ISSN (Print) 1226-1750
ISSN (Online) 2233-6656
Journal of Magnetics 30(4), 764-775 (2025) https://doi.org/10.4283/JMAG.2025.30.4.764

A Study on the Clinical Application of AI-based Auto-Segmentation for Target Volumes in Patients with Breast Cancer after Breast-Conserving Surgery

Jun-Taek Shin[1,2], Jun-Bong Shin[2], Minsik Lee[2], Sumin Lee[2], Nuri Hyun Jung[2], Qichao Zhou[3], Wonyoung Cho[4], Minji Koh[2]*, and Man-Seok Han[1,5]*

[1]*Department of Health Medical Science, Graduate School, Kangwon National University, Samcheok, Republic of Korea*
[2]*Department of Radiation Oncology, Kangwon National University Hospital, Baengnyeong-ro 156, Chuncheon, Republic of Korea*
[3]*Manteia Technologies Co., Ltd., 3001-3005, No.5 Huizhan North Road, Xiamen, Fujian, P.R. China*
[4]*Oncosoft Inc., 37 Myeongmul-gil, Seodaemun-gu, Seoul, Republic of Korea*
[5]*Department of Radiological Science, Kangwon National University, Samcheok, Republic of Korea*

(Received 31 October 2025, Received in final form 8 December 2025, Accepted 9 December 2025)

We evaluated artificial intelligence (AI)-based auto-segmentation models in patients with breast cancer undergoing surgery and radiation therapy. Radiation oncologists manually defined clinical target volume (CTV) and planning target volume (PTV) in 100 cases to train Acculearning 2.2.3.182 and OncoStudio 2.0.4, with models automatically contouring CTV and PTV, showing acceptable agreement with manual contours using Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and mean surface distance (MSD). In four cases, manual contouring took 548 ± 205 s, whereas AI-assisted contouring with manual revision took 187 ± 40 s. The paired *t*-test revealed significant accuracy improvements with the 3D approach for RNI cases ($p < 0.05$) and laterality-dependent differences in WBI cases ($p < 0.05$). These findings highlight the need to consider treatment extent and anatomical laterality in AI auto-segmentation. Deep learning segmentation speeds up contouring and enhances workflow efficiency in radiation therapy planning.

Keywords : AI-based auto-segmentation, electromagnetic radiation, breast cancer, regional nodal irradiation, clinical target volume, planning target volume

1. Introduction

In traditional radiation therapy planning, outlining the treatment volume on simulation CT (sim-CT) images is entirely the responsibility of radiation oncologists. The manual contouring process is time-consuming and susceptible to interobserver and intra- and interinstrument variability, leading to differences in target delineation [1-5]. With advances in artificial intelligence (AI), this technology has emerged as a potential tool for optimizing radiation therapy across all stages, from sim-CT segmentation to treatment planning and dose delivery [6-8]. AI is increasingly applied to dose prediction, adaptive planning, and adaptive radiation therapy, with its clinical utility rapidly evolving [9]. Among these applications, AI-based auto-segmentation automates the delineation of targets and organs at risk (OARs), improving workflow efficiency and mitigating interobserver variability [10-13]. This approach generates accurate and consistent contours and reduces the time required for manual contouring [14-18]. With the increasing incidence of breast cancer and the growing number of patients receiving radiation therapy, treatment techniques have evolved from two-dimensional (2D) planning to stereotactic approaches and intensity-modulated radiation therapy. As the use of comprehensive regional nodal irradiation (RNI) has expanded through clinical trials, the target volumes have become more complex, necessitating adherence to international guidelines [19,20]. Discrepancies in OAR contours increase dose uncertainty, compromising treatment efficacy and elevating the risk of complications in patients. Recent advances in computing performance, algorithms, and data collection have accelerated AI research. Men *et al.* [21] developed a deep learning-based model for breast target

*Corresponding author:
Tel: +82-33-540-3383, Fax: +82-33-540-3389,
e-mail: angio7896@naver.com(Man-Seok Han)
Tel: +82-33-258-9446, Fax: +82-33-258-9472
e-mail: mjkoh121@gmail.com(Minji Koh)

delineation, and several vendors have since released an AI-based auto-segmentation platform. As AI-assisted contouring becomes more common, validation is essential for safe clinical integration, and identifying the key components for model optimization remains critical [22]. This study evaluated two AI-based auto-segmentation models in patients undergoing breast-conserving surgery (BCS) and receiving whole breast irradiation (WBI) with or without regional node irradiation (RNI). The performance was assessed based on the contouring accuracy and time efficiency in the clinical workflow.

2. Materials and Methods

2.1. Patients with Breast Cancer

This study was reviewed and approved by the Institutional Review Board of Kangwon National University Hospital (IRB No. 2025-05-001). All imaging and contour data were deidentified prior to analysis.

The study included 100 patients with breast cancer who received adjuvant radiation therapy following BCS at our hospital between August 26, 2021, and April 1, 2025. The cohort comprised 50 patients with left-sided breast cancer and 50 with right-sided breast cancer. Among them, 60 patients underwent WBI, and 40 received RNI in addition to WBI. For patients treated with WBI alone, the distribution of cTNM staging at diagnosis was Tis in 7 (11.7%), T1 in 42 (70.0%), and T2 in 11 (18.3%); all were N0 (100%). For patients receiving WBI and RNI, the staging distribution was as follows: T1, 9 (22.5%), T2 22 (55%), T3 8 (20%); T4, 1 (2.5%); N0, 1 (2.5%), N1 29 (72.5%), N2 8 (20%); and N3, 2 (5%). The median patient age was 58 years (range, 34–86).

Sim-CT was performed in all 100 patients to delineate the clinical target volume (CTV) and planning target volume (PTV), and contrast enhancement was applied in patients undergoing RNI when clinically indicated. CT scans (Go-Sim, Siemens) were performed approximately 1 week before RT, with a slice thickness of 2 mm. The patients were positioned supine with both arms elevated on arm-support devices (CIVICO). Contrast-enhanced CT scans were obtained 1 min after intravenous injection of 100–115 mL of iodinated contrast (Ultraject, 320 mg/mL), which was adjusted according to the patient's weight.

2.2. Delineation

A radiation oncologist delineated the CTV and PTV for each patient with breast cancer. The target areas included the ipsilateral breast (CTVp_breast) and, depending on individual risk factors, axillary lymph node levels 1-3 (CTVn_L1-L3), supraclavicular lymph nodes (SCL; CTVn_L4), and/or internal mammary lymph nodes (CTVn_IMN). The CTV was delineated according to the ESTRO guidelines, with modifications applied by the radiation oncologist as clinically appropriate. The PTV was defined as a 4–5 mm expansion from the final CTV.

2.3. AI-based auto-segmentation (Deep learning framework)

We compared the AI-based auto-segmentation frameworks of two platforms: Acculearning 2.2.3 (Manteia Technologies Co.) and OncoStudio 2.0.4 (Oncosoft Inc., Seoul, South Korea). Both employ deep learning architectures based on the encoder-decoder U-Net framework with skip connections to preserve the spatial features. Dice loss is the primary loss function, with optional alternatives such as Focal Loss and Generalized Dice Loss. The Adam optimizer was used by default, and both platforms supported data augmentation and image normalization to enhance model generalizability. Validation strategies, including holdout testing and cross-validation, are consistently implemented.

