Multimodal Technologies and Interaction

MDPI

Systematic Review

Performance of Commercial Deep Learning-Based Auto-Segmentation Software for Breast Cancer Radiation Therapy Planning: A Systematic Review

Curtise K. C. Ng [1,2]

[1] Curtin Medical School, Curtin University, GPO Box U1987, Perth, WA 6845, Australia; curtise.ng@curtin.edu.au or curtise_ng@yahoo.com.hk; Tel.: +61-8-9266-7314; Fax: +61-8-9266-2377
[2] Curtin Health Innovation Research Institute (CHIRI), Faculty of Health Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

Abstract: As yet, no systematic review on commercial deep learning-based auto-segmentation (DLAS) software for breast cancer radiation therapy (RT) planning has been published, although NRG Oncology has highlighted the necessity for such. The purpose of this systematic review is to investigate the performances of commercial DLAS software packages for breast cancer RT planning and methods for their performance evaluation. A literature search was conducted with the use of electronic databases. Fifteen papers met the selection criteria and were included. The included studies evaluated eight software packages (Limbus Contour, Manteia AccuLearning, Mirada DLCExpert, MVision.ai Contour+, Radformation AutoContour, RaySearch RayStation, Siemens syngo.via RT Image Suite/AI-Rad Companion Organs RT, and Therapanacea Annotate). Their findings show that the DLAS software could contour ten organs at risk (body, contralateral breast, esophagus-overlapping area, heart, ipsilateral humeral head, left and right lungs, liver, and sternum and trachea) and three clinical target volumes (CTVp_breast, CTVp_chestwall, and CTVn_L1) up to the clinically acceptable standard. This can contribute to 45.4%–93.7% contouring time reduction per patient. Although NRO Oncology has suggested that every clinical center should conduct its own DLAS software evaluation before clinical implementation, such testing appears particularly crucial for Manteia AccuLearning, Mirada DLCExpert, and MVision.ai Contour+ as a result of the methodological weaknesses of the corresponding studies such as the use of small datasets collected retrospectively from single centers for the evaluation.

Keywords: artificial intelligence; artificial neural network; automatic; clinical target volumes; computed tomography; contouring; delineation; machine learning; organs at risk; radiotherapy

check for updates

Citation: Ng, C.K.C. Performance of Commercial Deep Learning-Based Auto-Segmentation Software for Breast Cancer Radiation Therapy Planning: A Systematic Review. *Multimodal Technol. Interact.* **2024**, *8*, 114. https://doi.org/10.3390/mti8120114

Academic Editor: Mark Billinghurst

Received: 1 November 2024
Revised: 16 December 2024
Accepted: 18 December 2024
Published: 20 December 2024

Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Breast cancer is the commonest cancer type, and also a leading cause of death in females in the world [1–3]. Sixty years ago, radiation therapy (RT) was solely used as an adjuvant therapy for high-risk breast cancer patients after mastectomy [4]. RT treatment was delivered based on two-dimensional (2D) imaging. Accurate organs at risk (OARs) and clinical target volumes (CTVs) control was not feasible [5]. Nowadays, breast RT is usually recommended for both postlumpectomy and postmastectomy patients to reduce recurrence of breast cancer [1,6–8]. Unlike other cancer types such as lung and head and neck cancers, field-based and 2D techniques are still used for RT in breast cancer [7,8]. However, use of intensity-modulated RT (IMRT) in breast cancer has become popular [1,7,8]. This technique relies on computed tomography (CT) to visualize OARs and CTVs for better OARs sparing and CTVs coverage [5].

Accurate segmentation of OARs and CTVs is essential for breast cancer IMRT planning to minimize treatment side effects and improve its effectiveness [1,7–10]. Typically, this contouring process takes between thirty minutes and one hour with a radiation therapist

(RTT) or a radiation oncologist (RO) manually handling each case [1,5,11,12]. Apart from the time-consuming aspect, it is also well known that the manual contouring in breast RT has notable inter- and intra-operator variability [1,8,13,14].

To address these issues, auto-segmentation solutions have been developed for breast RT [13,15–17]. These solutions include atlas- [2,4,13] and deep learning (DL)-based auto-segmentation [18–21]. Since use of DL in medical imaging has significantly increased recently [22–25], this has attracted commercial companies to focus on developing DL-based auto-segmentation (DLAS) software packages such as Limbus AI Inc. Contour (Regina, SK, Canada) [26]; Manteia Medical Technologies AccuLearning AI (Jiansheng, China) [27]; Mirada Medical Ltd. DLCExpert (Oxford, UK) [28]; MVision.ai Contour+ (Helsinki, Finland) [29]; Radformation Inc. AutoContour (New York, NY, USA) [29,30]; RaySearch Laboratories AB RayStation (Stockholm, Sweden) [31–34]; Siemens Healthineers AG syngo.via RT Image Suite [35,36] and AI-Rad Companion Organs RT (Erlangen, Germany) [37,38]; and Therapanacea Annotate (Paris, France) for breast RT planning [29,39].

Over the past few years, several articles have reviewed previous DLAS studies for breast cancer RT [40–43]. All except one were narrative reviews which covered DLAS approaches/architectures (such as convolutional neural network (CNN) variants, U-Net and V-Net), dataset source (e.g., public, private, etc.), type (such as CT) and size (e.g., 100 patients, etc.), structures segmented (such as heart (OAR) and breast (primary tumor) CTV (CTVp_breast)), and DLAS model performance (e.g., geometric accuracy, etc.) [40–43]. Although Matoska et al. [44] critically reviewed the performance of DLAS for breast RT planning in 2024, that review only included nine studies, and none of these used the commercial DLAS software. Hence, its findings are less relevant to clinical practice. The necessity of reviewing the commercial DLAS software packages is well recognized by the influential body NRG Oncology, which has formed a working group for this task [45,46]. In 2024, the working group published its outcomes as a narrative review on the commercial DLAS software packages for RT planning to enable clinical centers to recognize their strengths and limitations and make educated decisions on implementation of such packages in clinical practice. However, that narrative review did not cover breast cancer RT planning [46]. As such, no systematic review focused on the commercial DLAS software for breast cancer RT planning is available as yet. Without a systematic review to provide the gold standard evidence of the current performance of the commercial DLAS software for breast cancer RT planning with appraisal of associated methodology for evidence generation, it would be challenging for clinical centers to make any appropriate decision regarding its adoption. Furthermore, its potential benefits cannot be realized in a wider context. The purpose of this article is to systematically review original studies to answer this question: "What are the performances of commercial DLAS software packages for breast cancer RT planning and methods for their performance evaluation?"

2. Materials and Methods

This systematic review of the performance of the commercial DLAS software packages for breast cancer RT planning was conducted as per the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines and patient/population (breast cancer patients), intervention (use of commercial DLAS software for breast cancer RT planning), comparison (DLAS versus manual contouring (standard practice)), and outcome (structure segmentation performance) model (PICO) [47–50]. It involved four major processes, namely literature search, article selection, and data extraction and synthesis [47–49].

2.1. Literature Search

Seven electronic scholarly publication databases, namely Institute of Electrical and Electronics Engineers (IEEE) Xplore, PubMed, ScienceDirect, Scopus, SpringerLink, Web of Science, and Wiley Online Library, were employed for the literature search on 6 August 2024 to identify articles about the performance of the commercial DLAS software for breast cancer RT planning, with no publication year restriction. The following search statement

was used: "Commercial" AND ("Deep Learning" OR "Artificial Neural Network") AND ("Segmentation" OR "Delineation" OR "Contouring") AND "Breast Cancer" AND ("Radiotherapy" OR "Radiation Therapy"). The search keywords were derived from the review focus [40,41,44,51].

2.2. Article Selection

One reviewer was involved in the article selection. Table 1 illustrates inclusion and exclusion criteria for the articles [47–49].

Table 1. Inclusion and exclusion criteria for articles.

Inclusion Criteria	Exclusion Criteria
1. Written in English 2. Peer-reviewed, original research paper published in any year 3. Focused on use of commercial deep learning-based auto-segmentation software for breast cancer patients' radiation therapy planning	1. Commentary 2. Conference proceeding 3. Editorial 4. Grey literature 5. Non-peer-reviewed article (e.g., article on arXiv platform) 6. Opinion 7. Perspective 8. Review

The exclusion criteria of Table 1 were used as this systematic review focused on the performance of the commercial DLAS software for breast cancer RT planning and the appraisal of methodology for the performance evaluations that were reported in the peer-reviewed original research papers [47–49,52]. The conference proceedings were excluded because their academic rigor was generally lower. They were deemed unsuitable for the systematic review appraising the study methodology. This arrangement was in line with other systematic reviews on DL [52,53]. Figure 1 shows the details of the article selection process [47–49]. This involved duplicate article removal from the database search results. Titles of papers, their abstracts, and full texts were subsequently assessed based on the selection criteria. Every non-duplicate article within the search results was not removed unless an exclusion decision could be made. Additional papers were identified through checking reference lists of the included articles [47,48,54,55].

