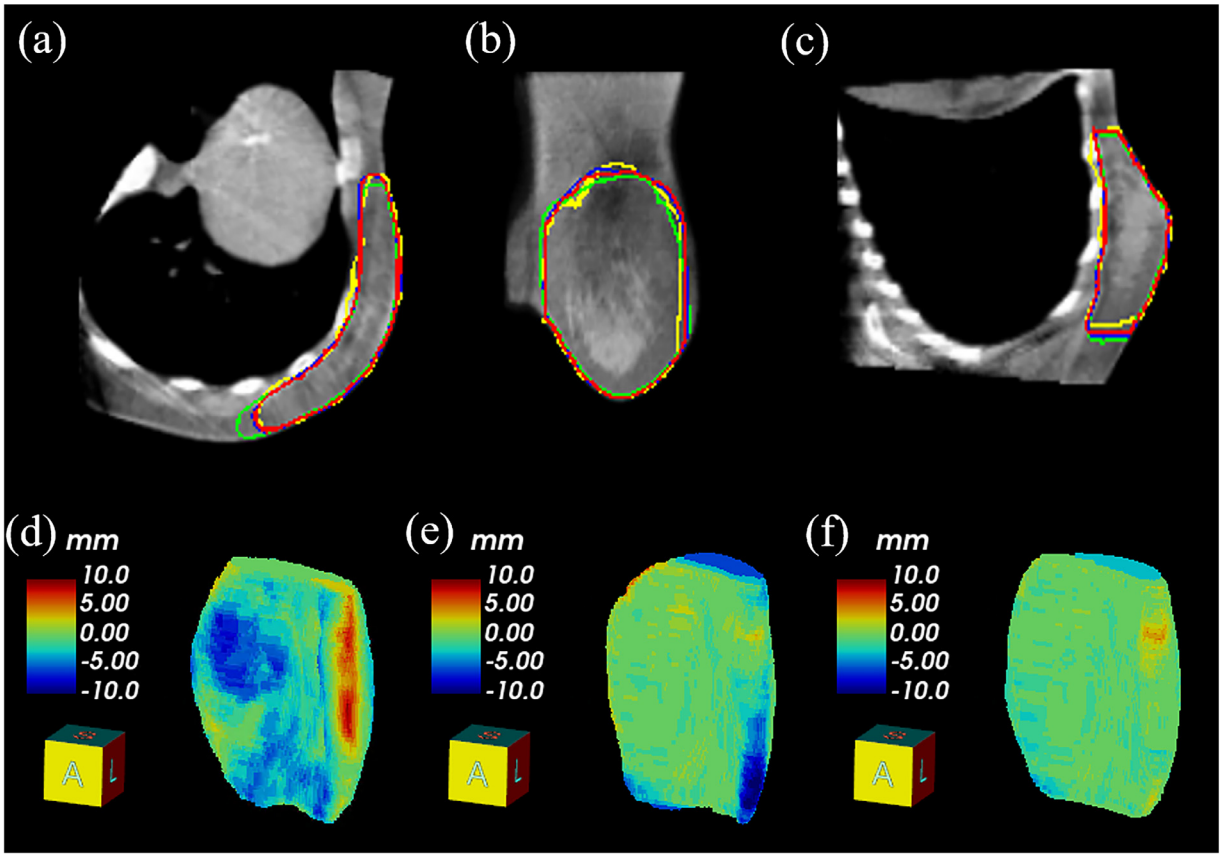

Figure 3 (a-c) Comparison of reference contour (red), deformed planning CT output (yellow), pretrained (Swin UNETR) network CTV output (green), and additional training with the first fraction CTV output (blue). (a) Horizontal axial view. (b) Sagittal axial view. (c) Coronal axial view. (d) Comparison between reference contour and deformed planning CT output. (e) Comparison between reference contour and network (Swin UNETR) CTV output. (f) Comparison between reference contour and additional training with the first fraction CTV output.
Abbreviation: CTV = clinical target volume.

IDOL framework in enhancing the precision and efficiency of CTV contouring in breast cancer RT. In actual clinical routine, the generalized DL model can be trained with a large data set initially. When a new patient comes in for treatment requiring more than 15 fractions, the model can be trained up to the third fraction. After this point, the IDOL network can be used, allowing physicians to make minor adjustments and apply the model clinically with ease. As demonstrated in Table 3 and through *P* value comparisons, the results obtained after training up to the third fraction were not significantly different from those obtained by training beyond the third fraction. Moreover, training and applying the model clinically after the third fraction showed better and more reliable results compared with training only up to the first or second fractions. Therefore, training up to the third fraction and then applying the model in clinical practice is highly recommended. These findings are also of importance for ART and offer advancements in IGRT for patients with breast cancer while mitigating the workload on physicians.

However, our study had some limitations. Reliance on a single physician for CTV contouring and evaluation introduced a potential constraint considering the possibility of interobserver variability stemming from different contouring techniques. Furthermore, our evaluation focused exclusively on breast CTV, limiting the applicability of the findings to other anatomic regions. To enhance the robustness and broad applicability of the IDOL framework, future investigations should involve multiple physicians and diverse CTVs to ensure comprehensive insights across various clinical contexts. Exploring the adaptability of the IDOL framework to different DL models, data sets, and cancer sites is a promising avenue for advancing IGRT and optimizing treatment outcomes. Additionally, the exploration of radiomics and volume studies of breast tumor growth prognosis provides intriguing opportunities for further research.