There are key differences between these platforms. Acculearning provides a customizable training environment with a flexible configuration of network architectures (e.g., UNet, VNet, HighRes3DNet), activation functions (e.g., ReLU, Leaky-ReLU, Swish), and normalization techniques (BatchNorm, InstanceNorm, GroupNorm). It uses balanced sampling to address class imbalance and includes a postprocessing module to refine the predictions. In contrast, OncoStudio standardizes its training pipeline for comparability across 2D and 3D models. It employs residual U-Net architectures with PReLU activation and integrates deep supervision through auxiliary decoder outputs to improve gradient propagation. The 2D models process slice-by-slice data with adaptive channel scaling, whereas the 3D models process full-volumetric data with fixed depth and channel configuration.

Acculearning emphasizes dataset management and provides tools for automatic ROI name normalization, voxel-level statistical analysis, and format conversion (DICOM/NIFTI to internal format). These features are absent in OncoStudio, which focuses on harmonized optimization and validation procedures across 2D and 3D models. Overall, although both platforms share core architectural principles, Acculearning prioritizes flexibility and fine-tuning options, whereas OncoStudio emphasizes standardization and model comparability across modalities.

The manually contoured dataset created by radiation oncologists was divided into three subsets for training and evaluation: 80% for model training, 10% for validation (used for hyperparameter tuning and model selection),

and 10% for independent testing to assess the generalization performance. The models were trained using the AI-based 2D and 3D frameworks of each manufacturer. Each model approach is characterized by training independent deep learning networks for individual anatomical structures (e.g., CTV, PTV, CTVn–L1/L2/L3, SCL, and IMN), enabling the models to learn structure-specific anatomical characteristics and boundary features. This strategy enhances the boundary precision, particularly for small or morphologically irregular targets. However, limited feature sharing across structures may reduce learning efficiency, and total training time increases as the number of models increases.

In contrast, the sum model approach trains a single multiclass network that learns all anatomical label categories concurrently. This approach encourages shared representation of anatomical features across structures, mitigates class imbalance, and preserves global topological consistency, which is particularly beneficial for nodal chains. However, local boundary delineation may be compromised, particularly for small or low-contrast structures, owing to overlapping feature learning among adjacent labels. Both approaches were explored because the RNI includes multiple topologically connected structures, where shared feature learning across nodal levels may improve generalization. However, individual models may provide superior contour fidelity in individual lymph node stations, which is clinically critical for accurate dose distribution. The results were compared with manually delineated contours from radiation oncologists using datasets in which the CTV and PTV were trained separately and in combination.

2.4. Analysis

Using patient sim-CT data, 10 test datasets were prepared, each consisting of manually delineated contours by radiation oncologists and AI-based auto-segmentation results. These datasets were used to compare the performance of the two AI-based auto-segmentation models. Evaluation was performed using established geometric metrics: Dice similarity coefficient (DSC) [23, 24] Eq. (1), and 95% Hausdorff distance (HD95) [25] Eq. (2), and the mean surface distance (MSD) Eq. (3). For objective evaluation, MATLAB scripts were implemented to input the data, calculate the metrics, and compare the results. Statistical analyses were performed using paired two-tailed *t*-tests, given that the paired measurements originated from the same patient cohort. Comparative assessments were performed between Acculearning and OncoStudio, 2D and 3D segmentation frameworks, CTV and PTV, each-structure and sum-model training strategies, and between the left and right sides to evaluate potential laterality differences. Statistical significance was set at $p < 0.05$.

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

* |A| is the number of elements in set A
* |B| is the number of elements in set B
* |A ∩ B| is the number of elements common to both sets A and B (intersection)

$$\mathrm{HD(Gt, Pd)} = mean_{p_{pd} \in P_d} min_{p_{gt} \in G_t} \| p_{gt} - p_{pd} \|_2 \quad (2)$$

* Ppd represents the predicted pixels, and Pgt represents the ground truth pixels used for segmentation.

$$MSD = \frac{1}{(n_A + n_B)} \times [\Sigma d(a_i, surface\ B) + \Sigma d(b_i, surface\ A)] \quad (3)$$

* n_A is the number of points on Surface A.
* n_B is the number of points on Surface B.
* d(a_i, Surface B) is the distance from point a_i on Surface A to the closest point on Surface B.
* d(b_j, Surface A) is the distance from point b_j on Surface B to the closest point on Surface A.
* Sum() represents the sum of all distances.

Additionally, to assess the clinical utility of AI-based auto-segmentation, we compared the time required for clinicians to complete manual contouring on the test dataset with that required for manual refinement after AI-based auto-segmentation.

2.5. Comparison of target contouring times between manual contouring and AI-assisted contouring with manual correction

We prepared an additional dataset of 10 patients, including 6 who received WBI only and 4 who received WBI and RNI, to evaluate the utility of AI-based auto-segmentation for adjuvant radiation therapy planning after BCS. A radiation oncologist quantitatively compared the time required for manual contouring alone with that required for manual refinement after AI-based auto-segmentation.

3. Results

3.1. Comparison results of manual segmentation and AI-based automatic segmentation (Acculearning) based on 2D/3D (Each/Sum) learning methods when the radiation treatment range is limited to the WBI

We quantitatively compared the 2D and 3D segmentation methods for WBI planning. Overall, the 3D approach outperformed the 2D method in all evaluation metrics. For both the left and right breasts, the 3D segmentation achieved higher DSC values and lower HD95 and MSD values, indicating improved boundary accuracy.

For the left breast PTV, the 3D method achieved a DSC of 0.951, compared with 0.923 for the 2D method. In the right breast CTV, the 3D method also demonstrated superior performance (DSC 0.942 vs 0.923). The HD95

Table 1. DSC, HD95, and MSD results based on 2D/3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (Acculearning) when the radiation treatment range is limited to the WBI.

WBI			PTV			CTV		
			DSC	HD95(mm)	MSD(mm)	DSC	HD95(mm)	MSD(mm)
Lt Breast	2D	Each	0.923 (±0.02)	7.800 (±1.93)	1.417 (±0.25)	0.954 (±0.01)	3.040 (±1.06)	0.810 (±0.19)
		Sum	0.961 (±0.00)	2.377 (±0.72)	0.740 (±0.10)	0.951 (±0.00)	2.467 (±0.61)	0.857 (±0.11)
	3D	Each	0.951 (±0.01)	3.353 (±0.37)	0.943 (±0.16)	0.951 (±0.01)	2.720 (±0.63)	0.853 (±0.14)
		Sum	0.950 (±0.01)	3.387 (±0.54)	0.957 (±0.20)	0.949 (±0.01)	3.203 (±0.35)	0.883 (±0.17)
Rt Breast	2D	Each	0.926 (±0.03)	10.990 (±7.99)	1.550 (±0.82)	0.923 (±0.02)	11.667 (±6.66)	1.513 (±0.67)
		Sum	0.930 (±0.02)	10.387 (±6.57)	1.437 (±0.57)	0.926 (±0.02)	10.000 (±5.29)	1.423 (±0.51)
	3D	Each	0.942 (±0.00)	4.927 (±0.32)	1.177 (±0.18)	0.942 (±0.01)	4.323 (±0.70)	1.083 (±0.25)
		Sum	0.944 (±0.01)	4.750 (±1.29)	1.140 (±0.33)	0.942 (±0.01)	4.360 (±1.23)	1.077 (±0.31)

