Hong Kong J Radiol. 2025;28(1):e23-34. Published 24 March 2025 | https://doi.org/10.12809/hkjr2417871

ORIGINAL ARTICLE

Artificial Intelligence for Contouring and Treatment Planning in Locoregional Radiotherapy Including Internal Mammary Nodal Irradiation for Breast Cancer

CCC Wong, YL Wong, KF Choi, SCY Ng, ASY Chan, MY Luk, KK Yuen

Department of Clinical Oncology, Queen Mary Hospital, Hong Kong SAR, China

ABSTRACT

Introduction: *Postoperative internal mammary node (IMN) irradiation in high-risk early-stage breast cancer is becoming a standard procedure. We assessed the performance of artificial intelligence (AI) in contouring and treatment planning in locoregional radiotherapy including IMN for breast cancer.*

Methods: *We included 27 patients who underwent postoperative locoregional radiotherapy. The study consisted of two parts. For the first part, AI-generated segmentations of target volumes and organs at risk (OARs) were compared to manual segmentations using four metrics, namely, volume ratio (VR), the 95th percentile Hausdorff distance (95% HD), surface Dice similarity coefficient (sDSC), and volume Dice similarity coefficient (vDSC). The time needed for auto-segmentation was recorded. For the second part, AI-generated volumetric modulated arc therapy (VMAT) and intensity modulated radiation therapy (IMRT) treatment plans using machine learning (ML) were compared to manually created plans, focusing on dosimetric parameters and planning times.*

Results: *AI demonstrated excellent performance in segmenting OARs, with high sDSC and vDSC, low 95% HD and a VR of 1.01 ± 0.03. However, performance for breast and nodal volumes was less robust. The time for auto-segmentation averaged 27.39 ± 2.53 seconds. In treatment planning, both ML and manual plans achieved satisfactory target volumes coverage and OAR constraints, except that IMRT (ML) plans did not meet the dose limit for the ipsilateral lung. VMAT (ML) planning took less time than manual planning, while IMRT (ML) planning took longer than manual planning.*

Conclusion: *AI is proficient in segmenting OARs but is less reliable for target volumes. It produces VMAT plans comparable to manual plans in a shorter timeframe.*

Key Words: *Artificial intelligence; Breast neoplasms; Organs at risk; Radiotherapy, intensity-modulated*

Correspondence: *Dr CCC Wong, Department of Clinical Oncology, Queen Mary Hospital, Hong Kong SAR, China*
Email: wcc742@ha.org.hk

Submitted: 10 July 2024; Accepted: 14 November 2024.

Contributors: All authors designed the study. CCCW, YLW, KFC, SCYN, ASYC and MYL acquired the data. CCCW and YLW analysed the data. CCCW drafted the manuscript and critically revised the manuscript for important intellectual content. All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of Interest: As an editor of the journal, CCCW was not involved in the peer review process. Other authors have disclosed no conflicts of interest.

Funding/Support: This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability: All data generated or analysed during the present study are available from the corresponding author on reasonable request.

Ethics Approval: This research was approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster, Hong Kong (Ref No.: UW 24-221). The requirement for informed patient consent was waived by the Committee due to the retrospective nature of the research.

中文摘要

人工智能在乳癌包含內乳淋巴區的局部放射治療的輪廓勾畫及治療規劃之應用

黃卓卓、黃逸霖、蔡國鋒、吳楚儀、陳心妍、陸美儀、袁國強

引言：在高風險早期乳癌進行術後內乳淋巴區放射治療逐漸成為標準程序。我們評估人工智能在乳癌包含內乳淋巴區的局部放射治療的輪廓勾畫及治療規劃方面之表現。

方法：我們納入27名接受了術後局部放射治療的患者。本研究分為兩部分，在第一部分，我們使用以下四個指標比較人工智能產生及人手制訂的靶區體積及危及器官勾畫：（1）體積比；（2）95% Hausdorff距離；（3）表面Dice相似系數；及（4）體積Dice相似系數。我們並紀錄了經人工智能勾畫所需的時間。至於第二部分，我們比較經機器學習人工智能產生的體積調控弧型放射治療（VMAT）及強度調控放射治療（IMRT）計劃與人手制訂計劃，重點為劑量測定參數及規劃時間。

結果：人工智能在勾畫危及器官方面表現優秀，表面Dice相似系數及體積Dice相似系數高，95% Hausdorff距離小，體積比為1.01 ± 0.03，但在乳房及淋巴結體積的表現較不穩定。經人工智能勾畫的時間平均為27.39 ± 2.53秒。在治療規劃，機器學習及人手制訂計劃均達至滿意的靶區體積覆蓋及危及器官劑量限制，唯機器學習IMRT計劃未能符合同側肺的劑量限制。機器學習VMAT計劃需要的時間較人手制訂計劃短，機器學習IMRT計劃需要的時間則較人手制訂計劃長。

結論：人工智能擅長勾畫危及器官，但在勾畫靶區體積方面則較遜色。它能在較短時間內製作可媲美人手制訂的VMAT計劃。

INTRODUCTION

Breast cancer is the most prevalent cancer affecting women worldwide. According to the World Health Organization, in 2022, there were approximately 2.3 million new breast cancer diagnoses made globally, and 670,000 deaths.[1] In Hong Kong, approximately 5500 new breast cancer cases were reported in 2021, with a crude incidence rate of 138.1 per 100,000 female population.[2]

The treatment of breast cancer typically involves a multimodal approach, including surgery, radiotherapy, and systemic therapies. Adjuvant radiotherapy plays a crucial role in reducing the risk of cancer recurrence and improving survival, particularly for patients who have undergone breast conserving surgery or have been diagnosed with locally advanced breast cancer.

The Danish Breast Cancer Group Internal Mammary Node study in 2022[3] showed that elective irradiation of the internal mammary nodes (IMNs) reduced the risk of distant recurrence and improved long-term survival of patients with node-positive early breast cancer, regardless of breast cancer laterality. However, the irradiation of IMNs has remained a controversial topic, and guidelines for IMN irradiation vary significantly across different institutes. One controversial issue is the potential for long-term cardiotoxicity,[4] as IMN irradiation tends to result in a higher cardiac dose.[5] The technical complexities involved in radiotherapy treatment can be challenging, leading to an increase in the time and resources required to treat each patient.

Over the past decade, there has been rapid advancement in artificial intelligence (AI) and deep learning technology, which have been increasingly utilised in radiation oncology.[6] The AI-powered tools can perform segmentation, treatment planning, and delivery,[7] with the aim of improving the quality and consistency of the treatment delivered, and enhancing the efficiency of the work of oncologists and radiation therapists. Ultimately, the incorporation of AI in radiation oncology seeks to streamline hospital operations, thus benefiting a greater number of patients within the demanding clinical environment. In the past few years, the use of deep learning and convolutional neural networks on segmentation as well as treatment planning in breast

cancer radiotherapy has been explored. These studies have shown promising results, with the automated techniques demonstrating comparable quality to manual approaches.[8-12]

In this article, we sought to compare the performance of AI to that of manual approaches, in the context of segmentation and treatment planning in radiotherapy of IMNs in breast cancer.