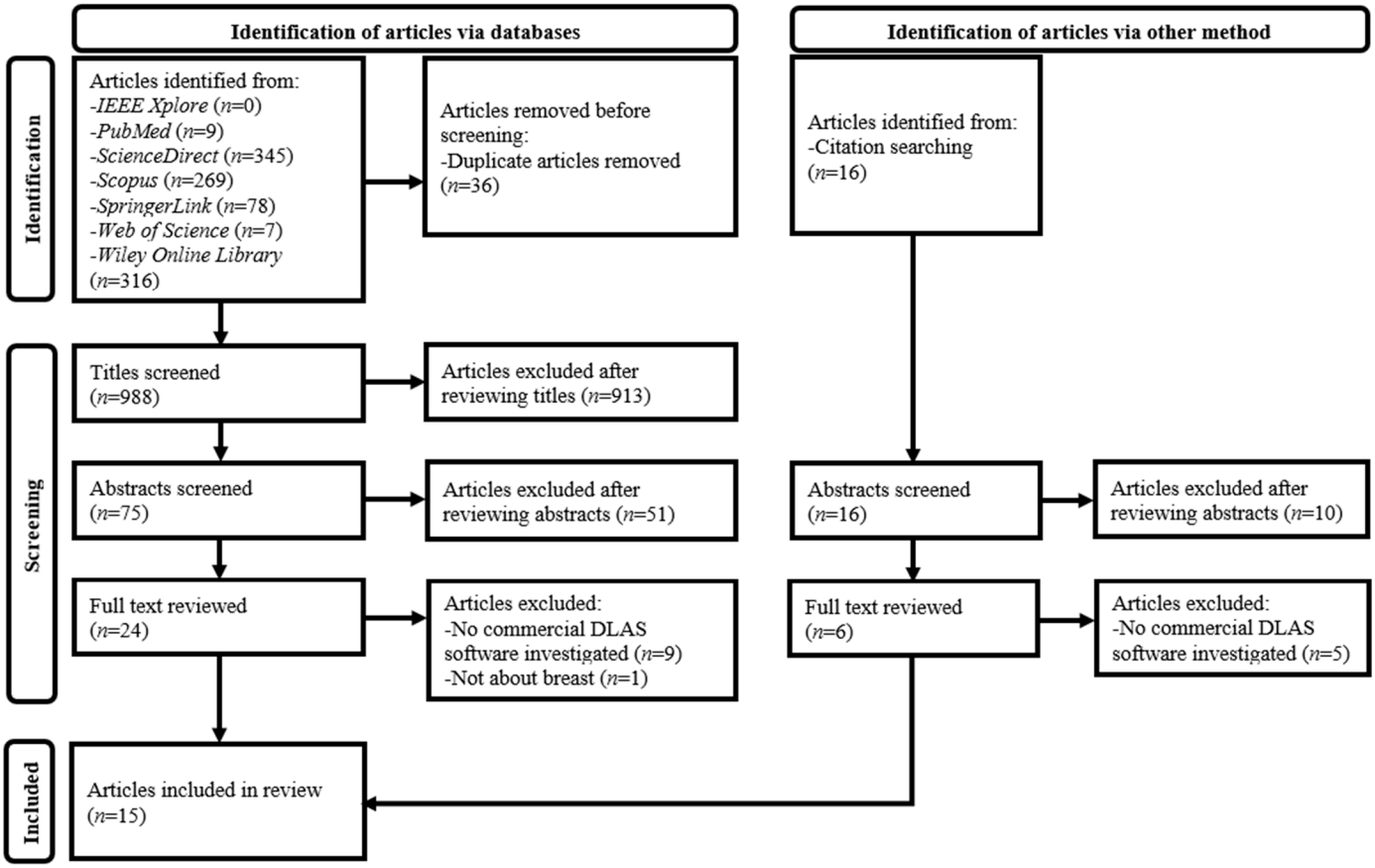

Figure 1. Preferred reporting items for systematic reviews and meta-analyses flow diagram for systematic review of performance of commercial deep learning-based auto-segmentation (DLAS) software for breast cancer radiation therapy planning. IEEE, Institute of Electrical and Electronics Engineers.

2.3. Data Extraction and Synthesis

A data extraction form was developed based on two narrative reviews (about DLAS for a range of RT plannings [46] including lung cancer RT [56]), one critical review (on DLAS for CTVs of different disease sites [44]), and two systematic reviews (about cervical [57] and prostate cancers [58]). The data extracted from each included article were author name and country; year of publication; software name and version (e.g., Limbus AI Inc. Contour v1.5.0, etc.); DLAS architecture (such as U-Net); design of study (either retrospective or prospective); any multi-center involvement; patient/population (e.g., left breast cancer patients after mastectomy, etc.); dataset source (such as public: US The Cancer Imaging Archive; and private: 1 Italian center) and size (e.g., 139 patients, etc.) for software development (model training) and evaluation (testing); any calculation of sample size for evaluation and external testing (i.e., software testing with use of dataset not involved in its development); modality of images for DLAS (such as CT); OARs (e.g., heart, etc.) and CTVs (such as interpectoral lymph node) segmented; reference contour (ground truth) source (e.g., 3 ROs and 3 RTTs with greater than 10 years of experience, etc.); contouring guidelines (such as European Society of Therapeutic Radiology and Oncology (ESTRO)) used; segmentation performances in terms of geometric accuracy (based on metrics such as Dice similarity coefficient (DSC)) for the DLAS and manual contouring (with regard to inter-observer variation) as well as any difference between these two; subjective accuracy evaluation (e.g., percentage of OARs and CTVs requiring no and minor corrections, etc.); efficiency evaluation (such as mean contouring time reduction percentage); and dosimetric impact (e.g., no clinically relevant difference of doses to OARs between plans by the DLAS and manual contouring) [44,56–58].

When a paper investigated both atlas- and DL-based auto-segmentation software packages or multiple disease sites, only data related to the DLAS for breast cancer RT planning were extracted. In addition, for facilitating the software comparison, only mean performance figures of the best model achieved in the external testing were included when multiple sets of evaluation findings were reported [47–49]. Furthermore, three data synthesis strategies were employed in this systematic review (if feasible): 1. taking the averages for figures regarding the OARs, left and right breasts as the contralateral breast figures [29,47–49]; 2. calculating the percentage of OARs and CTVs requiring no and minor corrections, as well as those considered unusable based on individual figures reported when an overall figure was not available [31,47–49]; and 3. calculating the percentage of mean contouring time reduction per patient with the use of reported absolute time required for correcting DLAS-generated contours (due to software processing time being hardware-dependent and insignificant) and manual contouring [37,47–49]. Meta-analysis was not conducted because the included studies employed a range of performance evaluation strategies leading to a high study heterogeneity and hence limiting its value [52,59,60]. The revised checklist for artificial intelligence in medical imaging (CLAIM) published in 2024 was used to assess quality of all included articles [61–64].

3. Results

Fifteen papers met the article selection criteria and were included in the review (Figure 1). Tables 2 and 3 illustrate the performances of the commercial DLAS software packages investigated by the included studies for breast cancer RT planning [26–39,65]. Eight software packages (Limbus AI Inc. Contour [26], Manteia Medical Technologies AccuLearning AI [27], Mirada Medical Ltd. DLCExpert [28,29], MVision.ai Contour+ [29], Radformation Inc. AutoContour [29,30], RaySearch Laboratories AB RayStation [29,31–34,65], Siemens Healthineers AG syngo.via RT Image Suite [35,36] and AI-Rad Companion Organs RT [35,37,38], and Therapanacea Annotate [29,39]) were covered in these studies.

DSC was employed in all studies to objectively evaluate the geometric accuracy of the structures contoured by the DLAS programs [26–39,65], and 95-percentile Hausdorff distance (HD95) was the second most common metric used (Table 2) [27,31–35,38,65]. All but two papers (86.7%) applied DLAS to contour the OARs. Fourteen structures were covered, collectively [26,28,29,31–38,65]. The reported mean/median DSC and/or HD95 for these structures were as follows: body (external contour): 0.99 [36]; contralateral breast: 0.72–0.94 [26,28,29,31,36–38,65] and 6.3–22.7 mm [31,39,65]; esophagus: 0.32–0.99 [28,29,31–33,36–38] and 4.0–161.0 mm [31–33,39]; esophagus-overlapping area: 0.85 and 2.3 mm [32]; heart: 0.88–0.96 [26,28,29,31–34,37–39,65] and 4.4–10.0 mm [31–35,39,65]; ipsilateral humeral head: 0.81–0.93 [29,31–33] and 4.4–8.3 mm [31–33]; left lung: 0.95–1.00 [26,28,29,31–39,65] and 1.4–8.8 mm [31–35,38,65]; right lung: 0.95–1.00 [26,28,29,31–39,65] and 1.4–8.8 mm [31–35,38,65]; left anterior descending artery (LAD): 0.39–0.54 and 7.3–18.2 mm [31,65]; liver: 0.96–0.97 [29]; sternum: 0.95 and 1.2 mm [31]; spinal canal: 0.69–0.98 [29,31,36–38] and 4.7 mm [31,38]; thyroid: 0.63–0.78 [28,31–33] and 4.5–8.2 mm [31–33]; and trachea: 0.93 and 4.9 mm [31]. Hence, the mean/median DSC and HD95 ranges of the DLAS for the OARs were 0.32 (esophagus) [32]−1.00 (lungs) [28] and 1.2 (sternum) [31]−161.0 mm (esophagus) [32], respectively.

Table 2. Geometric accuracy of commercial deep learning-based auto-segmentation (DLAS) software packages for breast cancer radiation therapy planning.