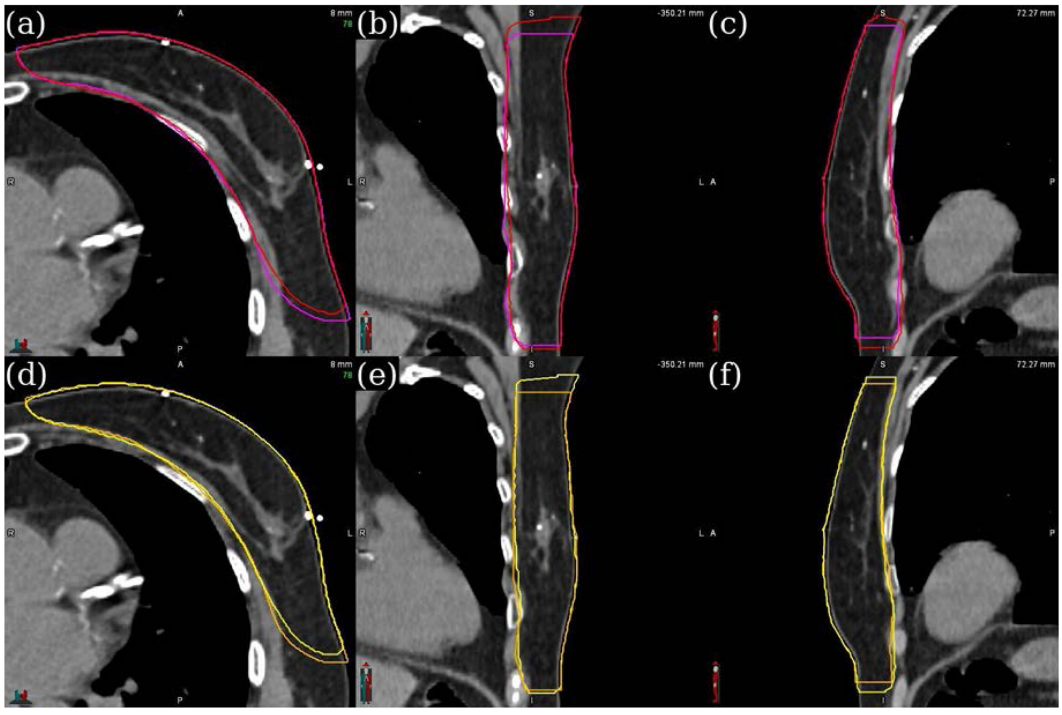

Fig. 1. (Color online) Results of 2D and 3D (Each/Sum) learning methods for manual and AI-based auto- segmentation (Acculearning) when the radiation treatment range was limited to the WBI.

and MSD values were consistently lower with the 3D technique, reflecting enhanced spatial agreement in contour delineation.

Notably, the right breast showed higher HD95 and MSD values than the left breast, suggesting greater segmentation complexity, possibly due to anatomical or motion-related factors. Additionally, CTVs generally showed higher DSC and lower distance metrics than PTVs, indicating greater delineation consistency. In the 3D Sum method, PTV results were: DSC 0.944 ± 0.01, HD95 4.750 ± 1.29 mm, and MSD 1.140 ± 0.33 mm; CTV results were: DSC 0.942 ± 0.01, HD95 4.360 ± 1.23 mm, and MSD 1.077 ± 0.31 mm.

In conclusion, 3D segmentation yielded superior quantitative accuracy compared to 2D segmentation. It may offer significant clinical advantages, particularly in cases requiring precise anatomical delineations (Table 1, Fig. 1-2).

3.2. Comparison results of manual segmentation and AI-based automatic segmentation (OncoStudio) based on 2D/3D (Each/Sum) learning methods when the radiation treatment range is limited to the WBI

This study also evaluated the performance of 2D and 3D segmentation methods in WBI radiation therapy planning. The 3D method consistently outperformed the 2D method across DSC, HD95, and MSD.