METHODS

Patient Enrolment

The study included the data from 27 breast cancer patients, comprising 14 patients with left-sided breast cancer and 13 with right-sided breast cancer. All patients had undergone either mastectomy or breast conserving surgery (including lumpectomy and cryotherapy), as well as axillary surgery (including sentinel lymph node biopsy and axillary dissection). Five of the mastectomy patients also had undergone breast reconstruction. All 27 patients had received postoperative locoregional radiotherapy including irradiation of clinically evident IMN (i.e., lymph nodes that are suspected to be involved by malignancy on imaging) at the dose of 40 Gy in 15 fractions over 3 weeks between 2017 and 2023. A tumour bed electron boost was administered at a dose of 10 Gy in five fractions over 1 week to patients who had undergone breast conserving surgery. Patient characteristics (including cancer staging[13]) are presented in Table 1.

Table 1. Patient characteristics (n = 27).*

Mean age (range), y	55.7 (41-86)
Laterality	
Left breast cancer	14 (51.9%)
Right breast cancer	13 (48.1%)
Tumour (T) staging[13] upon diagnosis	
T1	1 (3.7%)
T2	16 (59.3%)
T3	4 (14.8%)
T4	4 (14.8%)
Not applicable†	1 (3.7%)
Unknown	1 (3.7%)
Nodal (N) staging[13] upon diagnosis	
N2	4 (14.8%)
N3	23 (85.2%)
Overall staging[13] upon diagnosis	
Stage III	20 (74.1%)
Stage IV	7 (25.9%)
Neoadjuvant systemic therapy	
Yes	21 (77.8%)
No	6 (22.2%)
Type of breast surgery	
Mastectomy	
Without reconstruction	17 (63.0%)
With reconstruction	5 (18.5%)
Breast conserving surgery	
Lumpectomy	4 (14.8%)
Cryotherapy	1 (3.7%)
Type of axillary surgery	
Axillary dissection	22 (81.5%)
Sentinel lymph node biopsy	5 (18.5%)

* Data are shown as No. (%), unless otherwise specified.
† One patient had solitary recurrence in an internal mammary node.

Image Acquisition

All patients were positioned in the supine orientation on a breast board (Breastboard MT-350N; CIVCO, Orange City [FL], US), with the arms abducted and holding onto a central rod. Simulation was performed using computed tomography (CT), with a slice separation of 2.5 mm, covering the region from the jaw to the mid-abdomen. For patients who had undergone mastectomy without reconstruction, a 0.5-cm skin bolus was administered to the chest wall in the planning process.

The CT data were retrieved and exported to treatment planning software Eclipse[14] (version 15.6; Varian Medical Systems, Palo Alto [CA], US) and RayStation 12A deep learning segmentation software[15] (RaySearch Laboratories, Stockholm, Sweden) to perform contouring and treatment planning.

Contouring

This study comprises two parts. In the first part of the study, two sets of contours were generated from the patients' CT data. The performance of auto-segmentation was assessed by comparing its results to the manual contouring done by the oncologists.

Manual Contours

Contouring of target volume and organs at risk (OARs) were done on Eclipse with reference to the European Society for Radiotherapy and Oncology consensus guidelines.[16] They were reviewed and approved by two senior clinical oncologists, each having >10 years of experience in breast cancer.

Auto-contours or Auto-segmentation

Automatic contouring of target volume and OARs were created using the deep learning segmentation (RSL Breast CT_1) planning model within RayStation 12A. A 5-mm expansion was applied to the breast or chest wall, axillary levels 3 to 4, and the IMNs from manual contours to become the planning target volume (PTV). The PTV was then trimmed by 5 mm underneath the

skin, except in slim patients where the thoracic chest wall was too thin, causing the target volume to disappear when cropped. The time taken for generation of auto-segmentation was recorded.

The list and number of target volumes that were created are presented in Table 2.

Radiotherapy Treatment Planning

In the second part of the study, four virtual treatment plans were created. These included two plans created with machine learning (ML): a volumetric modulated arc therapy ML [VMAT (ML)] plan and an intensity modulated radiation therapy ML [IMRT (ML)] plan using a dynamic multileaf collimator with the ML model RSL-Breast-L-nodes-4005(1.0). Manual VMAT [VMAT (manual)] and manual IMRT [IMRT (manual)] plans were generated by the usual manual treatment planning methods.

Table 2. Number of target volumes created.*

	Auto-contouring	Manual contouring
Axillary level 1		
Left (LN_Ax_L1_L)	13	14
Right (LN_Ax_L1_R)	12	13
Axillary level 2		
Left (LN_Ax_L2_L)	14	14
Right (LN_Ax_L2_R)	13	13
Axillary level 3		
Left (LN_Ax_L3_L)	14	14
Right (LN_Ax_L3_R)	13	13
Axillary level 4		
Left (LN_Ax_L4_L)	14	14
Right (LN_Ax_L4_R)	13	13
Interpectoral nodes		
Left (LN_Ax_Pectoral_L)	14	14
Right (LN_Ax_Pectoral_R)	13	13
Internal mammary nodes		
Left (LN_IMC_L)	13	14
Right (LN_IMC_R)	10	13
Breast		
Left (Breast_L)	2	2
Right (Breast_R)	3	3
Chest wall		
Left (Chestwall_L)	N/A	12
Right (Chestwall_R)	N/A	10
Lungs		
Left (Lung_L)	27	27
Right (Lung_R)	27	27
Heart	27	27

Abbreviation: N/A = not applicable.
* Items in parentheses are the nomenclature of the target region on the treatment planning system.

The dosimetric objectives, constraints for target volumes and OARs used for treatment planning dose optimisation are presented in Table 3, in accordance with departmental consensus, The Royal College of Radiologists' UK Consensus Statements,[17] and the Danish Breast Cancer Cooperative Group HYPO II protocol.[18] We recorded the time needed for automated and manual treatment planning.

Parameters for Comparison

Contouring

The manually contoured and auto-segmented volumes were compared using the following four geometrical metrics:

1. Volume ratio (VR)[19]: This metric evaluates the ratio of the volumes calculated by the test contour (auto-segmentation) and reference contour (manual contouring). A VR close to 1 indicates good agreement between the two contours.

2. The 95th percentile Hausdorff distance (95% HD)[19,20]: HD is a surface-based metric of the maximum perpendicular distance between two sets of points (in mm). It is used to quantify the differences between the field edges by manual and auto-contouring. The purpose of using the 95% HD instead of maximum HD is to eliminate the impact of small outliers.