Software Name and Version	Author, Year and Country	Geometric Accuracy		
		DLAS	**IOV**	**DLAS VS IOV**
		Limbus AI Inc.		
Contour v1.5.0	Radici et al. (2022), Italy [26]	Mean DSC, DCOM (mm) and PVD: contralateral breast (0.72, 7.7 and −5.0%); heart (0.92, 4.2 and 12.0%); L (0.99, 0.1 and 1.0%) and R lungs (0.99, 0.2 and 1.0%)	NA	NA
		Manteia Medical Technologies		
AccuLearning AI	Hou et al. (2023), China [27]	Mean DSC and HD95 (mm) for U-Net: CTVp_breast (0.86 and 15.0)	NA	NA
		Mirada Medical Ltd.		
DLCExpert	Vaassen et al. (2022), The Netherlands and UK [28]	Median DSC, sDSC, APL (mm) and MSHD (mm): contralateral breast (0.90, 0.62, 2321.8 and 10.0); esophagus (0.60, 0.33, 756.0 and 3.3); heart (0.91, 0.61, 1497.5 and 7.5); L (1.00, 0.99, 513.0 and 1.3) and R lungs (1.00, 0.99, 601.5 and 1.4); thyroid (0.66, 0.34, 479.5 and 3.4); CTVp_breast (0.88, 0.57, 2665.8 and 12.4)	NA	NA
		Radformation Inc.		
AutoContour	Tsui et al. (2024), US [30]	Mean DSC, HD (mm) and MSD (mm): CTVn_L1-3 (0.70, 36.3 and 5.2); CTVn_L4 (0.54, 41.0 and 9.7); CTVn_IMN (0.33, 41.8 and 9.0); CTVp_breast (0.85, 38.1 and 4.3); CTVp_chestwall (0.71, 38.5 and 6.9)	NA	NA

Table 2. *Cont.*

Software Name and Version	Author, Year and Country	Geometric Accuracy		
		DLAS	**IOV**	**DLAS VS IOV**
		RaySearch Laboratories AB		
RayStation v9B	Almberg et al. (2022), Norway [31]	Mean DSC and HD95 (mm): CTVn_L1 (0.80 and 9.0); CTVn_L2 (0.76 and 10.0); CTVn_L3 (0.80 and 5.5); CTVn_L4 (0.80 and 4.3); CTVn_interpect (0.68 and 12.2); CTVn_IMN (0.71 and 8.0); CTVp_breast (0.95 and 5.3); contralateral breast (0.94 and 8.9); esophagus (0.85 and 4.0); heart (0.96 and 5.4); ipsilateral humeral head (0.93 and 4.4); L (0.98 and 1.6) and R lungs (0.98 and 1.5); LAD (0.54 and 7.3); spinal canal (0.91 and 4.7); sternum (0.95 and 1.2); thyroid (0.78 and 4.5); trachea (0.93 and 4.9)	Mean DSC and HD95 (mm): CTVn_L1 (0.74 and 14.6); CTVn_L2 (0.62 and 16.2); CTVn_L3 (0.67 and 9.4); CTVn_L4 (0.72 and 6.1); CTVn_interpect (0.61 and 14.5); CTVn_IMN (0.64 and 8.9); CTVp_breast (0.94 and 5.7); contralateral breast (0.91 and 11.2); esophagus (0.83 and 3.0); heart (0.95 and 6.7); LAD (0.44 and 20.7); spinal canal (0.85 and 8.8); thyroid (0.81 and 3.9); trachea (0.90 and 4.2)	DLAS outperforming ROs/RTTs for all CTVs and OARs with statistically significant differences ($p < 0.001$–0.022) except for DSC and HD95 of CTVp_breast and thyroid and HD95 of CTVn_interpect, contralateral breast, esophagus, heart and trachea
RayStation v9B/10B-SP1	Bakx et al. (2023), The Netherlands [32]	Mean DSC, HD95 (mm) and sDSC: CTVn_L1 (0.76, 13.3 and 0.65); CTVn_L2 (0.69, 10.1 and 0.80); CTVn_L3 (0.67, 8.7 and 0.75); CTVn_L4 (0.33, 16.4 and 0.43); CTVp_breast (0.92, 8.8 and 0.86); esophagus (0.32, 161.0 and 0.42); esophagus-overlapping area (0.85, 2.3 and 0.99); heart (0.93, 9.5 and 0.82); ipsilateral humeral head (0.85, 8.3 and 0.82); L (0.96, 4.6 and 0.93) and R lungs (0.96, 5.4 and 0.92); thyroid (0.71, 7.1 and 0.87)	NA	NA
RayStation v9B/10B-SP1	Bakx et al. (2023), The Netherlands [33]	Mean DSC, HD95 (mm) and sDSC: CTVn_L1 (0.78, 13.6 and 0.69); CTVn_L2 (0.71, 10.4 and 0.82); CTVn_L3 (0.73, 6.8 and 0.82); CTVn_L4 (0.57, 7.2 and 0.75); CTVp_breast (0.93, 14.4 and 0.83); esophagus (0.70, 10.4 and 0.88); heart (0.94, 7.1 and 0.81); ipsilateral humeral head (0.88, 7.6 and 0.86); L (0.98, 2.2 and 0.98) and R lungs (0.99, 2.2 and 0.98); thyroid (0.63, 8.2 and 0.81)	NA	NA

Table 2. *Cont.*

Software Name and Version	Author, Year and Country	Geometric Accuracy		
		DLAS	**IOV**	**DLAS VS IOV**
RayStation v11B-SP2	Mikalsen et al. (2023), Norway [34]	Mean DSC and HD95 (mm): CTVn_L1 (0.72 and 12.0); CTVn_L2 (0.66 and 12.0); CTVn_L3 (0.76 and 7.0); CTVn_L4 (0.70 and 7.7); CTVn_interpect (0.66 and 12.0); CTVn_IMN (0.67 and 12.0); CTVp_breast (0.91 and 9.8); heart (0.94 and 5.6); lungs (0.98 and 1.4)	NA	NA
RayStation v9B	Zeverino et al. (2024), Switzerland [65]	Median DSC, HD95 (mm), sDSC, HD (mm), HD99 (mm) and ΔV: contralateral breast (0.90, 6.3, 0.90, 15.5, 11.7 and −4.7%); heart (0.94, 6.8, 0.86, 11.0, 8.0 and −4.2%); L (0.98, 2.1, 0.98, 23.5, 7.2 and 0.2%) and R lungs (0.98, 2.0, 0.98, 24.3, 7.9 and −0.3%); LAD (0.39, 18.2, 0.73, 25.2, 23.1 and 1.9 cm^3)	NA	NA
		Siemens Healthineers AG		
syngo.via RT Image Suite VB50/AI-Rad Companion Organs RT VA20	Marschner et al. (2022), Germany and US [35]	Mean DSC, HD95 (mm), MSD (mm), ΔV, RMSD (mm), sensitivity, specificity, JCI, DI, GMI, CVD (mm) and L, R, anterior, posterior, superior and inferior boundaries (mm): heart (0.92, 4.4, 1.6, 2.1%, 2.2, 0.91, 0.99, 0.85, 0.06, 0.08, 4.7, −0.3, 0.0, −0.4, 0.0, −4.9 and −8.5); L (0.97, 2.7, 0.8, −0.9%, 1.8, 0.98, 0.99, 0.95, 0.03, 0.02, 2.0, −0.26, −4.9, 0.6, −0.1, −0.7 and −1.8) and R lungs (0.97, 2.9, 1.0, −0.9%, 1.8, 0.98, 0.99, 0.95, 0.03, 0.03, 2.1, 2.5, 0.8, 0.7, −0.2, −0.5 and −1.5)	NA	NA

Table 2. *Cont.*

Software Name and Version	Author, Year and Country	Geometric Accuracy		
		DLAS	**IOV**	**DLAS VS IOV**
AI-Rad Companion Organs RT VA31	Hu et al. (2023), Australia [37]	Mean DSC, HD (mm), sensitivity and precision: contralateral breast (0.89, 23.3, 0.91 and 0.88); esophagus (0.75, 15.9, 0.76 and 0.75); heart (0.93, 11.1, 0.89 and 0.98); L (0.96, 21.7, 0.99 and 0.93) and R lungs (0.97, 28.3, 0.99 and 0.95); spinal canal (0.69, 4.9, 0.98 and 0.54)	NA	NA
syngo.via RT Image Suite VB40	Pera et al. (2023), Germany and Spain [36]	Median DSC: body (0.99); contralateral breast (0.89); esophagus (0.99); L (0.98) and R lungs (0.98); spinal canal (0.98)	NA	NA
AI-Rad Companion Organs RT VA30	Yamauchi et al. (2024), Japan [38]	Median DSC, HD95 (mm) and MDA (mm): contralateral breast (0.89, 22.7 and 2.2); esophagus (0.80, 4.4 and 0.7); heart (0.95, 10.0 and 1.5); L and R lungs (0.97, 8.8 and 0.9); spinal canal (0.78, 4.7 and 1.0)	NA	NA
		Limbus AI Inc, RaySearch Laboratories AB, and Therapanacea		
Contour v1.5.0, RayStation v11B and Annotate v1.10.0	Heilemann et al. (2023), Austria [39]	Median DSC and HD (mm) for Contour/RayStation/Annotate: heart (0.88 and 1.6/0.91 and 1.3/0.88 and 1.9); L and R lungs (0.97 and 2.0/0.95 and 1.4/0.97 and 1.4)	NA	NA

Table 2. *Cont.*

Software Name and Version	Author, Year and Country	Geometric Accuracy		
		DLAS	**IOV**	**DLAS VS IOV**
Mirada Medical Ltd., MVision.ai, Radformation Inc., RaySearch Laboratories AB and Therapanacea				
DLCExpert v2.6.4.47181, Contour+ v1.2.1, AutoContour v1.0.25.0, RayStation v12.0.0.932 and Annotate v1.10.0	Doolan et al. (2023), Cyprus and Germany [29]	Median DSC, HD (mm), sDSC and APL (mm) for DLCExpert/Contour+/AutoContour/RayStation/Annotate: contralateral breast (0.86, 29.2, 0.25 and 28,963.0/0.90, 21.4, 0.34 and 25,807.5/0.82, 37.5, 0.16 and 34,386.0/0.84, 24.8, 0.15 and 34,985.5/0.89, 24.9, 0.31 and 28,250.0); esophagus (0.73, 21.2, 0.50 and 3522.0/0.79, 19.1, 0.59 and 2504.0/0.76, 19.4, 0.51 and 3200.0/0.81, 13.6, 0.64 and 2260.0/0.84, 9.9, 0.63 and 2441.0); heart (0.94, 16.7, 0.42 and 19236.0/0.95, 10.8, 0.49 and 17,922.0/0.95, 10.7, 0.46 and 20,252.0/0.95, 12.0, 0.46 and 19,262.0/0.94, 10.5, 0.48 and 18,115.0); ipsilateral humeral head (NA/0.91, 19.9, 0.66 and 2670.0/0.91, 20.6, 0.68 and 2216.0/0.81, 44.4, 0.53 and 4520.0/0.86, 36.9, 0.63 and 3567.0); L (0.97, 24.1, 0.56 and 35,206.0/0.97, 25.3, 0.61 and 32,763.0/0.96, 24.6, 0.54 and 37,397.0/0.96, 28.4, 0.55 and 37,980.0/0.97, 26.8, 0.61 and 32,844.0) and R lungs (0.96, 19.7, 0.57 and 29,982.0/0.96, 20.4, 0.60 and 27,896.0/0.95, 21.7, 0.48 and 35,266.0/0.96, 23.4, 0.57 and 31,733.0/0.96, 19.5, 0.60 and 28,851.0); liver (0.96, 17.4, 0.55 and 29,255.0/0.97, 18.4, 0.60 and 25,995.0/0.96, 24.9, 0.54 and 31,248.0/0.96, 22.7, 0.58 and 26,532.0/0.97, 22.1, 0.59 and 28,043.0); spinal canal (0.82, 9.2, 0.52 and 5239.0/0.83, 6.5, 0.48 and 5293.5/0.84, 6.3, 0.53 and 5092.0/0.84, 7.7, 0.53 and 4959.0/0.85, 6.7, 0.55 and 5175.0)	NA	NA