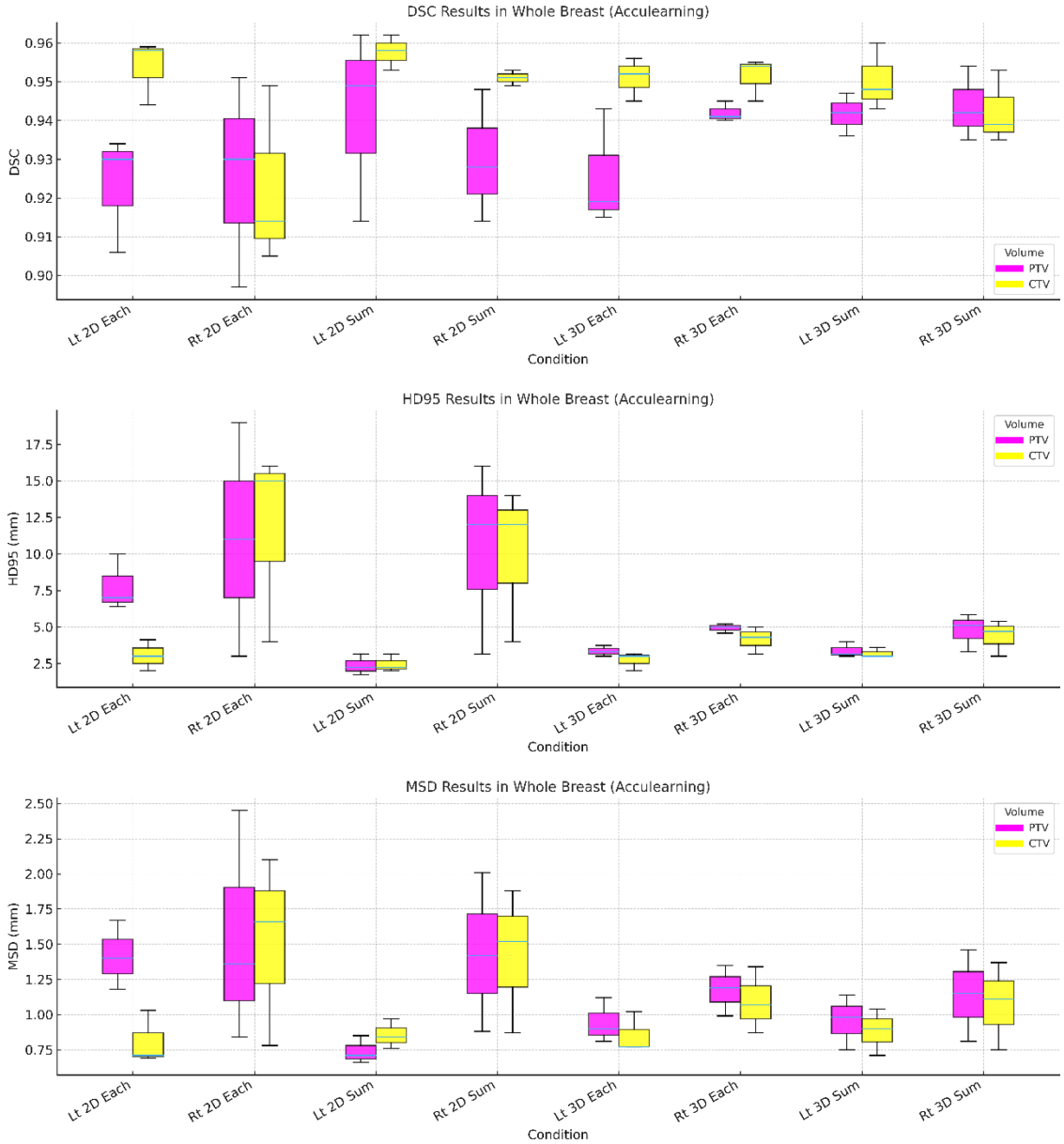

Fig. 2. (Color online) Results of DSC, HD95, and MSD based on 2D and 3D (Each/Sum) learning methods for manual and AI-based auto-segmentation (Acculearning) when the radiation treatment range was limited to the WBI.

Table 2. DSC, HD95, and MSD results based on 2D/3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (OncoStudio) when the radiation treatment range is limited to WBI.

WBI			PTV			CTV		
			DSC	HD95(mm)	MSD (mm)	DSC	HD95 (mm)	MSD (mm)
Lt Breast	2D	Each	0.940 (±0.01)	4.687 (±2.01)	1.097 (±0.25)	0.944 (±0.01)	3.667 (±0.58)	0.967 (±0.16)
		Sum	0.954 (±0.01)	2.817 (±0.32)	0.850 (±0.07)	0.952 (±0.01)	2.743 (±0.47)	0.820 (±0.06)
	3D	Each	0.954 (±0.01)	3.103 (±1.03)	0.830 (±0.23)	0.951 (±0.01)	3.243 (±1.49)	0.837 (±0.25)
		Sum	0.958 (±0.01)	3.173 (±0.94)	0.790 (±0.21)	0.956 (±0.01)	2.993 (±0.75)	0.740 (±0.17)
Rt Breast	2D	Each	0.901 (±0.04)	11.783 (±9.78)	2.063 (±1.14)	0.889 (±0.06)	13.200 (±13.35)	2.483 (±2.12)
		Sum	0.887 (±0.01)	20.743 (±4.20)	2.737 (±0.94)	0.903 (±0.03)	11.667 (±5.86)	1.923 (±0.76)
	3D	Each	0.947 (±0.01)	4.597 (±1.34)	1.093 (±0.39)	0.944 (±0.01)	4.673 (±1.51)	1.060 (±0.32)
		Sum	0.941 (±0.01)	5.837 (±4.52)	1.213 (±0.44)	0.940 (±0.01)	5.303 (±3.57)	1.123 (±0.39)

In the left breast, the 3D technique achieved higher DSC values and lower MSD than the 2D method, particularly for the PTV (DSC: 0.954 vs 0.940; MSD: 0.830 vs 1.097). Similar trends were observed for the CTV. In contrast, the right breast showed substantially higher HD95 and MSD values with the 2D method, especially in the PTV (HD95: 20.743 mm with 2D vs 5.837 mm with 3D), indicating poor boundary agreement.

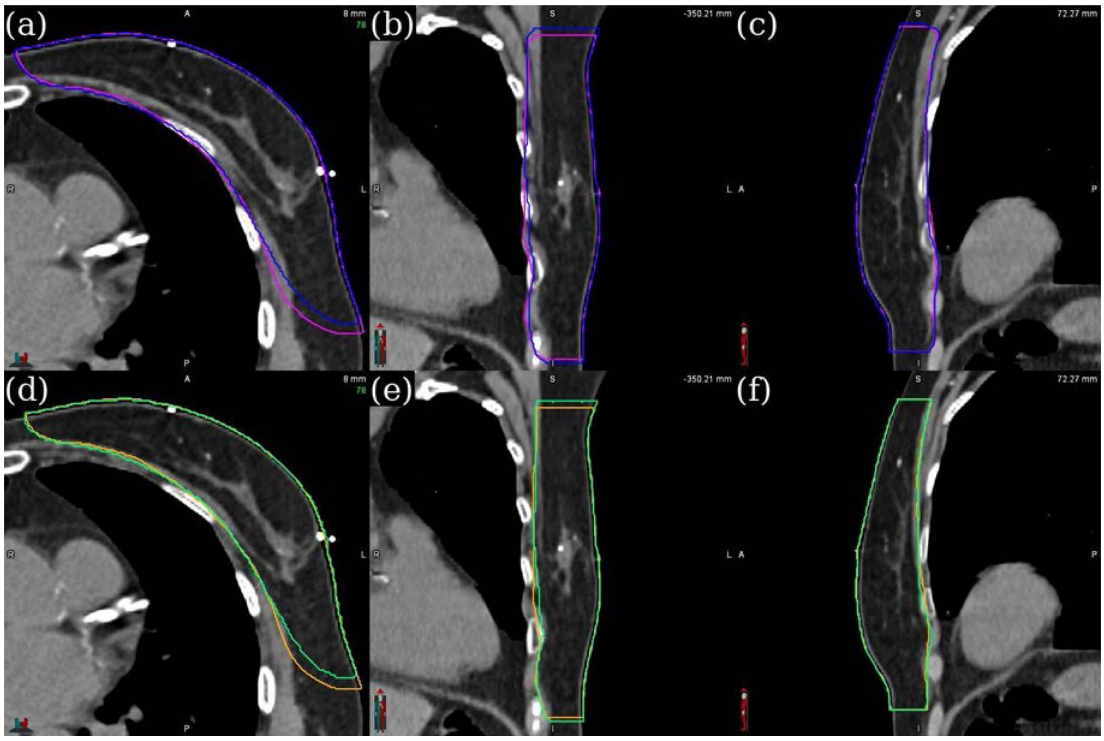

Fig. 3. (Color online) Results of 2D and 3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (OncoStudio) when the radiation treatment range was limited to the WBI.