3. Surface Dice similarity coefficient (sDSC)[19,21,22]: sDSC is a metric of spatial overlap between two volumes as a measure of estimated editing. It specifically looks

Table 3. Dosimetric objectives and constraints of target volumes and organs at risk for radiotherapy treatment planning.

		Objective	Constraint
Whole breast/ chest wall	$V_{95\%}$ † of 40 Gy	≥95%	≥90%
	$V_{105\%}$ † of 40 Gy	≤5%	≤7%
	$V_{107\%}$ † of 40 Gy	N/A	≤2%
Nodal volumes*	Maximum dose ($D_{0.5cc}$)	N/A	≤44 Gy
	$V_{90\%}$ † of 40 Gy	≥90%	N/A
	Maximum dose ($D_{0.5cc}$)	N/A	≤44 Gy
Ipsilateral lung	Mean dose	≤13 Gy	N/A
	Volume receiving 17 Gy	N/A	≤35 Gy
Contralateral lung	Mean dose	≤4 Gy	N/A
Heart	Mean dose	≤4 Gy	≤6 Gy
	Volume receiving 17 Gy	N/A	≤10%
	Mean dose	<1.5 Gy	≤3.5 Gy

Abbreviation: N/A = not applicable.
* Included axillary levels 3 and 4 and internal mammary nodes.
† Volume (%) of the planning target volume receiving at least the specified percentage of prescription dose.

at the overlap between the surfaces of the reference and test contours, providing an assessment of the contouring agreement. It has been shown to have a good correlation with the time required to edit contours.[23,24]

4. Volume Dice similarity coefficient (vDSC)[19,22]: vDSC evaluates the overall volumetric overlap between the reference and test contours. This gives an indication of how well the total volumes agree. However, it is not sensitive to complex boundaries and tends to give higher scores for larger volumes. It is generally accepted that volume DSC >0.7 indicates excellent agreement.[25,26]

Radiotherapy Treatment Planning

Dose-volume histograms (DVHs) were generated in each treatment plan for review of dosimetric parameters, including the PTV of the whole breast or chest wall ($V_{95\%}$ [i.e., volume (%) of the PTV receiving at least 95% of the prescription dose] of 40 Gy, $V_{105\%}$ of 40 Gy, $V_{107\%}$ of 40 Gy, and maximum dose [$D_{0.5cc}$]), the PTV of the nodal volumes ($V_{90\%}$ of 40 Gy and maximum dose [$D_{0.5cc}$]), the mean dose to the ipsilateral lung, the volume receiving 17 Gy in the ipsilateral lung, the mean dose to the contralateral lung, the mean dose to the heart, the volume receiving 17 Gy in the heart, and the mean dose to the contralateral breast.

Statistical Analyses

Descriptive analyses were stated as mean ± standard deviation; descriptive tables were created for the comparison of the contouring performance by AI. The dosimetric parameters for the target volumes and OARs in different treatment plans were compared using the Wilcoxon signed rank test for statistical significance. A p value < 0.05 was considered statistically significant.

RESULTS

Contouring

AI successfully performed auto-segmentation of parts of the target volumes including axillary level 2, axillary level 3 (infraclavicular volume), axillary level 4 (supraclavicular volume), interpectoral nodes, breast (in breast conserving surgery patients), as well as OARs including the heart and the lungs for all patients. Nevertheless, it failed in identifying axillary level 1 in two patients (one on the left side and one on the right side) and IMNs in four patients (one on the left side and three on the right side). AI misclassified the region of the chest wall in mastectomy patients as breast due to the absence of a specific chest wall segmentation function on RayStation 12A (Table 4). Thus, the result of comparison between manual and auto-contouring of the chest wall was disregarded. Table 2 shows the number of structures created by auto- and manual contouring, respectively.

Table 4 shows the results of the VR, 95% HD, sDSC, and vDSC for all structures.

Table 4. Results of comparison of manual and auto-contouring.*

	Left/right	No.	VR	95% HD, mm	sDSC	vDSC
Axillary level 1	Left	13	0.75 ± 0.22	19.29 ± 12.11	0.25 ± 0.16	0.68 ± 0.14
	Right	12	0.67 ± 0.30	19.89 ± 10.04	0.19 ± 0.10	0.61 ± 0.18
Axillary level 2	Left	14	0.90 ± 0.28	8.59 ± 3.88	0.51 ± 0.22	0.72 ± 0.16
	Right	13	0.77 ± 0.18	7.96 ± 4.01	0.61 ± 0.17	0.78 ± 0.14
Axillary level 3	Left	14	0.89 ± 0.33	8.59 ± 3.62	0.31 ± 0.14	0.68 ± 0.16
	Right	13	0.85 ± 0.16	6.74 ± 2.73	0.42 ± 0.15	0.76 ± 0.08
Axillary level 4	Left	14	0.62 ± 0.26	10.52 ± 6.34	0.31 ± 0.16	0.63 ± 0.19
	Right	13	0.56 ± 0.18	9.68 ± 6.39	0.32 ± 0.15	0.62 ± 0.17
Interpectoral nodes	Left	14	0.92 ± 0.20	7.80 ± 11.10	0.71 ± 0.32	0.78 ± 0.28
	Right	13	0.81 ± 0.16	6.82 ± 4.46	0.72 ± 0.25	0.78 ± 0.21
IMNs	Left	13	0.63 ± 0.21	15.62 ± 16.94	0.48 ± 0.20	0.66 ± 0.16
	Right	10	0.76 ± 0.15	13.37 ± 12.26	0.71 ± 0.10	0.81 ± 0.08
Breast	Left	2	0.84 ± 0.02	21.00 ± 0.99	0.61 ± 0.04	0.78 ± 0.08
	Right	3	0.92 ± 0.03	9.95 ± 1.53	0.90 ± 0.03	0.93 ± 0.03
Lung	Left	27	1.04 ± 0.03	3.03 ± 1.97	0.96 ± 0.02	0.98 ± 0.01
	Right	27	1.03 ± 0.03	3.54 ± 2.91	0.95 ± 0.03	0.98 ± 0.01
Heart	N/A	27	0.97 ± 0.06	5.38 ± 2.75	0.86 ± 0.08	0.94 ± 0.03

Abbreviations: 95% HD = 95th percentile Hausdorff distance; IMN = internal mammary node; N/A = not applicable; sDSC = surface Dice similarity coefficient; vDSC = volume Dice similarity coefficient; VR = volume ratio.

* Data are shown as mean ± standard deviation, unless otherwise specified.

Overall, AI demonstrated excellent performance in delineating the lungs and heart across all patients. In contrast, the performance for the target structures, i.e., nodal volumes and the breast, was less robust compared to the OARs, based on the four evaluation metrics used.