APL, added path length; CTV, clinical target volume; CTVn_IMN, internal mammary lymph node clinical target volume; CTVn_interpect, interpectoral lymph node clinical target volume; CTVn_L1, axillary level 1 lymph node clinical target volume; CTVn_L2, axillary level 2 lymph node clinical target volume; CTVn_L3, axillary level 3 lymph node clinical target volume; CTVn_L4, supraclavicular lymph node clinical target volume; CTVp_breast, breast (primary tumor) clinical target volume; CTVp_chestwall, post-mastectomy chestwall (primary tumor) clinical target volume; CVD, center of volume distance; DCOM, displacement of center of mass; DI, discordance index; DSC, Dice similarity coefficient; GMI, geographical miss index; HD, Hausdorff distance; HD95, 95-percentile Hausdorff distance; HD99, 99-percentile Hausdorff distance; IOV, inter-observer variation; JCI, Jaccard conformity index; L, left; LAD, left anterior descending artery; MDA, mean distance to agreement; MSD, mean surface distance; MSHD, mean slice-wise Hausdorff distance; NA, not available; OAR, organ at risk; PVD, percentage volume difference; R, right; RMSD, residual mean surface distance; RO, radiation oncologist; RTT, radiation therapist; sDSC, surface Dice similarity coefficient; UK, United Kingdom; US, United States; VS, versus; ΔV, difference of volumes segmented by deep learning and manual approaches.

Table 3. Subjective accuracy, efficiency and dosimetric evaluation results of commercial deep learning-based auto-segmentation (DLAS) software packages for breast cancer radiation therapy planning.

Software Name and Version	Author, Year and Country	Evaluation Results		
		Subjective	Efficiency	Dosimetric
		Limbus AI Inc.		
Contour v1.5.0	Radici et al. (2022), Italy [26]	NA	Mean time reduction/patient: 46.0% (7.0 min)	No clinically relevant difference of doses to OARs between DLAS and manual contouring
		Manteia Medical Technologies		
AccuLearning AI	Hou et al. (2023), China [27]	No/minor corrections required for 13.0%/75.0% of CTVs; No unusable CTVs contours	NA	NA
		Radformation Inc.		
AutoContour	Tsui et al. (2024), United States [30]	NA	NA	ΔV90/95% < 5% with DSC > 0.70: 94.1%, 67.7%, 14.7% and 0.0% for CTVp_breast, CTVn_L1-3, CTVn_L4 and CTVn_IMN of BCS patients; 62.5%, 56.3%, 9.4% and 3.1% for CTVp_chestwall, CTVn_L1-3, CTVn_L4 and CTVn_IMN of mastectomy patients, respectively. ΔV95% used for all structures except CTVn_IMN
		RaySearch Laboratories AB		
RayStation v9B	Almberg et al. (2022), Norway [31]	No/minor corrections required for 72.0%/26.0% of OARs and 14.0%/71.0% of CTVs; No unusable OARs and CTVs contours	Estimation of time reduction/patient: 75.0% (manual: 60.0 min VS DLAS: 15.0 min)	CTV coverage (D98 > 95%): 100.0% for breast and 89.0% for lymph nodes; no clinically relevant difference of doses to OARs between DLAS and manual contouring

Table 3. *Cont.*

Software Name and Version	Author, Year and Country	Evaluation Results		
		Subjective	**Efficiency**	**Dosimetric**
RayStation v9B/10B-SP1	Bakx et al. (2023), The Netherlands [33]	No/some corrections required for 39.0%/56.0% of OARs and 7.0%/75.0% of CTVs; Unusable contours: CTVp_breast (35.0%), CTVn_L1 (30.0%), CTVn_L4 (25.0%), heart (5.0%) and thyroid (5.0%)	Mean time reduction/patient: 58.2% (manual: 58.6 min VS DLAS: 24.5 min)	NA
RayStation v11B-SP2	Mikalsen et al. (2023), Norway [34]	No/minor corrections required for 85.0%/10.0% of OARs and 8.0%/77.0% of CTVs; Unusable contours: CTVs (6.0%) and OARs (2.0%)	Mean time reduction/patient: 68.0% (manual: 47.2 min VS DLAS: 15.1 min)	CTV coverage (D98 > 95%): 70.0% for breast and 85.0% for lymph nodes; no clinically relevant difference of doses to OARs between DLAS and manual contouring
Siemens Healthineers AG				
AI-Rad Companion Organs RT VA31	Hu et al. (2023), Australia [37]	No/minor OARs corrections required: 87.3%/12.7%; No unusable OARs contours	Mean time reduction/patient: 82.2% (manual: 16.0 min VS DLAS: 2.9 min)	NA
syngo.via RT Image Suite VB40	Pera et al. (2023), Germany and Spain [36]	No/minor OARs corrections required: 75.7%/17.7%; Unusable OARs contours: 0.7%	Mean time reduction/patient: 88.6% (manual: 32.7 min VS DLAS: 3.7 min)	NA
AI-Rad Companion Organs RT VA30	Yamauchi et al. (2024), Japan [38]	Mean score: 3.6 out of 4.0 (indicating no/minor OARs corrections required); No unusable OARs contours	Mean time reduction/patient: 45.4% (manual: 18.6 min VS DLAS: 10.1 min)	NA
Limbus AI Inc, RaySearch Laboratories AB and Therapanacea				
Contour v1.5.0, RayStation v11B and Annotate v1.10.0	Heilemann et al. (2023), Austria [39]	Median score for Limbus Contour/RayStation/Annotate: 3.5/3.0/3.5 out of 4.0 (indicating no/minor OARs corrections required); No unusable OARs contours	NA	No clinically relevant difference of doses to OARs between DLAS and manual contouring

Table 3. *Cont.*

Software Name and Version	Author, Year and Country	Evaluation Results		
		Subjective	**Efficiency**	**Dosimetric**
Mirada Medical Ltd., MVision.ai, Radformation Inc., RaySearch Laboratories AB and Therapanacea				
DLCExpert v2.6.4.47181, Contour+ v1.2.1, AutoContour v1.0.25.0, RayStation v12.0.0.932 and Annotate v1.10.0	Doolan et al. (2023), Cyprus and Germany [29]	NA	Mean time reduction/patient for DLCExpert/Contour+/AutoContour/RayStation/Annotate: 66.0%/92.8%/64.4%/86.0%/93.7% (manual: 22.0 min VS DLAS: 7.5/1.6/7.8/3.1/1.4 min)	NA

Vaassen et al. [28], Bakx et al. [32], Zeverino et al. [65], and Marschner et al. [35] did not conduct any subjective accuracy, efficiency, or dosimetric evaluations. BCS, breast-conserving surgery; CTV, clinical target volume; CTVn_IMN, internal mammary lymph node clinical target volume; CTVn_L1, axillary level 1 lymph node clinical target volume; CTVn_L2, axillary level 2 lymph node clinical target volume; CTVn_L3, axillary level 3 lymph node clinical target volume; CTVn_L4, supraclavicular lymph node clinical target volume; CTVp_breast, breast (primary tumor) clinical target volume; CTVp_chestwall, post-mastectomy chestwall (primary tumor) clinical target volume; D98, dose received by 98% of structure; DSC, Dice similarity coefficient; min, minutes; NA, not available; OAR, organ at risk; VS, versus; ΔV90/95%, % difference of structures segmented by deep learning and manual approaches that received 90/95% of prescribed dose.

Ten (body [36], contralateral breast [26,28,29,31,36–38,65], esophagus-overlapping area [32], heart [26,28,29,31–35,37–39,65], ipsilateral humeral head [29,31–33], left and right lungs [26,28,29,31–39,65], liver [29], sternum and trachea [31]) out of fourteen (71.4%) OARs had the mean/median DSC values of at least 0.70, indicating acceptable for clinical practice (Table 2) [26,29,30,36–39,66]. Using the same DSC cutoff ($\geq$0.70), more than three quarters of the OARs delineated by the following seven software packages were deemed clinically acceptable: Limbus AI Inc. Contour (all (four) OARs (contralateral breast, heart, and left and right lungs)) [26], Mirada Medical Ltd. DLCExpert (seven (contralateral breast, heart, ipsilateral humeral head, left and right lungs, liver and spinal canal) out of nine (77.8%) OARs) [28,29], RaySearch Laboratories AB RayStation (eleven (contralateral breast, esophagus-overlapping area, heart, ipsilateral humeral head, left and right lungs, liver, spinal canal, sternum, thyroid and trachea) out of thirteen (84.6%) OARs) [29,31–34,39,65], Siemens Healthineers AG syngo.via RT Image Suite/AI-Rad Companion Organs RT (five (contralateral breast, esophagus, heart, and left and right lungs) out of six (83.3%) OARs) [35–38], MVision.ai Contour+, Radformation Inc. AutoContour and Therapanacea Annotate (all (eight) OARs (contralateral breast, esophagus, heart, ipsilateral humeral head, left and right lungs, liver and spinal canal)) [29].