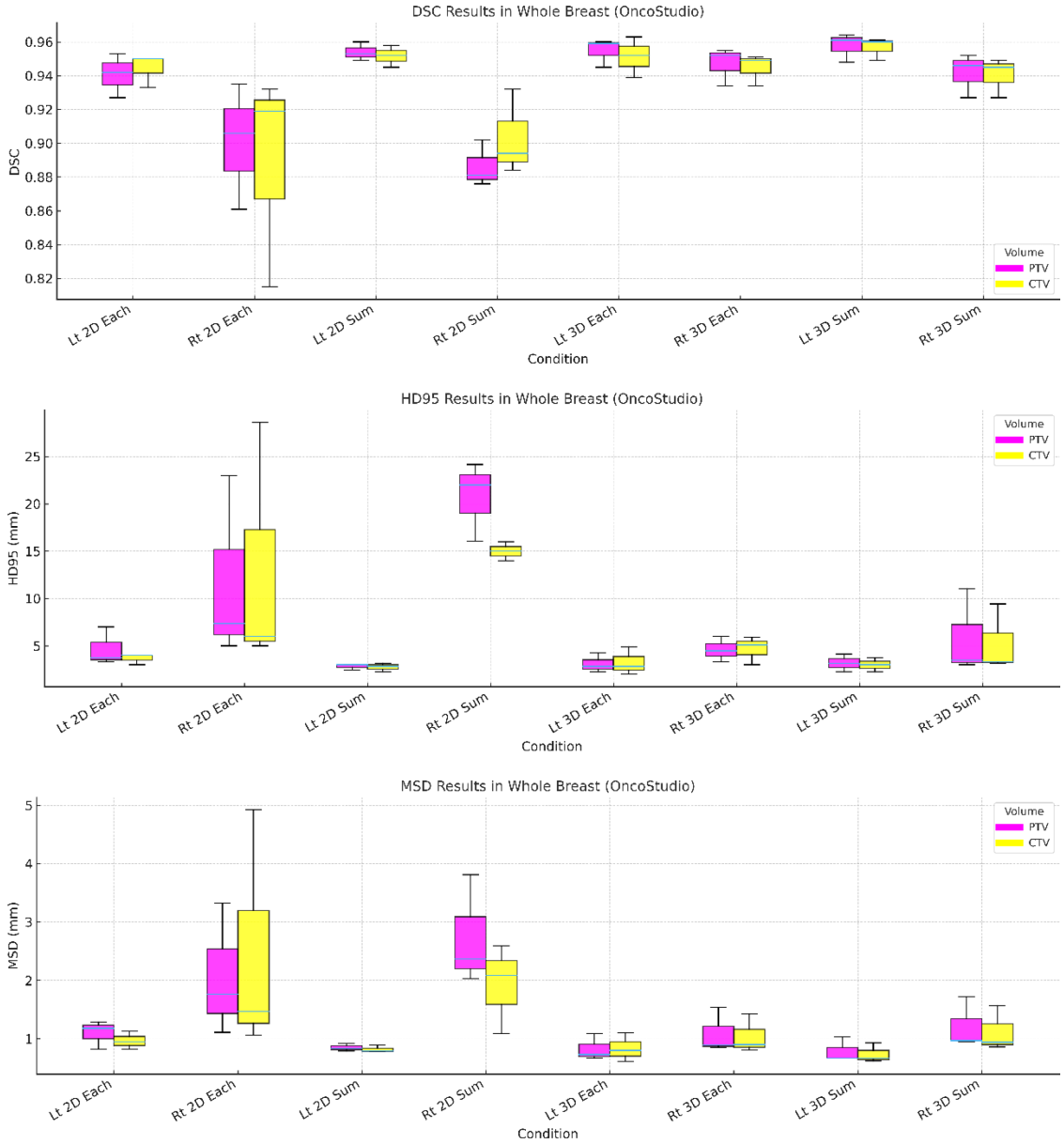

Fig. 4. (Color online) Results of DSC, HD95, and MSD based on 2D and 3D (Each/Sum) learning methods for manual and AI-based auto-segmentation (OncoStudio) when the radiation treatment range was limited to the WBI.

The 3D approach mitigated this issue and achieved better spatial conformity.

CTV structures generally exhibited higher DSC values and lower distance metrics than PTVs, suggesting greater contouring consistency. The large performance gap in the right breast with the 2D method implies greater anatomical complexity or variability, which was better managed using the 3D method.

In conclusion, even in the WBI setting, 3D segmentation demonstrated superior quantitative accuracy compared with 2D, offering improved delineation performance, particularly in anatomically complex regions. This finding may have important implications for clinical planning (Table 2, Fig. 3-4).

3.3. Comparison results of manual segmentation and AI-based automatic segmentation (Acculearning) based on 2D/3D (Each/Sum) learning methods when the radiation treatment range is extended to RNI

For WBI with RNI cases, the analysis showed that the 3D segmentation technique outperformed the 2D approach across all evaluation metrics. For WBI with RNI cases,

Table 3. DSC, HD95, and MSD results based on 2D/3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (Acculearning) when the radiation treatment range is extended to include the RNI.

WBI+RNI			PTV			CTV		
			DSC	HD95(mm)	MSD(mm)	DSC	HD95(mm)	MSD(mm)
Lt RNI	2D	Each	0.921	7.770	1.455	0.907	11.905	1.665
		Sum	0.918	7.450	1.520	0.914	5.925	1.400
	3D	Each	0.937	3.535	0.915	0.928	3.500	0.980
		Sum	0.935	3.735	0.950	0.928	3.680	0.990
Rt RNI	2D	Each	0.925	4.695	1.315	0.918	5.000	1.275
		Sum	0.925	4.550	1.345	0.915	5.500	1.330
	3D	Each	0.937	4.195	1.115	0.930	4.285	1.100
		Sum	0.939	3.845	1.080	0.934	3.950	1.035

the DSC was consistently higher with 3D for both the PTV and CTV on the left and right sides, whereas the HD95 and MSD were generally lower, indicating improved spatial alignment and boundary conformity.

For the left WBI and RNI PTVs, the 3D method achieved DSC of 0.937, HD95 of 3.535 mm, and MSD of 0.915 mm, compared to the 2D method at 0.921, 7.770, and 1.455 mm, respectively. This comparison shows a clear performance advantage. The CTV showed similar improvements, particularly in HD95, which decreased markedly from 11.905 mm (2D) to 3.500 mm (3D), suggesting that the 2D approach may be subject to considerable boundary errors.

The right WBI and RNI also demonstrated the consistent superiority of the 3D method. For instance, in the right CTV, HD95 decreased from 5.500 mm (2D) to 3.950 mm (3D), and MSD decreased from 1.330 mm to 1.035 mm, reflecting a more precise contour alignment. These findings confirm that 3D segmentation provides more accurate and robust delineation in WBI in RNI cases. This finding is particularly valuable in radiation therapy planning involving complex nodal structures, supporting the clinical utility of