Figure 1 shows the manual contouring and auto-segmentation of the axillary nodal volume and breast on selected CT slices. When comparing the contouring of the nodal volumes, looking at VRs, the AI-generated structures tended to have smaller volumes compared to the manual contours, with VRs ranging from 0.56 ± 0.18 to 0.92 ± 0.03. This discrepancy was more pronounced for contouring of axillary level 4, axillary level 1, and IMNs (Table 4).

The largest 95% HD measurements were observed at axillary level 1, reaching up to 19.29 ± 12.11 mm for the left axillary level 1 and 19.89 ± 10.04 mm for the right axillary level 1, and left breast with the result of 21.00 ± 0.99 mm. This was followed by the IMNs, with measurements of 15.62 ± 16.94 mm for the left side and 13.37 ± 12.26 mm for the right side. The 95% HD measurements for the remaining structures ranged from 6.74 ± 2.73 mm to 10.52 ± 6.34 mm (Table 4).

High sDSC values were observed for interpectoral nodes, axillary level 2 and the breast. The lowest agreement was found for axillary level 1. The best results for vDSC were again observed in the interpectoral nodes, axillary level 2 and the breast. Conversely, both axillary level 1 and axillary level 4 had the lowest vDSC values (Table 4).

Results of nodal volumes in patients who underwent axillary dissection and sentinel lymph node biopsy (SLNB) are shown in Tables 5 and 6, respectively. In patients who underwent SLNB instead of axillary dissection, there was a tendency towards better performance of auto-segmentation, as indicated by higher sDSC and vDSC, in comparison to patients who underwent axillary dissection. However, it is important to note that a limitation of this analysis is the small sample size of patients who underwent SLNB.

The time required for auto-segmentation was recorded, showing an average of 27.39 ± 2.53 seconds for generation of one contouring.

Treatment Planning

DVH parameters of VMAT (ML) and VMAT (manual) are presented in Table 7, while DVH parameters of IMRT (ML) and IMRT (manual) are shown in Table 8. Figure 2 illustrates the dose distribution comparisons between VMAT (ML), VMAT (manual), IMRT (ML) and IMRT (manual) on selected slices.

All of the four virtual plans achieved satisfactory performance in covering the target volumes, including the breast or chest wall and nodal volumes, by fulfilling the dose constraints. For OARs, IMRT (ML) failed to achieve the dose limit for mean dose to the ipsilateral lung, while the other parameters were fulfilled in all of the plans.

Volumetric Modulated Arc Therapy Created with Machine Learning Versus Created Manually

Both VMAT (ML) plans and VMAT (manual) plans demonstrated clinically acceptable treatment plans. VMAT (manual) showed a significantly better performance for target coverage of the breast or chest wall and the nodal volumes [$p < 0.001$ for both PTV (IMN) and PTV (axillary levels 3 and 4); $p = 0.034$ for PTV (whole breast/chest wall)]. VMAT (ML) plans had

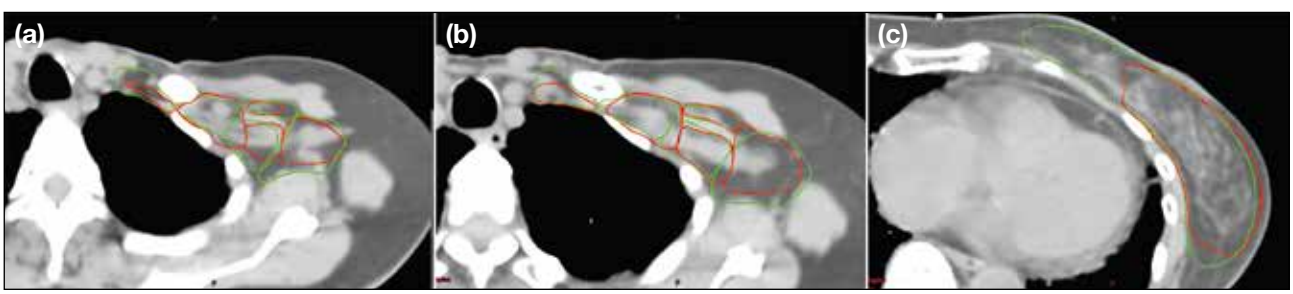

Figure 1. Selected cuts of the radiotherapy plans comparing auto-segmentation (red contours) and manual contouring (green contours) of target volumes on a patient following left lumpectomy and sentinel lymph node biopsy.

Table 5. Results of comparison of manual and auto-contouring of nodal volumes in patients who underwent axillary dissection.*

Structure†	Left/right	No.	VR	95% HD, mm	sDSC	vDSC
LN_Ax_L1	Left	10	0.73 ± 0.21	18.90 ± 12.29	0.25 ± 0.17	0.65 ± 0.14
	Right	10	0.68 ± 0.32	20.86 ± 10.82	0.18 ± 0.11	0.60 ± 0.19
LN_Ax_L2	Left	11	0.90 ± 0.31	9.18 ± 4.17	0.49 ± 0.24	0.71 ± 0.17
	Right	11	0.75 ± 0.19	8.56 ± 4.04	0.59 ± 0.18	0.76 ± 0.15
LN_Ax_L3	Left	11	0.87 ± 0.36	8.98 ± 4.01	0.31 ± 0.15	0.68 ± 0.18
	Right	11	0.86 ± 0.17	7.05 ± 2.85	0.41 ± 0.16	0.75 ± 0.09
LN_Ax_L4	Left	11	0.62 ± 0.28	10.51 ± 7.10	0.32 ± 0.18	0.63 ± 0.21
	Right	11	0.52 ± 0.17	10.24 ± 6.68	0.30 ± 0.13	0.60 ± 0.16
LN_Ax_Pectoral	Left	11	0.94 ± 0.22	8.22 ± 12.62	0.67 ± 0.35	0.75 ± 0.31
	Right	11	0.78 ± 0.15	7.50 ± 4.52	0.69 ± 0.27	0.76 ± 0.22
LN_IMN	Left	10	0.60 ± 0.21	16.39 ± 19.38	0.47 ± 0.21	0.66 ± 0.18
	Right	8	0.77 ± 0.16	15.25 ± 13.14	0.71 ± 0.11	0.80 ± 0.09

Abbreviations: 95% HD = 95th percentile Hausdorff distance; sDSC = surface Dice similarity coefficient; vDSC = volume Dice similarity coefficient; VR = volume ratio.
* Data are shown as mean ± standard deviation, unless otherwise specified.
† Nomenclature of the target region on the treatment planning system.