Conclusion

Our study introduced a transformative patient-specific DL-based algorithm that significantly enhanced the precision and efficiency of breast CTV segmentation in CBCT

scans. These findings may elevate IGRT in adjuvant breast cancer treatment and are relevant for various ART applications, offering benefits such as labor reduction, time efficiency, and heightened segmentation accuracy. In the future, the adaptability of the IDOL framework to diverse DL models, data sets, and cancer sites will be explored, with the potential to substantially enhance treatment outcomes across a broader patient spectrum.

Disclosures

None.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.adro.2024.101580.

References

1. Group EBCTC. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. *Lancet*. 2011;378:1707-1716.
2. McGale P, Correa C, Cutter D, et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet*. 2014;383:2127-2135.
3. Haffty B. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Breast Dis Year Book Q*. 2014;25:343-344.
4. Whelan TJ, Olivotto IA, Parulekar WR, et al. Regional nodal irradiation in early-stage breast cancer. *N Engl J Med*. 2015;373:307-316.
5. Poortmans PM, Collette S, Kirkove C, et al. Internal mammary and medial supraclavicular irradiation in breast cancer. *N Engl J Med*. 2015;373:317-327.
6. Poortmans PM, Weltens C, Fortpied C, et al. Internal mammary and medial supraclavicular lymph node chain irradiation in stage I–III breast cancer (eortc 22922/10925): 15-year results of a randomised, phase 3 trial. *Lancet Oncol*. 2020;21:1602-1610.
7. Chang JS, Chang JH, Kim N, Kim YB, Shin KH, Kim K. Intensity modulated radiotherapy and volumetric modulated arc therapy in the treatment of breast cancer: an updated review. *J Breast Cancer*. 2022;25:349-365.
8. Ahmad SS, Duke S, Jena R, Williams MV, Burnet NG. Advances in radiotherapy. *BMJ*. 2012;345:e7765.
9. Smith GL, Smith BD. Sea change: a decade of intensity-modulated radiation therapy for treatment of breast cancer. Editor, editor editors. *Book Sea Change: A Decade of Intensity-Modulated Radiation Therapy for Treatment of Breast Cancer*. Oxford University Press; 2020:221-223.
10. Offersen BV, Boersma LJ, Kirkove C, et al. Estro consensus guideline on target volume delineation for elective radiation therapy of early stage breast cancer. *Radiother Oncol*. 2015;114:3-10.
11. White J, Tai A, Arthur D, et al. Breast cancer atlas for radiation therapy planning: consensus definitions. *Book Breast Cancer Atlas Radiat Ther Plan*. 2009;73:944-951.
12. Sá AC, Fermento A, Neves D, et al. Radiotherapy setup displacements in breast cancer patients: 3D surface imaging experience. *Rep Pract Oncol Radiother*. 2018;23:61-67.
13. Laaksomaa M, Sarudis S, Rossi M, et al. Alignrt® and catalyst™ in whole-breast radiotherapy with dibh: Is igrt still needed? *J Appl Clin Med Phys*. 2019;20:97-104.
14. Juan-Senabre XJ, López-Tarjuelo J, Conde-Moreno A, et al. Uncertainties and ctv to ptv margins quantitative assessment using cone-beam CT technique in clinical application for prostate, and head and neck irradiation tumours. *Clin Transl Oncol*. 2011;13:819-825.
15. Maund I, Benson R, Fairfoul J, Cook J, Huddart R, Poynter A. Image-guided radiotherapy of the prostate using daily cbct: The feasibility and likely benefit of implementing a margin reduction. *Br J Radiol*. 2014;87: 20140459.
16. Webb JM, Adusei SA, Wang Y, et al. Comparing deep learning-based automatic segmentation of breast masses to expert interobserver variability in ultrasound imaging. *Comput Biol Med*. 2021;139: 104966.
17. Yue W, Zhang H, Zhou J, et al. Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging. *Front Oncol*. 2022;12: 984626.
18. Kim YJ, Kim KG. Detection and weak segmentation of masses in gray-scale breast mammogram images using deep learning. *Yonsei Med J*. 2022;63:S63.
19. Kim S, Kim E-H, Kim H-S. Physician knowledge base: clinical decision support systems. *Yonsei Med J*. 2022;63:8.
20. Chun J, Park JC, Olberg S, et al. Intentional deep overfit learning (idol): a novel deep learning strategy for adaptive radiation therapy. *Med Phys*. 2022;49:488-496.
21. Kim N, Chang JS, Shah C, et al. Hypofractionated volumetric-modulated arc therapy for breast cancer: a propensity-score-weighted comparison of radiation-related toxicity. *Int J Cancer*. 2021;149:149-157.
22. Group ST. The uk standardisation of breast radiotherapy (start) trial b of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *Lancet*. 2008;371:1098-1107.
23. Haviland JS, Owen JR, Dewar JA, et al. The uk standardisation of breast radiotherapy (start) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. *Lancet Oncol*. 2013;14:1086-1094.
24. Tang Y, Yang D, Li W, et al. Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 202220730-20740.
25. Hatamizadeh A, Nath V, Tang Y, et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in MRI images. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part I*. Springer; 2022:272-284.
26. Falk T, Mai D, Bensch R, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nat Method*. 2019;16:67-70.
27. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*. Springer; 2019:311-320.
28. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 1993;15:850-863.