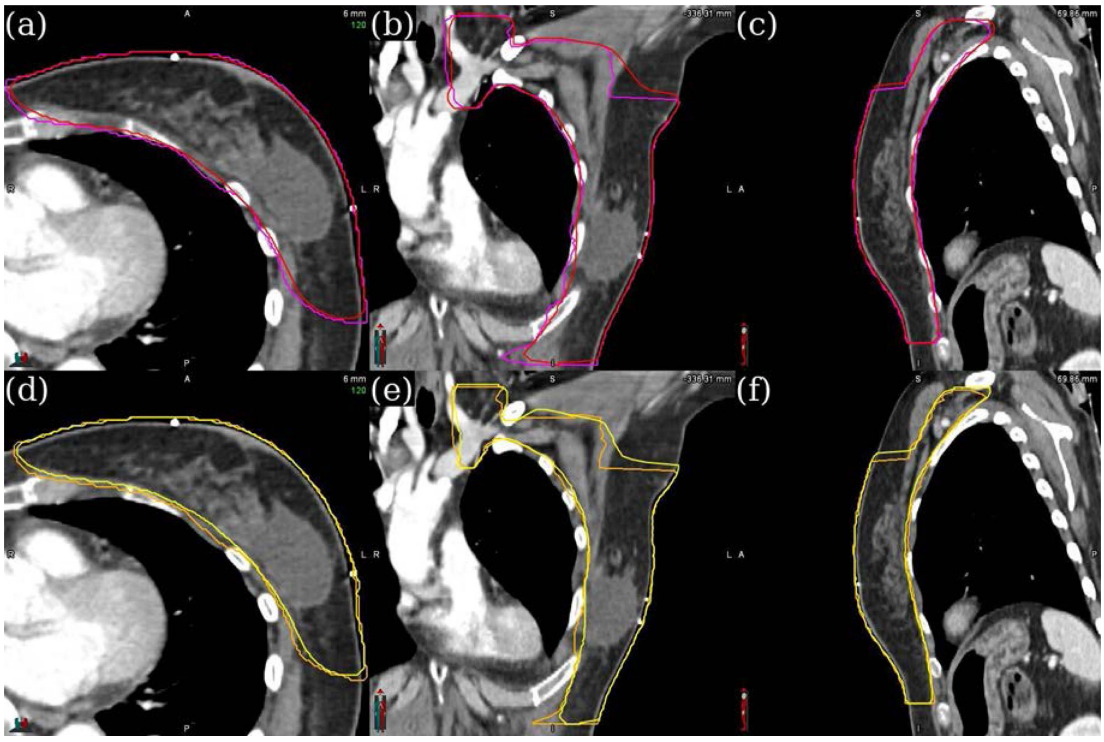

Fig. 5. (Color online) Results based on 2D and 3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (Acculearning) when the radiation treatment range is extended to include the RNI.

3D techniques in improving treatment precision for RNI (Table 3, Fig. 5).

3.4. Comparison results of manual segmentation and AI-based automatic segmentation (OncoStudio) based on 2D/3D (Each/Sum) learning methods when the radiation treatment range is extended to include the RNI

For WBI with RNI cases, the 3D segmentation technique demonstrated superior quantitative performance compared with the 2D approach. Across all comparisons, the DSC values were consistently higher for 3D, whereas the HD95 and MSD values were generally lower, reflecting more accurate and consistent boundary delineation. For the left WBI and RNI PTV, the 3D method achieved a DSC of 0.955, HD95 of 2.560 mm, and MSD of 0.730 mm, compared to the 2D method, which achieved 0.911, 5.680 mm, and 1.250 mm, respectively. These values show marked improvements in quantitative accuracy. In the CTV, 3D segmentation achieved better HD95 (3.415 mm vs 5.820 mm) and MSD (0.990 mm vs 1.365 mm), further confirming enhanced spatial conformity in complex anatomical regions.

The right WBI and RNI exhibited similar trends. For instance, in the CTV, the 3D approach achieved a DSC of

Table 4. DSC, HD95, and MSD results based on 2D/3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (OncoStudio) when the radiation treatment range is extended to include the RNI.

WBI+RNI			PTV			CTV		
			DSC	HD95(mm)	MSD(mm)	DSC	HD95(mm)	MSD(mm)
Lt RNI	2D	Each	0.911	5.680	1.250	0.901	5.820	1.365
		Sum	0.917	4.560	1.110	0.905	5.450	1.250
	3D	Each	0.955	2.560	0.730	0.928	3.415	0.990
		Sum	0.927	4.235	1.070	0.918	4.735	1.165
Rt RNI	2D	Each	0.927	5.180	1.315	0.928	3.735	1.125
		Sum	0.934	4.290	1.185	0.912	6.035	1.390
	3D	Each	0.941	3.870	1.040	0.929	4.000	1.095
		Sum	0.941	3.965	1.030	0.936	3.760	0.985

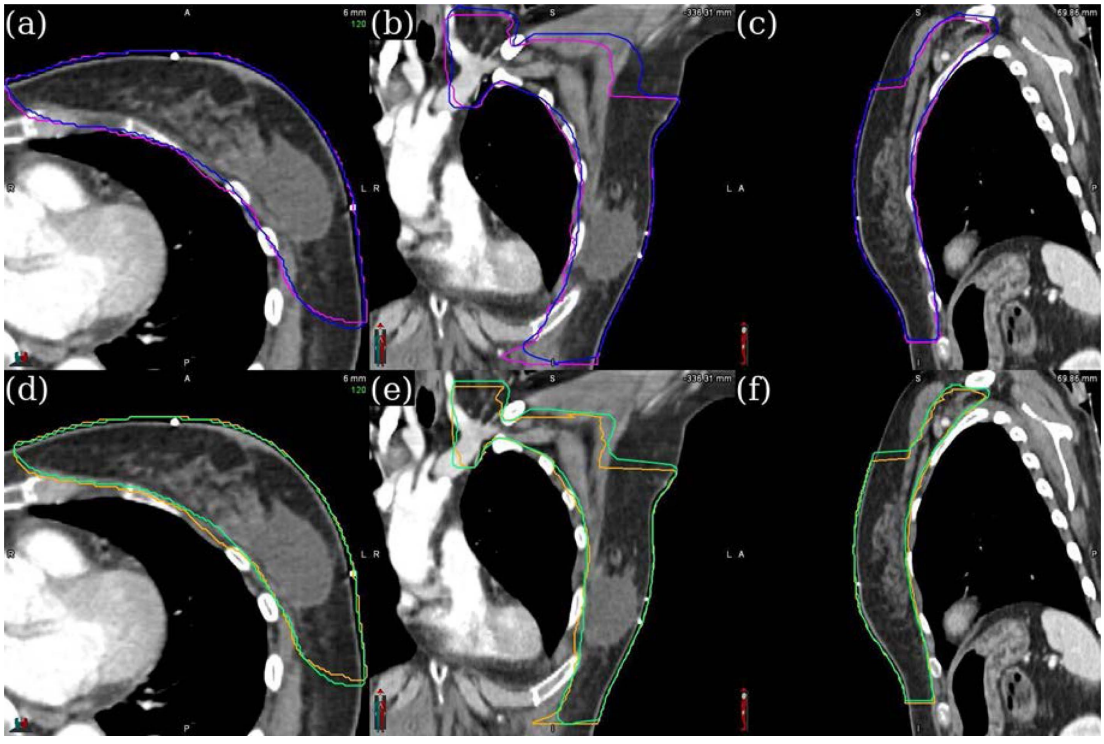

Fig. 6. (Color online) Results based on 2D and 3D (Each/Sum) learning methods for manual segmentation and AI-based auto-segmentation (OncoStudio) when the radiation treatment range is extended to include the RNI.

0.936, HD95 of 3.760 mm, and MSD of 0.985 mm, outperforming the 2D method (DSC: 0.912, HD95: 6.035 mm, MSD: 1.390 mm). These improvements in distance-based metrics underscore the advantages of the 3D method in reducing contouring errors and enhancing geometric precision.

In conclusion, this analysis reinforces that 3D segmentation achieves higher structural accuracy and boundary consistency in WBI for RNI cases, highlighting its strong clinical applicability in radiation therapy planning for nodal targets (Table 4, Fig. 6).