Table 6. Results of comparison of manual and auto-contouring of nodal volumes in patients who underwent sentinel lymph node biopsy.*

Structure†	Left/right	No.	VR	95% HD, mm	sDSC	vDSC
LN_Ax_L1	Left	3	0.81 ± 0.29	20.56 ± 14.01	0.26 ± 0.11	0.77 ± 0.08
	Right	2	0.77 ± 0.27	11.70 ± 5.80	0.29 ± 0.16	0.75 ± 0.13
LN_Ax_L2	Left	3	0.91 ± 0.09	6.42 ± 1.45	0.58 ± 0.18	0.77 ± 0.09
	Right	2	0.87 ± 0.05	4.71 ± 1.22	0.77 ± 0.10	0.88 ± 0.04
LN_Ax_L3	Left	3	0.98 ± 0.19	7.18 ± 1.05	0.32 ± 0.14	0.73 ± 0.06
	Right	2	0.81 ± 0.07	6.00 ± 1.86	0.56 ± 0.21	0.83 ± 0.04
LN_Ax_L4	Left	3	0.59 ± 0.17	10.55 ± 2.98	0.29 ± 0.12	0.62 ± 0.10
	Right	2	0.72 ± 0.15	6.39 ± 3.30	0.44 ± 0.19	0.70 ± 0.16
LN_Ax_Pectoral	Left	3	0.84 ± 0.04	6.24 ± 0.34	0.87 ± 0.04	0.89 ± 0.04
	Right	2	0.94 ± 0.03	3.89 ± 1.51	0.88 ± 0.03	0.90 ± 0.03
LN_IMN	Left	3	0.73 ± 0.18	13.05 ± 4.30	0.49 ± 0.18	0.66 ± 0.10
	Right	2	0.71 ± 0.07	5.82 ± 0.78	0.72 ± 0.02	0.83 ± 0.04

Abbreviations: 95% HD = 95th percentile Hausdorff distance; N/A = not applicable; sDSC = surface Dice similarity coefficient; vDSC = volume Dice similarity coefficient; VR = volume ratio.
* Data are shown as mean ± standard deviation, unless otherwise specified.
† Nomenclature of the target region on the treatment planning system.

significantly lower mean doses to the contralateral lung and the heart compared to VMAT (manual) [$p = 0.027$ and $p = 0.009$, respectively] (Table 7). The time required for treatment planning of a VMAT (ML) plan was 11.65 ± 2.00 minutes, while that of a VMAT (manual) plan was 15.42 ± 5.24 minutes ($p = 0.023$).

Intensity Modulated Radiation Therapy Created with Machine Learning Versus Created Manually

IMRT (ML) plans and VMAT (manual) plans had comparable performance in terms of target coverage for the breast or chest wall and PTV (axillary levels 3 and 4); however, IMRT (ML) has a higher coverage of the PTV (IMN) than IMRT (manual) [99.06% vs. 96.96%; $p = 0.013$]. For OARs, IMRT (ML) plans had a higher dose to the ipsilateral lung in terms of both mean dose and V_{17Gy} (both $p < 0.001$). The mean dose to ipsilateral lung (14.81 ± 1.26 Gy) failed to fulfil the dose constraint. The differences in mean dose and V_{17Gy} to the heart were not statistically significant (Table 8). The time required for treatment planning of an IMRT (ML) plan was 51.47 ± 12.46 minutes, while that of a IMRT (manual) plan was 30.84 ± 5.20 minutes ($p < 0.001$).

DISCUSSION

Contouring

AI has been revolutionising various sectors, including the medical field, impacting the system from daily patient care to policy formulation.[27,28] The advancements in AI have significantly enhanced radiological imaging

Table 7. Parameters of dose-volume histograms of volumetric modulated arc therapy created with machine learning and created manually.*

	Parameter†	VMAT (ML)	VMAT (manual)	p Value‡
PTV (whole breast/chest wall)	V_{38Gy}, %	96.19 ± 2.48	97.50 ± 1.42	0.034
PTV (whole breast/chest wall)	V_{42Gy}, %	5.54 ± 1.53	4.88 ± 1.44	0.034
PTV (whole breast/chest wall)	$V_{42.8Gy}$, %	0.05 ± 0.08	0.09 ± 0.08	0.013
PTV (whole breast/chest wall)	V_{44Gy}, cc	0.00 ± 0.00	0.00 ± 0.00	0.344
PTV (L3 + L4§)	V_{36Gy}, %	98.50 ± 1.49	99.93 ± 0.10	< 0.001
PTV (L3 + L4§)	V_{44Gy}, cc	0.00 ± 0.00	0.00 ± 0.00	0.344
PTV (IMN)	V_{36Gy}, %	96.91 ± 2.54	99.06 ± 1.89	< 0.001
PTV (IMN)	V_{44Gy}, cc	0.00 ± 0.00	0.00 ± 0.01	0.169
Ipsilateral lung	Mean dose, Gy	12.31 ± 0.72	12.49 ± 0.38	0.400
Ipsilateral lung	V_{17Gy}, %	29.33 ± 2.09	31.31 ± 1.57	< 0.001
Contralateral lung	Mean dose, Gy	2.98 ± 0.71	3.26 ± 0.55	0.027
Heart	Mean dose, Gy	3.16 ± 0.78	3.65 ± 0.62	0.009
Heart	V_{17Gy}, %	1.86 ± 1.68	2.24 ± 2.10	0.595
Contra-lateral breast	Mean dose, Gy	3.30 ± 0.99	3.13 ± 0.49	0.242

Abbreviations: IMN = internal mammary node; ML = machine learning; PTV = planning target volume; VMAT = volumetric modulated arc therapy.

* Data are shown as mean ± standard deviation, unless otherwise specified.

† Expressed as relative volume (%) or volume (cc) of the PTV or organ at risk receiving at least the radiation dose specified, unless otherwise specified.

‡ VMAT (ML) vs. VMAT (manual).

§ Axillary levels 3 and 4.