About half of the included studies investigated the DLAS for contouring the CTVs with the use of four programs, namely, Manteia Medical Technologies AccuLearning AI [27], Mirada Medical Ltd. DLCExpert [28,29], Radformation Inc. AutoContour [30], and RaySearch Laboratories AB RayStation [31–34] (Table 2). Eight CTVs were covered, collectively [27,28,30–34]. The reported mean/median DSC and/or HD95 for these structures were as follows: CTVp_breast: 0.85–0.95 [27,28,30–34] and 5.3–15.0 mm [27,31–34]; post-mastectomy chestwall (primary tumor) clinical target volume (CTVp_chestwall): 0.71 [30]; axillary level 1 lymph node clinical target volume (CTVn_L1): 0.70–0.80 [30–34] and 9.0–13.6 mm [31–34]; axillary level 2 lymph node clinical target volume (CTVn_L2): 0.66–0.76 [30–34] and 10.0–12.0 mm [31–34]; axillary level 3 lymph node clinical target volume (CTVn_L3): 0.67–0.80 [30–34] and 5.5–8.7 mm [31–34]; supraclavicular lymph node clinical target volume (CTVn_L4): 0.33–0.80 [30–34] and 4.3–16.4 mm [31–34]; internal mammary lymph node clinical target volume (CTVn_IMN): 0.33–0.71 [30,31,34] and 8.0–12.0 mm; and interpectoral lymph node clinical target volume (CTVn_interpect): 0.66–0.68 and 12.0–12.2 mm [31,34]. Although only a little more than one third of the CTVs covered in these studies could be considered clinically acceptable (DSC $\geq$ 0.70), all four software packages were able to provide the satisfactory results for the primary tumors, CTVp_breast, and CTVp_chestwall [27,28,30–34].

Table 3 shows that about three quarters of the studies conducted more clinically meaningful (subjective, efficiency, and/or dosimetric) evaluations for the DLAS programs [26,27,29–31,33,34,36–39]. Except for Bakx et al.'s [33] study on RaySearch Laboratories AB RayStation v9B/10B-SP1, the other seven (87.5%) revealed that a large proportion of the OARs and CTVs delineated by Limbus AI Inc. Contour [39], Manteia Medical Technologies AccuLearning AI [27], RaySearch Laboratories AB RayStation [31,34,39], Siemens Healthineers AG syngo.via RT Image Suite [36] and AI-Rad Companion Organs RT [37,38], and Therapanacea Annotate [39] required no or minor corrections, with unusable contours up to a maximum of 6.0% [27,31,34,36–39]. Hence, this contributed to the notable mean contouring time reduction per patient in the range of 45.4% (Siemens Healthineers AG AI-Rad Companion Organs RT VA30 [38]) to 93.7% (Therapanacea Annotate [29]) for the DLAS. Five studies [26,30,31,34,39] evaluated the dosimetric impact of the DLAS. No study revealed any clinically relevant dose difference (>5%) to the OARs between the DLAS and manual contouring [26,31,34,39,67]. However, three papers demonstrated that the DLAS had notable dosimetric impacts on the CTVs, especially for the lymph nodes [30,31,34].

Table 4 illustrates the characteristics of the included studies. Although all articles were published within the last two years, they had a number of methodological weaknesses [26–39,65]. For example, only two studies (13.3%) collected the data prospectively [26,34], and one fifth (three out of fifteen) were multi-center [31,32,35]. No in-

cluded paper calculated the required sample size for the testing dataset [26–39,65]. This resulted in only tens of patients being involved in 13 out of 15 (86.7%) studies for DLAS program testing [26,27,29–34,36–39,65]. Two (13.3%) papers even had a dataset size below six [26,37]. Nonetheless, external testing was conducted in 12 out of 15 (80.0%) articles [26,28–30,32,34–39,65]. In addition, the datasets used in the included papers were collected from twelve different countries [26–32,34–39,65] on four continents: Europe [26,28,31,32,34–36,39,65], Asia [27,29,38], North America [30], and Australia [37]. The mix of the strengths and the weaknesses of the articles led to their quality scores ranging between 44.0% and 65.0% with the mean and median values of 53.9% and 53.0%, respectively [26–39,65].

Table 4. Characteristics of studies on commercial deep learning-based auto-segmentation (DLAS) software packages for breast cancer radiation therapy planning.

Author, Year and Country	DLAS Architecture	Study Design	Multi-Center	Patient/ Population	Training Dataset		Testing Dataset		Sample Size Calculation	External Testing	Reference Contour Source	Contouring Guidelines	Article Quality (%)
					Source	Size (Number of Patient)	Source	Size (Number of Patient)					
Limbus AI Inc. Contour													
Radici et al. (2022), Italy [26]	U-Net	Prospective	No	L breast cancer patients after BCS	Public: US TCIA and Iranian dataset by Rezaei et al. [68]	At least hundreds	Private: 1 Italian center	3	No	Yes	4 expert ROs	DBCG	53
Manteia Medical Technologies AccuLearning AI													
Hou et al. (2023), China [27]	4 CNN variants (encoder-decoder-based CNN, residual U-Net, U-Net, and V-Net)	Retrospective	No	L and R breast cancer patients after BCS and mastectomy	Private: 1 Chinese center	139	Private: 1 Chinese center	83 (27 L and 26 R BCS, and 16 L and 14 R mastectomy)	No	No	Senior ROs with >8-year experience	NA	65
Mirada Medical Ltd. DLCExpert													
Vaassen et al. (2022), The Netherlands and UK [28]	CNN	Retrospective	No	Breast cancer patients	NA	486	Private: 1 Dutch center	362	No	Yes	All (40) RTTs	NA	63
Radformation Inc. AutoContour													
Tsui et al. (2024), US [30]	NA	Retrospective	No	Breast cancer patients after BCS/mastectomy	NA	NA	Private: 1 US center	66 (34 BCS and 32 mastectomy)	No	Yes	2 ROs with 20- and 30-year experience	NA	58
RaySearch Laboratories AB RayStation													
Almberg et al. (2022), Norway [31]	3D CNN U-Net	Retrospective	Yes	L breast cancer patients after BCS	Private: 2 Norwegian centers	170	Private: 2 Norwegian centers	30	No	No	3 ROs and 3 RTTs with >10-year experience	ESTRO except heart based on Feng et al.'s atlas [69]	58
Bakx et al. (2023), The Netherlands [32]	3D CNN U-Net	Retrospective	Yes	L and R breast cancer patients	Private: 2 Norwegian and 1 Dutch centers for RayStation original and in-house models, respectively	Original model: 170 and in-house model: 160 (80/side)	Private: 1 Dutch center	30	No	Only for original model	ROs and RTTs with final review by 1 experienced RO	ESTRO for CTVs, and Feng et al.'s [69] and Kong et al.'s [70] atlases for OARs	53
Bakx et al. (2023), The Netherlands [33]	3D CNN U-Net	Retrospective	NA	L and R breast cancer patients after BCS	NA	160	NA	20	No	No	ROs and RTTs with final review by 1 experienced RO	ESTRO	56

Table 4. *Cont.*

Author, Year and Country	DLAS Architecture	Study Design	Multi-Center	Patient/ Population	Training Dataset		Testing Dataset		Sample Size Calculation	External Testing	Reference Contour Source	Contouring Guidelines	Article Quality (%)
					Source	Size (Number of Patient)	Source	Size (Number of Patient)					
Mikalsen et al. (2023), Norway [34]	3D CNN U-Net	Prospective	No	L and R breast cancer patients	Private: 2 Norwegian centers	170	Private: 1 Norwegian center	30	No	Yes	2 experienced ROs and 1 RTT	ESTRO except heart based on Feng et al.'s [69] atlas	51
Zeverino et al. (2024), Switzerland [65]	3D CNN U-Net	Retrospective	No	L breast cancer patients	Private: 2 Norwegian centers	170	Private: 1 Swiss center	20	No	Yes	1 senior RO	ESTRO and DBCG	51
Siemens Healthineers AG syngo.via RT Image Suite/AI-Rad Companion Organs RT													
Marschner et al. (2022), Germany and US [35]	U-Net variant	NA	Yes	Breast cancer patients	Private: multi-centers	10,386	Private: 1 German center	237	No	Yes	1 experienced RO	RTOG	56
Hu et al. (2023), Australia [37]	NA	Retrospective	No	Breast cancer patients	NA	NA	Private: 1 Australian center	5	No	Yes	1 RTT with >10-year experience	RTOG	49
Pera et al. (2023), Germany and Spain [36]	U-Net variant	NA	No	L and R breast cancer patients	Private: multi-centers in Asia, Europe, and North and South America	Thousands	Private: 1 Spanish center	30 (15/side)	No	Yes	1 expert RTT with final review by 1 RO	NA	47
Yamauchi et al. (2024), Japan [38]	U-Net variant	Retrospective	No	Breast cancer patients	Private: multi-centers in Europe and America	NA	Private: 1 Japanese center	30 (5 with implants and 5 with mastectomy)	No	Yes	6 expert ROs	RTOG	44
Limbus AI Inc Contour, RaySearch Laboratories AB RayStation and Therapanacea Annotate													
Heilemann et al. (2023)-Austria [39]	NA	Retrospective	No	Breast cancer patients	NA	NA	Private: 1 Austrian center	15	No	Yes	RTTs and 1 RO	NA	51
Mirada Medical Ltd. DLCExpert, MVision.ai Contour+, Radformation Inc. AutoContour, RaySearch Laboratories AB RayStation and Therapanacea Annotate													
Doolan et al. (2023), Cyprus and Germany [29]	NA	Retrospective	No	Bilateral, L and R breast cancer patients	NA	NA	Private: 1 Cypriot center	20 (1 bilaterial, 10 R and 9 L)	No	Yes	3 ROs with >10-year experience	RTOG	53

All studies used computed tomography images for DLAS. 3D, 3-dimensional; BCS, breast-conserving surgery; CNN, convolutional neural network; CTV, clinical target volume; DBCG, Danish Breast Cancer Cooperative Group; ESTRO, European Society of Therapeutic Radiology and Oncology; L, left; NA, not available; OAR, organ at risk; R, right; RO, radiation oncologist; RTOG, Radiation Therapy Oncology Group; RTT, radiation therapist; TCIA, The Cancer Imaging Archive; UK, United Kingdom; US, United States.