A paired *t*-test confirmed that 3D segmentation significantly outperformed the 2D method under RNI settings, with p-values < 0.05 across the geometric evaluation metrics (DSC, HD95, MSD). In the WBI cases, statistically significant differences were also observed between the left and right breast segmentations ($p < 0.05$), indicating that anatomical laterality contributed to segmentation variability.

3.5. Comparison results of target contouring times between manual contouring and AI-assisted contouring with manual correction

In this study, the average time required for manual contouring was 548 ± 205 s (9.1 min), whereas AI-assisted contouring, followed by manual revision, was significantly reduced to 187 ± 40 s (3.1 min), with AI-assisted contouring alone taking an average of 32 ± 7 s. This corresponded to an average time saving of approximately 361 s (6.0 min), representing a 66% reduction per patient. These findings highlight the clinical utility of AI-based auto-segmentation in reducing the workload of radiation oncologists and improving workflow efficiency (Table 5, Fig. 7).

Table 5. Comparison of target contouring times between manual contouring and AI-assisted contouring with manual correction.

Method	Mean ± SD (sec)
Manual contouring	548 ± 205
AI auto-segmentation + revision	187 ± 40

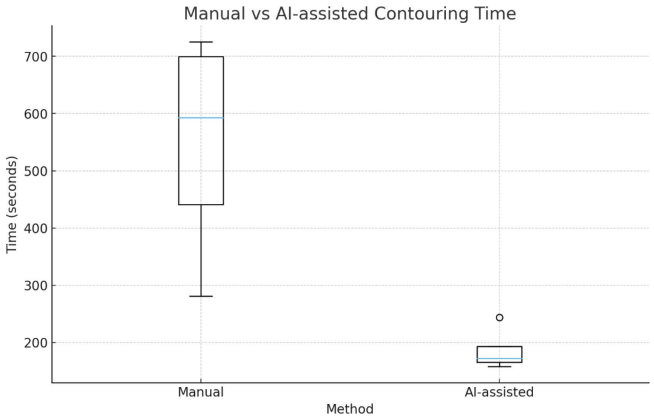

Fig. 7. (Color online) Boxplot comparison of contouring times between manual contouring and AI-assisted contouring with manual revision.

4. Discussion

This study demonstrated the clinical feasibility of AI-based auto-segmentation in radiation therapy planning for patients with breast cancer, particularly in addressing the limitations of manual contouring on sim-CT images. Both the Acculearning and OncoStudio models consistently achieved superior 3D segmentation performance compared to 2D across all evaluated metrics, including DSC, HD95, and MSD. These advantages were particularly evident in WBI cases with RNI, which involved anatomically complex target regions.

These statistical findings further reinforce that 3D segmentation consistently improves geometric accuracy and boundary conformity compared to 2D segmentation, particularly in complex nodal targets. Additionally, the significant laterality-dependent differences observed in WBI cases highlight the importance of accounting for anatomical variability during auto-segmentation in clinical practice.

The average DSC between the manual and AI-based contours was approximately 0.9, indicating a clinically acceptable level of agreement. Furthermore, the time required for manual correction after AI-based segmentation was significantly reduced, with manual-only segmentation averaging 548 ± 205 s (9.1 min) and AI-assisted segmentation requiring 187 ± 40 s (3.1 min), resulting in a 66% reduction. The AI-assisted contouring process also demonstrated high processing efficiency, with an average processing time of 32 ± 7 s, producing clinically applicable results in near-real time. However, the analysis of contouring time was based on only four representative cases, which limited the statistical power; therefore, the results were described using mean time reduction rather than statistical significance. These findings highlight the potential of AI tools to streamline clinical workflows and reduce the workload of radiation oncologists.

However, complete automation has not been achieved for all anatomical structures. In smaller or low-contrast regions, such as the optic nerve or rectum, segmentation accuracy was lower, with HD95 values occasionally being unstable or infinite [26]. These limitations likely

reflect the intrinsic resolution constraints of CT imaging, the limited diversity of the training dataset, and the irregular geometry of specific target volumes. However, the present study demonstrated that AI-based auto-segmentation can achieve satisfactory contouring performance even in anatomically less distinct regions, such as the CTV, suggesting its potential for broader clinical applicability. Future developments may benefit from integrating multimodal imaging approaches, such as magnetic resonance imaging or positron emission tomography, to improve both anatomical delineation and functional characterization.

To ensure generalizability, it is essential to validate AI models using multi-institutional datasets, as variations in contouring protocols and patient anatomy across institutions can substantially influence the performance of the models. While vendor-provided models performed adequately in specific anatomical regions, such as the chest, abdomen, and pelvis, transfer learning yielded notable improvements in the head and neck, underscoring the importance of fine-tuning anatomically complex and variable regions [27]. Although consensus guidelines exist for the CTV in WBI and RNI, substantial inter-institutional and inter-physician variability remains in clinical practice. Using our institutional dataset, this study demonstrated that AI-based training can generate auto-segmentation results optimized for institutional contouring practices, highlighting the importance of institution-specific model adaptation. Moreover, selecting appropriate validation tools tailored to the anatomical region of interest may be necessary to ensure robust, clinically relevant model evaluation.

Furthermore, our findings suggest that AI systems capable of adapting to patient-specific anatomical changes are critical for advancing adaptive radiation therapy protocols. This is attributable to our finding that AI-based auto-segmentation achieved accurate and time-efficient contouring, demonstrating its feasibility for integration into adaptive radiation therapy workflows. In our study, the entire contouring process, including manual revision after auto-segmentation, was completed in an average of 5 minutes. This efficiency suggests that, with further advances in automated planning systems, adaptive radiation therapy could become more time-efficient, less labor-intensive, and more precise in clinical practice.

Ultimately, fully automated workflows that integrate segmentation, dose planning, and adaptive adjustments have the potential to reduce inter-observer variability, enhance treatment precision, and personalize radiation therapy. However, achieving this vision will require concurrent advancements in algorithm interoperability, seamless integration with hospital PACS/TPS environments, and clinician-centered designs. Continuous clinical validation and multidisciplinary coordination among AI, data science, and radiation oncology experts are essential to ensure safe and effective implementation in clinical practice.

5. Conclusion

This study evaluated two AI-based auto-segmentation models for adjuvant radiation therapy planning in patients with breast cancer following BCS. Both models demonstrated high clinical utility, showing excellent agreement with manual contours for CTV and PTV, with mean DSCs exceeding 0.9 and consistently stable HD95 and MSD values. The average time required for manual correction after AI-based segmentation was reduced to 187 ± 40 s (3.1 min), compared with 548 ± 205 s (9.1 min) for manual segmentation, reflecting a substantial improvement in efficiency.

These findings support the role of AI-based auto-segmentation as a valuable adjunct for radiation oncologists, reducing manual workload and interobserver variability while laying the groundwork for adaptive radiation therapy. The superior performance of 3D segmentation, particularly in WBI with RNI cases involving complex anatomical structures, highlights its potential to improve anatomical accuracy and enhance clinical precision in treatment planning.