Table 8. Parameters of dose-volume histograms of intensity modulated radiation therapy created with machine learning and created manually.*

	Parameter†	IMRT (ML)	IMRT (manual)	p Value‡
PTV (whole breast/chest wall)	V_{38Gy}, %	96.47 ± 3.22	96.53 ± 2.58	0.863
PTV (whole breast/chest wall)	V_{42Gy}, %	5.05 ± 2.24	4.94 ± 2.62	0.556
PTV (whole breast/chest wall)	$V_{42.8Gy}$, %	0.42 ± 0.43	0.31 ± 0.46	0.156
PTV (whole breast/chest wall)	V_{44Gy}, cc	0.16 ± 0.45	0.02 ± 0.09	0.003
PTV (L3 + L4§)	V_{36Gy}, %	99.85 ± 0.25	99.75 ± 0.27	0.093
PTV (L3 + L4§)	V_{44Gy}, cc	0.00 ± 0.00	0.00 ± 0.00	1.000
PTV (IMN)	V_{36Gy}, %	99.06 ± 1.52	96.96 ± 4.20	0.013
PTV (IMN)	V_{44Gy}, cc	0.00 ± 0.01	0.00 ± 0.00	0.336
Ipsilateral lung	Mean dose, Gy	14.81 ± 1.26	12.39 ± 0.68	< 0.001
Ipsilateral lung	V_{17Gy}, %	31.97 ± 1.62	28.16 ± 2.39	< 0.001
Contralateral lung	Mean dose, Gy	2.40 ± 0.68	2.97 ± 0.62	< 0.001
Heart	Mean dose, Gy	4.36 ± 0.91	3.94 ± 1.39	0.081
Heart	V_{17Gy}, %	2.65 ± 2.16	3.22 ± 3.72	0.755
Contralateral breast	Mean dose, Gy	2.27 ± 0.81	2.24 ± 0.37	0.619

Abbreviations: IMN = internal mammary node; IMRT = intensity modulated radiation therapy; ML = machine learning; PTV = planning target volume.

* Data are shown as mean ± standard deviation, unless otherwise specified.

† Expressed as relative volume (%) or volume (cc) of the PTV or organ at risk receiving at least the radiation dose specified, unless otherwise specified.

‡ IMRT (ML) vs. IMRT (manual).

§ Axillary levels 3 and 4.

through improved image recognition using deep learning algorithms. The clinical utilisation of AI in radiotherapy delivery is on the rise, encompassing tasks such as segmentation and treatment planning, in malignancies arising from different regions including head and neck, thorax and pelvis.[29] The primary goals are to enhance quality, decrease inter-observer variability, standardise contouring, and increase efficiency.

Based on our study findings, AI demonstrated encouraging results in segmentation of OARs. The results of 95% HD and vDSC for heart and lungs

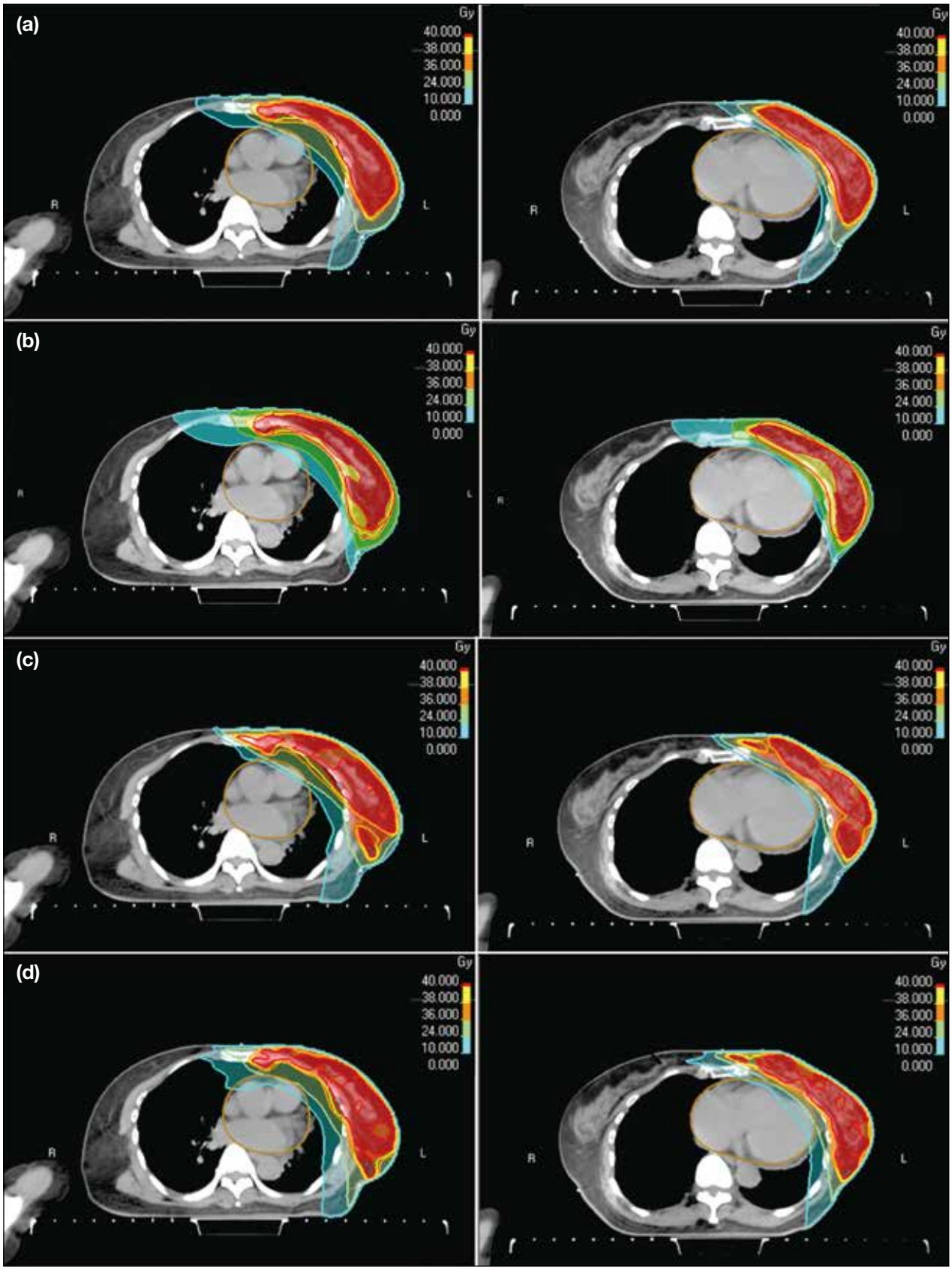

Figure 2. Images of the same computed tomography slice at two different levels in the same patient showing the dose distribution achieved by (a) volumetric modulated arc therapy (VMAT) created by machine learning, (b) VMAT created manually, (c) intensity modulated radiation therapy (IMRT) created by machine learning and (d) IMRT created manually.

were comparable to those of other AI models.[10,30] This achievement can be attributed to the clear visualisation and differentiation of the heart and lungs, which also exhibit minimal variability across different patients. For clinical target volume, favourable performance was observed for axillary level 2 and interpectoral nodes. Delineation of these structures is closely associated with identification of the pectoralis minor muscle, which is a readily identified structure on CT, thus contributing to improved consistency in contouring.