4. Discussion

This article is the first systematic review on the performance of the commercial DLAS software for breast cancer RT planning, covering eight software packages [26–39,65]. Hence, it advances the previous four narrative reviews that covered DLAS for breast cancer RT planning [40–43], the NRO Oncology's narrative review on the DLAS without the inclusion of the breast cancer RT [46], and Matoska et al.'s [44] critical review that did not include any commercial DLAS software for this purpose. Another merit of this systematic review paper is that all included studies were published in the last two years, and hence can provide more recent and relevant findings to inform clinical practice (Tables 2–4) [26–39,65].

All included papers reported the mean/median DSC figures to objectively indicate the geometric accuracy of the OARs and/or CTVs delineated by the respective commercial DLAS programs (Table 2) [26–39,65]. Based on the commonly used DSC cutoff of 0.70 [26,29,30,36–39,66], the commercial DLAS programs were able to contour ten OARs (body [36], contralateral breast [26,28,29,31,36–38,65], esophagus-overlapping area [32], heart [26,28,29,31–35,37–39,65], ipsilateral humeral head [29,31–33], left and right lungs [26,28,29,31–39,65], liver [29], sternum and trachea [31]) and three CTVs (CTVp_breast [27,28,30–34], CTVp_chestwall [30] and CTVn_L1 [30–34]) up to the clinically acceptable standard. The performance of the commercial DLAS software for the CTV contouring reported in this review is comparable to the geometric accuracy of the in-house DLAS models reviewed by Matoska et al. [44] that CTVp_breast and CTVp_chestwall contours were deemed satisfactory for clinical use (DSC $\geq$ 0.70), but those for lymph nodes were not acceptable because of their small volumes [30,34]. This also explains the DSC of LAD being lower than 0.70 in some situations [34,65]. The suboptimal DSC values of the OARs, esophagus, spinal canal, and thyroid can be attributed to the difference in the contouring guidelines used by the commercial companies for the software development and individual clinical centers [28,29,32,33].

The DSC was the most common geometric accuracy metric for DLAS because it is a popular loss function for model training, with the goal of achieving the highest DSC value for minimizing loss to attain the greatest DLAS performance [40,46]. The use of DSC also facilitates the comparison of DLAS model performance [26–39,65]. It is noted that every geometric accuracy metric has its strengths and weaknesses, but such discussion is out of the scope of this review, and these details are available from other review articles [40,46,71]. Table 2 shows that all included papers used multiple geometric accuracy metrics to address the limitations of individual ones [26–39,65]. However, it has been criticized that all these geometric metrics are less clinically relevant as the ultimate goal of DLAS implementation is for auto-delineation of the structures that are acceptable by local clinicians and effective for the subsequent treatments, resulting in a reduction of contouring time due to minimal manual corrections required for these contours [26,27,29–31,33,34,36–39,46].

NRO Oncology has suggested that commercial DLAS software should be evaluated by multiple ways including the geometric accuracy, subjective, efficiency and dosimetric evaluations for ensuring that it meets the needs of individual clinical centers [46]. A good example of the required evaluations was shown in Almberg et al.'s study [31] which covered the geometric accuracy for both DLAS and manual contouring as well as any difference between these two, subjective accuracy, efficiency and dosimetric assessments. Nevertheless, Table 3 illustrates that most of the included studies were compliant with the NRO Oncology's suggestion, and their subjective, efficiency and/or dosimetric evaluation results were in line with the corresponding geometric accuracy findings overall (Table 2) [26,27,29–31,33,34,36–39]. These indicate that all eight software packages covered in this review, namely Limbus AI Inc. Contour [26,39], Manteia Medical Technologies AccuLearning AI [27], Mirada Medical Ltd. DLCExpert [29], MVision.ai Contour+, Radformation Inc. AutoContour [29], RaySearch Laboratories AB RayStation [29,31,33,34,39], Siemens Healthineers AG syngo.via RT Image Suite [36] and AI-Rad Companion Organs RT [37,38], and Therapanacea Annotate [29,39] should be useful for clinical practice.

Table 4 demonstrates that the included papers only have quality scores of 44.0–65.0% (mean: 53.9% and median: 53.0%). In 2023, Sivanesan et al. [62] and Bhandari et al. [63] conducted two literature reviews on studies about AI in medical imaging and magnetic resonance imaging with the use of CLAIM to assess their included paper quality. They reported that the median and mean quality scores of their reviewed papers were 57.0% [62] and 47.6% (range: 23.8%–73.8%) [63], respectively, which are in line with the findings of this review. In addition, the issues identified in their included articles which contributed to the low quality scores were similar to the weaknesses of the papers covered in this review, e.g., a lack of sample size calculation, etc. [26–39,62,63,65].

Although NRO Oncology has suggested that every clinical center should conduct its own DLAS software evaluation before clinical implementation due to a variation of clinical protocols (needs) across the centers for RT treatment simulation and planning locally [46], such testing appears particularly crucial for the following three software packages: Manteia Medical Technologies AccuLearning AI [27], Mirada Medical Ltd. DLCExpert [28,29], and MVision.ai Contour+ [29] as per the findings given in Tables 2–4. This is because only Hou et al. [27] and Doolan et al. [29] evaluated Manteia Medical Technologies AccuLearning AI and MVision.ai Contour+, with retrospective collections of 83 and 20 breast cancer cases, respectively. In addition, Manteia Medical Technologies AccuLearning AI was not externally tested [27]. Furthermore, Vaassen et al. [28] merely assessed the geometric accuracy of Mirada Medical Ltd. DLCExpert despite Doolan et al. [29] determining its efficiency based on a retrospective dataset with 20 patients.

For future studies, the commercial DLAS software performance should be assessed based on Almberg et al.'s [31] approach with the external testing dataset size determined through the traditional power calculations suggested by the CLAIM 2024 Update Panel and prospective data collection from multiple centers to ensure that it can provide optimal outcomes in future clinical practice [48,49,64]. If any suboptimal outcome is revealed by the evaluation, the clinical centers should approach the commercial software developer for potential model finetuning with use of their local datasets to improve the DLAS performance due to the aforementioned clinical protocol difference issue [32].

This systematic review has two main limitations. The article selection and the data extraction and synthesis processes were handled by one author with greater than 20 years of experience in conducting literature reviews [47–49,55]. According to a recent systematic review on methodology, this approach is considered appropriate when the single reviewer is experienced [47–49,72]. Furthermore, the use of PRISMA guidelines [50], the data extraction form derived from two narrative reviews (about DLAS for a range of RT plannings [46] including lung cancer RT [56]), one critical review (on DLAS for CTVs of different disease sites [44]), and two systematic reviews (about cervical [57] and prostate cancers [58]), and the CLAIM would further address any potential bias [61,64]. In addition, only papers written in English were covered in this review, potentially affecting its comprehensiveness [47–49,55]. Nonetheless, fifteen articles were included, and that number is greater than the number of the DLAS studies on breast RT covered in Matoska et al.'s [44] critical review. In addition, this review covers the studies from four continents [26–39,65].

5. Conclusions

This systematic review covers eight software packages, namely Limbus AI Inc. Contour, Manteia Medical Technologies AccuLearning AI, Mirada Medical Ltd. DLCExpert, MVision.ai Contour+, Radformation Inc. AutoContour, RaySearch Laboratories AB RayStation, Siemens Healthineers AG syngo.via RT Image Suite and AI-Rad Companion Organs RT, and Therapanacea Annotate, for the DLAS of breast cancer RT planning. Based on the geometric accuracy assessment results reported by the included studies, collectively, these programs can contour ten OARs (body, contralateral breast, esophagus-overlapping area, heart, ipsilateral humeral head, left and right lungs, liver, sternum and trachea) and three CTVs (CTVp_breast, CTVp_chestwall and CTVn_L1) up to the clinically acceptable standard. Their subjective and dosimetric evaluation results match the corresponding

geometric accuracy findings in general. Hence, they can significantly reduce the mean contouring time per patient by 45.4%–93.7% for breast cancer RT planning.

Although NRO Oncology has suggested that every clinical center should conduct its own DLAS software evaluation before clinical implementation due to variations in the clinical protocols across centers for RT treatment simulation and planning locally, such testing appears particularly crucial for Manteia Medical Technologies AccuLearning AI, Mirada Medical Ltd. DLCExpert, and MVision.ai Contour+ as a result of the methodological weaknesses of the corresponding studies such as the use of small datasets collected retrospectively from single centers for the evaluation. For future studies, multiple aspects of the commercial DLAS software's performance should be assessed including their geometric accuracy, subjective, efficiency and dosimetric evaluations with the external testing dataset size determined through the traditional power calculations and the data collected prospectively from multiple centers. It is expected that the findings of this review will support clinical centers in screening commercial DLAS software for further evaluation before implementing it for breast cancer RT planning.