Statistical comparisons confirmed that 3D segmentation significantly outperformed 2D in RNI cases ($p < 0.05$) and demonstrated significant laterality-dependent variability in WBI cases ($p < 0.05$). Therefore, incorporating anatomical complexity and laterality into auto-segmentation workflows may further optimize radiation therapy planning and improve its clinical accuracy.

References

[1] H. Baroudi, K. K. Brock, W. Cao, X. Chen, C. Chung, L. E. Court, M. D. El Basha, M. Farhat, S. Gay, M. P. Gronberg, A. C. Gupta, S. Hernandez, K. Huang, D. A. Jaffray, R. Lim, B. Marquez, K. Nealon, T. J. Netherton, C. M. Nguyen, B. Reber, D. J. Rhee, R. M. Salazar, M. D. Shanker, C. Sjogreen, M. Woodland, J. Yang, C. Yu, and Y. Zhao, Diagnostics (Basel) **13**, 667 (2023).

[2] O. M. Dona Lemus, M. Cao, B. Cai, M. Cummings, and D. Zheng, Cancers (Basel) **16**, 1206 (2024).

[3] L. Joskowicz, A. Argenone, G. I. Boboc, F. Cucciarelli, F. De Rose, and M. C. De Santis. Eur. Radiol. **29**, 1391 (2019).

[4] G. Lappas, N. Staut, N. G. Lieuwes, R. Biemans, C. J. Wolfs, S. J. van Hoof, and F. Verhaegen, Phys. Imaging

Radiat. Oncol. **21**, 11 (2022).
[5] C. E. Cardenas, S. E. Blinde, A. S. R. Mohamed, S. P. Ng, C. Raaijmakers, M. Philippens, A. Kotte, A. A. Al-Mamgani, I. Karam, D. J. Thomson, J. Robbins, K. Newbold, C. D. Fuller, and C. Terhaard, Int. J. Radiat. Oncol. Biol. Phys. **113**, 426 (2022).
[6] Y. B. Fu, H. Zhang, E. D. Morris, C. K. Glide-Hurst, S. Pai, A. Traverso, L. Wee, I. Hadzic, P. I. Lønne, C. Y. Shen, T. Liu, and X. F. Yang, IEEE Trans. Radiat. Plasma Med. Sci. **6**, 158 (2022).
[7] C. Jeong, Y. Goh, and J. Kwak, Ewha Med. J. **47**, e49 (2024).
[8] R. Krishnamurthy, N. Mummudi, J. S. Goda, S. Chopra, B. Heijmen, and J. Swamidas, JCO Glob. Oncol. **8**, e2100393 (2022).
[9] F. Kouhen, H. E. Gouach, K. Saidi, Z. Dahbi, N. Errafiy, H. Elmarrachi, and N. Ismaili, Gulf J. Oncol. **1**, 94 (2024).
[10] M. Kawamura, T. Kamomae, M. Yanagawa, K. Kamagata, S. Fujita, D. Ueda, Y. Matsui, Y. Fushimi, T. Fujioka, T. Nozaki, A. Yamada, K. Hirata, R. Ito, N. Fujima, F. Tatsugami, T. Nakaura, T. Tsuboyama, and S. Naganawa, J. Radiat. Res. **65**, 1 (2024).
[11] J. Wong, V. Huang, D. Wells, J. Giambattista, C. Kolbeck, K. Otto, E. P. Saibishkumar, and A. Alexander, Radiat. Oncol. **16**, 101 (2021).
[12] A. C. Erdur, D. Rusche, D. Scholz, J. Kiechle, S. Fischer, Ó. Llorián-Salvador, J. A. Buchner, M. Q. Nguyen, L. Etzel, J. Weidner, M.-C. Metz, B. Wiestler, J. Schnabel, D. Rueckert, S. E. Combs, and J. C. Peeken Strahlenther. Onkol. **201**, 236 (2024).
[13] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, Nat. Methods **18**, 203 (2021).
[14] P. J. Doolan, S. Charalambous, Y. Roussakis, A. Leczynski, M. Peratikou, M. Benjamin, K. Ferentinos, I. Strouthos, C. Zamboglou, and E. Karagiannis, Front. Oncol. **13**, 1213068 (2023).
[15] S. M. H. Hoque, G. Pirrone, F. Matrone, A. Donofrio, G. Fanetti, A. Caroli, R. S. Rista, R. Bortolus, M. Avanzo, A. Drigo, and P. Chiovati, Cancers (Basel) **15**, 5735 (2023).
[16] F. Shi, W. Hu, J. Wu, M. Han, J. Wang, W. Zhang, Q. Zhou, J. Zhou, Y. Wei, Y. Shao, Y. Yu, X. Cao, Y. Zhan, X. S. Zhou, Y. Gao, and D. Shen, Nat. Commun. **13**, 6566 (2022).
[17] Z. Smine, S. Poeta, A. De Caluwé, A. Desmet, C. Garibaldi, K. Brou Boni, H. Levillain, D. Van Gestel, N. Reynaert, and J. Dhont, Radiother. Oncol. **202**, 110615 (2025).
[18] S. Warren, N. Richmond, A. Wowk, M. Wilkinson, and K. Wright, IPEM Transl. **6-8**, 100020 (2023).
[19] B. V. Ofersen, L. J. Boersma, C. Kirkove, S. Hol, M. C. Aznar, A. B. Sola, J. Mauritsen, and P. Schultheiss, Radiother. Oncol. **114**, 3 (2015).
[20] M. S. Gentile, A. A. Usman, E. I. Neuschler, V. Sathiaseelan, J. P. Hayes, and W. Small Jr., Int. J. Radiat. Oncol. Biol. Phys. **93**, 257 (2015).
[21] K. Men, T. Zhang, X. Chen, B. Chen, Y. Tang, S. Wang, Y. Li, and J. Dai, Phys. Med. **50**, 13 (2018).
[22] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Zeng, A. Weber, O. Székely, M. Macyszyn, W. Pedano, D. M. Prince, E. S. Haar Romeny, M. K. van Leemput, M. J. A. M. van Ginneken, and M. Reyes, Med. Image Anal. **35**, 18 (2017).
[23] T. A. Sorensen, Biol. Skar. **5**, 1 (1948).
[24] W. Groß, Monatsh. Math. Phys. **26**, A34 (1915).
[25] T. Heimann, B. van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. M. M. Cashman, Y. Chi, A. Córdova, B. M. Dawant, M. Fidrich, J. D. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmüller, R. I. Kitney, H. Kobatake, H. Lamecker, T. Lange, J. Lee, B. Lennon, R. Li, S. Li, H.-P. Meinzer, G. Németh, D. S. Raicu, A.-M. Rau, E. M. van Rikxoort, M. Rousson, L. Rusko, K. A. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. M. Waite, A. Wimmer, and I. Wolf, IEEE Trans. Med. Imaging **28**, 1251 (2009).
[26] W. Wang, Q. Wang, M. Jia, Z. Wang, C. Yang, D. Zhang, S. Wen, D. Hou, N. Liu, P. Wang, and J. Wang, Front. Phys. **9**, 743190 (2021).
[27] B. Clark, N. Hardcastle, L. A. Johnston, and J. Korte, Med. Phys. **51**, 4767 (2024).