The deep learning segmentation model employed in our study exhibited limitations in segmentation in some of the patients, failing to delineate axillary level 1 in two patients (7.4%) and the IMN in four patients (14.8%). In addition, the model's inability to identify the chest wall underscores its constraints in clinical application. The agreement between auto-segmentation and manual contouring was found to be less consistent for axillary level 1 compared to axillary levels 2 to 4, with larger 95% HD and smaller sDSC. From the analysis of SLNB and AD patients, better outcomes were more likely in those who underwent SLNB instead of AD, despite the small number of SLNB cases. Contouring of axillary level 1 is typically influenced by axillary surgery to encompass the postoperative OARs or areas with surgical clips based on clinical judgement rather than solely relying on anatomical structures, leading to possible inter-observer variability. The results suggest that disrupted anatomy may also pose a challenge for auto-segmentation.

Apart from individual factors, variations between manual contouring and auto-segmentation can also be attributed to systemic factors. For example, patients from our institute may have had different physiques compared to the patients recruited in the AI training dataset. Additionally, the quality of the CT imaging can be a contributing factor, as it plays a significant role in the visibility of structures such as the axillary and subclavian vessels. These structures are crucial for defining the boundaries of the axillary levels and consequently influence the precision of target volume delineation by both manual and auto-contouring.

Regarding the time spent for auto-segmentation and the potential of AI to improve clinical efficiency, in our study, the mean duration for auto-segmentation was found to be <30 seconds. Although the time spent on manual contouring was not documented, clinical experience from our institute and existing literature[31,32] suggest that it typically requires approximately 35 minutes for manual contouring of a breast radiotherapy plan, which is significantly longer than the time required by AI.

At present, although vDSC with 0.7 indicates excellent agreement in general,[25] there are no established acceptable thresholds for the four metrics for clinical use. In this study, the metrics provided only objective numerical analysis, without incorporating qualitative review by oncologists to gauge user satisfaction. This is a limitation of this study, as the ultimate goal of reviewing performance of AI in this domain is to augment manual processes, where user acceptance is crucial for clinical implementation of the technology.[33] Additionally, the small number of oncologists reviewing the manual contours as ground truth may not adequately represent the intra- and inter-observer variability in reality. This could be better addressed by creating a consensus manual contour developed by a larger group of oncologists.

Another limitation of this study is that, while the findings indicate that auto-segmentation requires significantly less time than manual contouring, further assessment of the time needed for manual review and correction of auto-segmentation is necessary to fully evaluate its time-saving benefits. Surface DSC is a relatively novel metric for assessing the performance of auto-segmentation and has been demonstrated to be a more effective indicator of correction time compared to traditional metrics, such as vDSC.[23,24] Generally, it is anticipated that no time savings will be realised if an observer expects to alter an automatically generated contour by more than approximately 40%.[23] Future research is needed to investigate the relationship between surface DSC and other relevant metrics with time savings, and establish acceptable thresholds for clinical practice.

Treatment Planning

Adjuvant radiotherapy for breast cancer has been associated with long-term complications. One notable concern is the increased risk of cardiac toxicity,[34] and the risk is particularly pronounced in women with existing cardiac risk factors, or women receiving radiotherapy for left-sided breast cancer. Additionally, there is a risk of pneumonitis[35] and the development of second primary cancers.[36] Although a recent meta-analysis found no significant differences in the occurrence of cardiac events and pneumonitis between patients who received IMN irradiation and those who did not,[37] it is still recommended to prioritise minimising the radiation dose to OARs as much as possible, following the ALARA (As

Low As Reasonably Achievable) principle.[38]

Facing the expected growing numbers of patients undergoing radiotherapy including the IMN, a shift from three-dimensional conformal radiation therapy to inverse planning is needed to improve the quality of radiotherapy plan. As it is likely to impose a substantial increase in the workload associated with treatment planning, there is a need to see if AI can facilitate the radiotherapy planning process. The goal is to carry out treatment planning efficiently while maintaining established standards.

The virtual plans in this study, including VMAT (ML), VMAT (manual), IMRT (ML), and IMRT (manual), were created based on contouring performed by oncologists, and all of them demonstrated satisfactory target coverage. VMAT (ML), VMAT (manual) and IMRT (manual) successfully met all of the dose constraints, while IMRT (ML) exceeded the objective for the ipsilateral lung mean dose, which was measured at 14.81 ± 1.26 Gy, surpassing the desired limit of ≤13 Gy.

In the evaluation of VMAT plans, VMAT (manual) demonstrated superior target coverage compared to VMAT (ML) plans, which exhibited a trend towards reduced doses to OARs. Both planning approaches, however, are deemed acceptable from a dosimetric perspective.

In the context of IMRT plans, IMRT (manual) is regarded as more acceptable, as it successfully meets the specified dose objectives and constraints. In contrast, IMRT (ML) will require manual adjustments and optimisation before clinical application.

In our study, the time required to generate a VMAT (ML) plan was faster by 3.77 minutes compared to VMAT (manual) plan. The time taken to create an IMRT (manual) plan was faster than the IMRT (ML) plan by 20.63 minutes. Notably, during the AI planning process for both IMRT and VMAT, planners can attend to other responsibilities while the system generates the treatment plans; thus, the actual clinical time dedicated to AI planning may be less than the recorded duration. On the other hand, it is also important to acknowledge that AI-generated plans are not inherently perfect and may necessitate further adjustments to tailor the treatment for individual patients. A limitation of this study is that it does not quantify the time required for optimising an AI-generated plan after its initial creation. These factors may influence the actual clinical time saving benefits derived from the use of AI in treatment planning.

CONCLUSION

In the first part of the study, AI has demonstrated reliable and excellent performance in auto-segmentation of OARs including heart and lungs for breast cancer radiotherapy. While the technology may also be applied to the delineation of the breast and regional nodal targets, it is crucial to bear in mind the need for case-by-case adjustment, particularly for postoperative changes.

In the second part, the study showed that satisfactory VMAT plans can be efficiently generated using an automated approach, achieving performance comparable to that of manual planning in a relatively short timeframe. This method holds potential for clinical application, warranting further investigation into its impact on improving workflow in radiotherapy practices.