Funding: This work received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Buelens, P.; Willems, S.; Vandewinckele, L.; Crijns, W.; Maes, F.; Weltens, C.G. Clinical evaluation of a deep learning model for segmentation of target volumes in breast cancer radiotherapy. *Radiother. Oncol.* **2022**, *171*, 84–90. [CrossRef]
2. Choi, M.S.; Choi, B.S.; Chung, S.Y.; Kim, N.; Chun, J.; Kim, Y.B.; Chang, J.S.; Kim, J.S. Clinical evaluation of atlas- and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiother. Oncol.* **2020**, *153*, 139–145. [CrossRef] [PubMed]
3. Duma, M.N.; Kulms, T.; Knippen, S.; Teichmann, T.; Wittig, A. Breast clinical target volume: HU-based glandular CTVs and ESTRO CTVs in modern and historical radiotherapy treatment planning. *Strahlenther. Onkol.* **2022**, *198*, 229–235. [CrossRef] [PubMed]
4. Harris, J.R. Fifty years of progress in radiation therapy for breast cancer. *Am. Soc. Clin. Oncol. Educ. Book* **2014**, *34*, 21–25. [CrossRef]
5. Chang, J.S.; Chang, J.H.; Kim, N.; Kim, Y.B.; Shin, K.H.; 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. [CrossRef]
6. Milo, M.L.H.; Nyeng, T.B.; Lorenzen, E.L.; Hoffmann, L.; Møller, D.S.; Offersen, B.V. Atlas-based auto-segmentation for delineating the heart and cardiac substructures in breast cancer radiation therapy. *Acta Oncol.* **2022**, *61*, 247–254. [CrossRef]
7. Byun, H.K.; Chang, J.S.; Choi, M.S.; Chun, J.; Jung, J.; Jeong, C.; Kim, J.S.; Chang, Y.; Chung, S.Y.; Lee, S.; et al. Evaluation of deep learning-based autosegmentation in breast cancer radiotherapy. *Radiat. Oncol.* **2021**, *16*, 203. [CrossRef]
8. Chung, S.Y.; Chang, J.S.; Choi, M.S.; Chang, Y.; Choi, B.S.; Chun, J.; Keum, K.C.; Kim, J.S.; Kim, Y.B. Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiat. Oncol.* **2021**, *16*, 44. [CrossRef] [PubMed]
9. Poortmans, P.M.P.; Takanen, S.; Marta, G.N.; Meattini, I.; Kaidar-Person, O. Winter is over: The use of artificial intelligence to individualise radiation therapy for breast cancer. *Breast* **2020**, *49*, 194–200. [CrossRef]
10. van Pelt, V.W.J.; Gerrets, S.; Simões, R.; Elkhuizen, P.H.M.; Janssen, T.M. Evaluation of delineating the target volume by radiation therapists in breast cancer patients. *Tech. Innov. Patient Support Radiat. Oncol.* **2021**, *17*, 78–81. [CrossRef]
11. Liu, Z.; Liu, F.; Chen, W.; Tao, Y.; Liu, X.; Zhang, F.; Shen, J.; Guan, H.; Zhen, H.; Wang, S.; 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*, 8209–8217. [CrossRef] [PubMed]
12. Im, J.H.; Lee, I.J.; Choi, Y.; Sung, J.; Ha, J.S.; Lee, H. Impact of denoising on deep-learning-based automatic segmentation framework for breast cancer radiotherapy planning. *Cancers* **2022**, *14*, 3581. [CrossRef]
13. Jung, J.W.; Mille, M.M.; Ky, B.; Kenworthy, W.; Lee, C.; Yeom, Y.S.; Kwag, A.; Bosch, W.; MacDonald, S.; Cahlon, O.; et al. Application of an automatic segmentation method for evaluating cardiac structure doses received by breast radiotherapy patients. *Phys. Imaging Radiat. Oncol.* **2021**, *19*, 138–144. [CrossRef]
14. Xie, X.; Song, Y.; Ye, F.; Yan, H.; Wang, S.; Zhao, X.; Dai, J. Prior information guided auto-contouring of breast gland for deformable image registration in postoperative breast cancer radiotherapy. *Quant. Imaging Med. Surg.* **2021**, *11*, 4721–4730. [CrossRef]
15. Zhou, H.; Li, Y.; Gu, Y.; Shen, Z.; Zhu, X.; Ge, Y. A deep learning based automatic segmentation approach for anatomical structures in intensity modulation radiotherapy. *Math. Biosci. Eng.* **2021**, *18*, 7506–7524. [CrossRef]

16. Zeleznik, R.; Weiss, J.; Taron, J.; Guthier, C.; Bitterman, D.S.; Hancox, C.; Kann, B.H.; Kim, D.W.; Punglia, R.S.; Bredfeldt, J.; et al. Deep-learning system to improve the quality and efficiency of volumetric heart segmentation for breast cancer. *NPJ Digit. Med.* **2021**, *4*, 43. [CrossRef] [PubMed]
17. Saha, M.; Jung, J.W.; Lee, S.W.; Lee, C.; Lee, C.; Mille, M.M. A deep learning segmentation method to assess dose to organs at risk during breast radiotherapy. *Phys. Imaging Radiat. Oncol.* **2023**, *28*, 100520. [CrossRef] [PubMed]
18. Zhong, Y.; Guo, Y.; Fang, Y.; Wu, Z.; Wang, J.; Hu, W. Geometric and dosimetric evaluation of deep learning based auto-segmentation for clinical target volume on breast cancer. *J. Appl. Clin. Med. Phys.* **2023**, *24*, e13951. [CrossRef]
19. Kazemimoghadam, M.; Yang, Z.; Chen, M.; Rahimi, A.; Kim, N.; Alluri, P.; Nwachukwu, C.; Lu, W.; Gu, X. A deep learning approach for automatic delineation of clinical target volume in stereotactic partial breast irradiation (S-PBI). *Phys. Med. Biol.* **2023**, *68*, 105011. [CrossRef]
20. Colbert, Z.M.; Ramachandran, P. Auto-segmentation of thoracic organs in CT scans of breast cancer patients using a 3D U-net cascaded into 2D patchGANs. *Biomed. Phys. Eng. Express.* **2023**, *9*, 055011. [CrossRef]
21. Choi, M.S.; Chang, J.S.; Kim, K.; Kim, J.H.; Kim, T.H.; Kim, S.; Cha, H.; Cho, O.; Choi, J.H.; Kim, M.; et al. Assessment of deep learning-based auto-contouring on interobserver consistency in target volume and organs-at-risk delineation for breast cancer: Implications for RTQA program in a multi-institutional study. *Breast* **2024**, *73*, 103599. [CrossRef]
22. Sun, Z.; Ng, C.K.C. Artificial intelligence (enhanced super-resolution generative adversarial network) for calcium deblooming in coronary computed tomography angiography: A feasibility study. *Diagnostics* **2022**, *12*, 991. [CrossRef] [PubMed]
23. Ng, C.K.C.; Leung, V.W.S.; Hung, R.H.M. Clinical evaluation of deep learning and atlas-based auto-contouring for head and neck radiation therapy. *Appl. Sci.* **2022**, *12*, 11681. [CrossRef]
24. Sun, Z.; Ng, C.K.C. Finetuned super-resolution generative adversarial network (artificial intelligence) model for calcium deblooming in coronary computed tomography angiography. *J. Pers. Med.* **2022**, *12*, 1354. [CrossRef] [PubMed]
25. Leung, V.W.S.; Ng, C.K.C.; Lam, S.K.; Wong, P.T.; Ng, K.Y.; Tam, C.H.; Lee, T.C.; Chow, K.C.; Chow, Y.K.; Tam, V.C.W.; et al. Computed tomography-based radiomics for long-term prognostication of high-risk localized prostate cancer patients received whole pelvic radiotherapy. *J. Pers. Med.* **2023**, *13*, 1643. [CrossRef] [PubMed]
26. Radici, L.; Ferrario, S.; Borca, V.C.; Cante, D.; Paolini, M.; Piva, C.; Baratto, L.; Franco, P.; La Porta, M.R. Implementation of a commercial deep learning-based auto segmentation software in radiotherapy: Evaluation of effectiveness and impact on workflow. *Life* **2022**, *12*, 2088. [CrossRef]
27. Hou, Z.; Gao, S.; Liu, J.; Yin, Y.; Zhang, L.; Han, Y.; Yan, J.; Li, S. Clinical evaluation of deep learning-based automatic clinical target volume segmentation: A single-institution multi-site tumor experience. *Radiol. Med.* **2023**, *128*, 1250–1261. [CrossRef] [PubMed]
28. Vaassen, F.; Boukerroui, D.; Looney, P.; Canters, R.; Verhoeven, K.; Peeters, S.; Lubken, I.; Mannens, J.; Gooding, M.J.; van Elmpt, W. Real-world analysis of manual editing of deep learning contouring in the thorax region. *Phys. Imaging Radiat. Oncol.* **2022**, *22*, 104–110. [CrossRef] [PubMed]
29. Doolan, P.J.; Charalambous, S.; Roussakis, Y.; Leczynski, A.; Peratikou, M.; Benjamin, M.; Ferentinos, K.; Strouthos, I.; Zamboglou, C.; Karagiannis, E. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front. Oncol.* **2023**, *13*, 1213068. [CrossRef] [PubMed]
30. Tsui, T.; Podgorsak, A.; Roeske, J.C.; Small, W., Jr.; Refaat, T.; Kang, H. Geometric and dosimetric evaluation for breast and regional nodal auto-segmentation structures. *J. Appl. Clin. Med. Phys.* **2024**, *25*, e14461. [CrossRef]
31. Almberg, S.S.; Lervåg, C.; Frengen, J.; Eidem, M.; Abramova, T.M.; Nordstrand, C.S.; Alsaker, M.D.; Tøndel, H.; Raj, S.X.; Wanderås, A.D. Training, validation, and clinical implementation of a deep-learning segmentation model for radiotherapy of loco-regional breast cancer. *Radiother. Oncol.* **2022**, *173*, 62–68. [CrossRef] [PubMed]
32. Bakx, N.; van der Sangen, M.; Theuws, J.; Bluemink, H.; Hurkmans, C. Comparison of the output of a deep learning segmentation model for locoregional breast cancer radiotherapy trained on 2 different datasets. *Tech. Innov. Patient Support Radiat. Oncol.* **2023**, *26*, 100209. [CrossRef] [PubMed]
33. Bakx, N.; Rijkaart, D.; van der Sangen, M.; Theuws, J.; van der Toorn, P.P.; Verrijssen, A.S.; van der Leer, J.; Mutsaers, J.; van Nunen, T.; Reinders, M.; 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. [CrossRef]
34. Mikalsen, S.G.; Skjøtskift, T.; Flote, V.G.; Hämäläinen, N.P.; Heydari, M.; Rydén-Eilertsen, K. Extensive clinical testing of deep learning segmentation models for thorax and breast cancer radiotherapy planning. *Acta Oncol.* **2023**, *62*, 1184–1193. [CrossRef] [PubMed]
35. Marschner, S.; Datar, M.; Gaasch, A.; Xu, Z.; Grbic, S.; Chabin, G.; Geiger, B.; Rosenman, J.; Corradini, S.; Niyazi, M.; et al. A deep image-to-image network organ segmentation algorithm for radiation treatment planning: Principles and evaluation. *Radiat. Oncol.* **2022**, *17*, 129. [CrossRef] [PubMed]
36. Pera, Ó.; Martínez, Á.; Möhler, C.; Hamans, B.; Vega, F.; Barral, F.; Becerra, N.; Jimenez, R.; Fernandez-Velilla, E.; Quera, J.; et al. Clinical validation of Siemens' syngo.via automatic contouring system. *Adv. Radiat. Oncol.* **2023**, *8*, 101177. [CrossRef]
37. Hu, Y.; Nguyen, H.; Smith, C.; Chen, T.; Byrne, M.; Archibald-Heeren, B.; Rijken, J.; Aland, T. Clinical assessment of a novel machine-learning automated contouring tool for radiotherapy planning. *J. Appl. Clin. Med. Phys.* **2023**, *24*, e13949. [CrossRef]
38. Yamauchi, R.; Itazawa, T.; Kobayashi, T.; Kashiyama, S.; Akimoto, H.; Mizuno, N.; Kawamori, J. Clinical evaluation of deep learning and atlas-based auto-segmentation for organs at risk delineation. *Med. Dosim.* **2024**, *49*, 167–176. [CrossRef] [PubMed]