REFERENCES

1. World Health Organization. Breast cancer. 2024 Mar 13. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Accessed 2 Jul 2024.
2. Centre for Health Protection, Department of Health, Hong Kong SAR Government. Breast cancer. 2025 Feb 24. Available from: https://www.chp.gov.hk/en/healthtopics/content/25/53.html. Accessed 5 Mar 2025.
3. Thorsen LB, Overgaard J, Matthiessen LW, Berg M, Stenbygaard L, Pedersen AN, et al. Internal mammary node irradiation in patients with node-positive early breast cancer: fifteen-year results from the Danish Breast Cancer Group Internal Mammary Node study. J Clin Oncol. 2022;40:4198-206.
4. Boekel NB, Schaapveld M, Gietema JA, Russell NS, Poortmans P, Theuws JC, et al. Cardiovascular disease risk in a large, population-based cohort of breast cancer survivors. Int J Radiat Oncol Biol Phys. 2016;94:1061-72.
5. Chargari C, Castadot P, Macdermed D, Vandekerkhove C, Bourgois N, Van Houtte P, et al. Internal mammary lymph node irradiation contributes to heart dose in breast cancer. Med Dosim. 2010;35:163-8.
6. Siddique S, Chow JC. Artificial intelligence in radiotherapy. Rep Prac Oncol Radiother. 2020;25:656-66.
7. Francolini G, Desideri I, Stocchi G, Salvestrini V, Ciccone LP, Garlatti P, et al. Artificial intelligence in radiotherapy: state of the art and future directions. Med Oncol. 2020;37:50.
8. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Phys Med. 2018;50:13-9.
9. Liu Z, Liu F, Chen W, Tao Y, Liu X, Zhang F, et al. Automatic segmentation of clinical target volume and organs-at-risk for breast conservative radiotherapy using a convolutional neural network. Cancer Manag Res. 2021;13:8209-17.
10. Almberg SS, Lervåg C, Frengen J, Eidem M, Abramova TM, Nordstrand CS, et al. Training, validation, and clinical implementation of a deep-learning segmentation model for radiotherapy of loco-regional breast cancer. Radiother Oncol. 2022:173:62-8.

11. van de Sande D, Sharabiani M, Bluemink H, Kneepkens E, Bakx N, Hagelaar E, et al. Artificial intelligence–based treatment planning of radiotherapy for locally advanced breast cancer. Phys Imaging Radiat Oncol. 2021;20:111-6.
12. Kneepkens E, Bakx N, van der Sangen M, Theuws J, van der Toorn PP, Rijkaart D, et al. Clinical evaluation of two AI models for automated breast cancer plan generation. Radiat Oncol. 2022;17:25.
13. American Joint Committee on Cancer. AJCC Stating Manual Eight Edition. 2018. Available from: http://www.breastsurgeonsweb.com/wp-content/uploads/downloads/2020/10/AJCC-Breast-Cancer-Staging-System.pdf. Accessed 4 Mar 2025.
14. Varian. Eclipse. Available from: https://www.varian.com/products/radiotherapy/treatment-planning/eclipse. Accessed 2 Jul 2024.
15. RaySearch Laboratories. RayStation. Available from: https://www.raysearchlabs.com/raystation/. Accessed 2 Jul 2024.
16. Offersen BV, Boersma LJ, Kirkove C, Hol S, Aznar MC, Biete Sola A, et al. ESTRO consensus guideline on target volume delineation for elective radiation therapy of early-stage breast cancer. Radiother Oncol. 2015;114:3-10.
17. The Royal College of Radiologists. Postoperative radiotherapy for breast cancer: UK consensus statements. November 2016. Available from: https://www.rcr.ac.uk/our-services/all-our-publications/clinical-oncology-publications/postoperative-radiotherapy-for-breast-cancer-uk-consensus-statements/. Accessed 2 Jul 2024.
18. Danish Breast Cancer Cooperative Group. The SKAGEN Trial 1. Moderately hypofractionated loco-regional adjuvant radiation therapy of early breast cancer combined with a simultaneous integrated boost in patients with an indication for boost: DBCG HYPO II, a randomised clinically controlled trial. 2015. Available from: https://www.dbcg.dk/PDF%20Filer/SKAGEN%20Trial%201_%20protokol.pdf. Accessed 2 Jul 2024.
19. Mackay K, Bernstein D, Glocker B, Kamnitsas K, Taylor A. A review of the metrics used to assess auto-contouring systems in radiotherapy. Clin Oncol (R Coll Radiol). 2023;35:354-69.
20. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Mach Intell. 1993;15:850-63.
21. Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, et al. Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. J Med Internet Res. 2021;23:e26151.
22. Doolan PJ, Charalambous S, Roussakis Y, Leczynski A, Peratikou M, Benjamin M, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. Front Oncol. 2023;13:1213068.
23. Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. Phys Imaging Radiat Oncol. 2019;13:1-6.
24. Kiser KJ, Barman A, Stieb S, Fuller CD, Giancardo L. Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. J Digit Imaging. 2021;34:541-53.
25. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging. 1994;13:716-24.
26. Bartko JJ. Measurement and reliability: statistical thinking considerations. Schizophr Bull. 1991;17:483-9.
27. Ramezani M, Takian A, Bakhtiari A, Rabiee HR, Ghazanfari S, Mostafavi H. The application of artificial intelligence in health policy: a scoping review. BMC Health Serv Res. 2023;23:1416.
28. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019;8:2328-31.
29. Wang C, Zhu X, Hong JC, Zheng D. Artificial intelligence in radiotherapy treatment planning: present and future. Technol Cancer Res Treat. 2019;18:153303381987392.
30. Matoska T, Patel M, Liu H, Beriwal S. Review of deep learning based autosegmentation for clinical target volume: current status and future directions. Adv Radiat Oncol. 2024;9:101470.
31. Andrianarison VA, Laouiti M, Fargier-Bochaton O, Dipasquale G, Wang X, Nguyen NP, et al. Contouring workload in adjuvant breast cancer radiotherapy. Cancer Radiother. 2018;22:747-53.
32. Bakx N, Rijkaart D, van der Sangen M, Theuws J, van der Toorn PP, Verrijssen AS, et al. Clinical evaluation of a deep learning segmentation model including manual adjustments afterwards for locally advanced breast cancer. Tech Innov Patient Support Radiat Oncol. 2023;26:100211.
33. Buelens P, Willems S, Vandewinckele L, Crijns W, Maes F, Weltens CG. Clinical evaluation of a deep learning model for segmentation of target volumes in breast cancer radiotherapy. Radiother Oncol. 2022;171:84-90.
34. Bradshaw PT, Stevens J, Khankari N, Teitelbaum SL, Neugut AI, Gammon MD. Cardiovascular disease mortality among breast cancer survivors. Epidemiology. 2016;27:6-13.
35. Lingos TI, Recht A, Vicini F, Abner A, Silver B, Harris JR. Radiation pneumonitis in breast cancer patients treated with conservative surgery and radiation therapy. Int J Radiat Oncol Biol Phys. 1991;21:355-60.
36. Stovall M, Smith SA, Langholz BM, Boice JD Jr, Shore RE, Andersson M, et al. Dose to the contralateral breast from radiotherapy and risk of second primary breast cancer in the WECARE study. Int J Radiat Oncol Biol Phys. 2008;72:1021-30.
37. Shaikh PM, Mulherkar R, Khasawneh MT, Clump D, Hazard-Jenkins H, Hafez M, et al. Treatment of internal mammary nodes is associated with improved overall survival in breast cancer: a meta-analysis. Am J Clin Oncol. 2024;47:81-7.
38. Hendee WR, Edwards FM. ALARA and an integrated approach to radiation protection. Semin Nucl Med. 1986;16:142-50.