39. Heilemann, G.; Buschmann, M.; Lechner, W.; Dick, V.; Eckert, F.; Heilmann, M.; Herrmann, H.; Moll, M.; Knoth, J.; Konrad, S.; et al. Clinical implementation and evaluation of auto-segmentation tools for multi-site contouring in radiotherapy. *Phys. Imaging Radiat. Oncol.* **2023**, *28*, 100515. [CrossRef]
40. Lin, H.; Xiao, H.; Dong, L.; Teo, K.B.; Zou, W.; Cai, J.; Li, T. Deep learning for automatic target volume segmentation in radiation therapy: A review. *Quant. Imaging Med. Surg.* **2021**, *11*, 4847–4858. [CrossRef]
41. Isaksson, L.J.; Summers, P.; Mastroleo, F.; Marvaso, G.; Corrao, G.; Vincini, M.G.; Zaffaroni, M.; Ceci, F.; Petralia, G.; Orecchia, R.; et al. Automatic segmentation with deep learning in radiotherapy. *Cancers* **2023**, *15*, 4389. [CrossRef] [PubMed]
42. Samarasinghe, G.; Jameson, M.; Vinod, S.; Field, M.; Dowling, J.; Sowmya, A.; Holloway, L. Deep learning for segmentation in radiation therapy planning: A review. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 578–595. [CrossRef] [PubMed]
43. Shen, J.; Gu, P.; Wang, Y.; Wang, Z. Advances in automatic delineation of target volume and cardiac substructure in breast cancer radiotherapy (review). *Oncol. Lett.* **2023**, *25*, 110. [CrossRef] [PubMed]
44. 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. [CrossRef]
45. Curran, W.J., Jr.; DiSaia, P.J.; Wolmark, N. NRG Oncology research opportunities within the new National Clinical Trials Network. *Semin. Oncol.* **2014**, *41*, 553–555. [CrossRef]
46. Rong, Y.; Chen, Q.; Fu, Y.; Yang, X.; Al-Hallaq, H.A.; Wu, Q.J.; Yuan, L.; Xiao, Y.; Cai, B.; Latifi, K.; et al. NRG Oncology assessment of artificial intelligence deep learning-based auto-segmentation for radiation therapy: Current developments, clinical considerations, and future directions. *Int. J. Radiat. Oncol. Biol. Phys.* **2024**, *119*, 261–280. [CrossRef] [PubMed]
47. Ng, C.K.C. Artificial intelligence for radiation dose optimization in pediatric radiology: A systematic review. *Children* **2022**, *9*, 1044. [CrossRef]
48. Ng, C.K.C. Diagnostic performance of artificial intelligence-based computer-aided detection and diagnosis in pediatric radiology: A systematic review. *Children* **2023**, *10*, 525. [CrossRef]
49. Ng, C.K.C. Generative adversarial network (generative artificial intelligence) in pediatric radiology: A systematic review. *Children* **2023**, *10*, 1372. [CrossRef] [PubMed]
50. PRISMA Statement. Available online: https://www.prisma-statement.org/ (accessed on 24 October 2024).
51. Mridha, M.F.; Hamid, M.A.; Monowar, M.M.; Keya, A.J.; Ohi, A.Q.; Islam, M.R.; Kim, J.M. A comprehensive survey on deep-learning-based breast cancer diagnosis. *Cancers* **2021**, *13*, 6116. [CrossRef] [PubMed]
52. Aggarwal, R.; Sounderajah, V.; Martin, G.; Ting, D.S.W.; Karthikesalingam, A.; King, D.; Ashrafian, H.; Darzi, A. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digit. Med.* **2021**, *4*, 65. [CrossRef] [PubMed]
53. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
54. Shah, K.A.; Ng, C.K.C. Workplace violence in medical radiation science: A systematic review. *Radiography* **2024**, *30*, 440–447. [CrossRef]
55. Ng, C.K.C. A review of the impact of the COVID-19 pandemic on pre-registration medical radiation science education. *Radiography* **2022**, *28*, 222–231. [CrossRef]
56. Liu, X.; Li, K.W.; Yang, R.; Geng, L.S. Review of deep learning based automatic segmentation for lung cancer radiotherapy. *Front. Oncol.* **2021**, *11*, 717039. [CrossRef]
57. Yang, C.; Qin, L.H.; Xie, Y.E.; Liao, J.Y. Deep learning in CT image segmentation of cervical cancer: A systematic review and meta-analysis. *Radiat. Oncol.* **2022**, *17*, 175. [CrossRef] [PubMed]
58. Fassia, M.K.; Balasubramanian, A.; Woo, S.; Vargas, H.A.; Hricak, H.; Konukoglu, E.; Becker, A.S. Deep learning prostate MRI segmentation accuracy and robustness: A systematic review. *Radiol. Artif. Intell.* **2024**, *6*, e230138. [CrossRef]
59. Vasey, B.; Ursprung, S.; Beddoe, B.; Taylor, E.H.; Marlow, N.; Bilbro, N.; Watkinson, P.; McCulloch, P. Association of clinician diagnostic performance with machine learning-based decision support systems: A systematic review. *JAMA Netw. Open.* **2021**, *4*, e211276. [CrossRef]
60. Imrey, P.B. Limitations of meta-analyses of studies with high heterogeneity. *JAMA Netw. Open.* **2020**, *3*, e1919325. [CrossRef] [PubMed]
61. Mongan, J.; Moy, L.; Kahn, C.E., Jr. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. [CrossRef]
62. Sivanesan, U.; Wu, K.; McInnes, M.D.F.; Dhindsa, K.; Salehi, F.; van der Pol, C.B. Checklist for artificial intelligence in medical imaging reporting adherence in peer-reviewed and preprint manuscripts with the highest Altmetric Attention Scores: A meta-research study. *Can. Assoc. Radiol. J.* **2023**, *74*, 334–342. [CrossRef]
63. Bhandari, A.; Scott, L.; Weilbach, M.; Marwah, R.; Lasocki, A. Assessment of artificial intelligence (AI) reporting methodology in glioma MRI studies using the Checklist for AI in Medical Imaging (CLAIM). *Neuroradiology* **2023**, *65*, 907–913. [CrossRef]
64. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, C.E., Jr.; CLAIM 2024 Update Panel. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. [CrossRef]
65. Zeverino, M.; Piccolo, C.; Marguet, M.; Jeanneret-Sozzi, W.; Bourhis, J.; Bochud, F.; Moeckli, R. Sensitivity of automated and manual treatment planning approaches to contouring variation in early-breast cancer treatment. *Phys. Med.* **2024**, *123*, 103402. [CrossRef] [PubMed]

66. van Dijk, L.V.; Van den Bosch, L.; Aljabar, P.; Peressutti, D.; Both, S.; Steenbakkers, R.J.H.M.; Langendijk, J.A.; Gooding, M.J.; Brouwer, C.L. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiother. Oncol.* **2020**, *142*, 115–123. [CrossRef] [PubMed]
67. Wong, J.Y.K.; Leung, V.W.S.; Hung, R.H.M.; Ng, C.K.C. Comparative study of Eclipse and RayStation multi-criteria optimization-based prostate radiotherapy treatment planning quality. *Diagnostics* **2024**, *14*, 465. [CrossRef]
68. Rezaei, M.; Mohammadbeigi, A.; Khoshgard, K.; Haghparast, A. CT images and radiotherapy treatment planning of patients with breast cancer: A dataset. *Data Brief* **2017**, *13*, 390–395. [CrossRef] [PubMed]
69. Feng, M.; Moran, J.M.; Koelling, T.; Chughtai, A.; Chan, J.L.; Freedman, L.; Hayman, J.A.; Jagsi, R.; Jolly, S.; Larouere, J.; et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *79*, 10–18. [CrossRef]
70. Kong, F.M.; Ritter, T.; Quint, D.J.; Senan, S.; Gaspar, L.E.; Komaki, R.U.; Hurkmans, C.W.; Timmerman, R.; Bezjak, A.; Bradley, J.D.; et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *81*, 1442–1457. [CrossRef] [PubMed]
71. 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.* **2023**, *35*, 354–369. [CrossRef] [PubMed]
72. Waffenschmidt, S.; Knelangen, M.; Sieben, W.; Bühn, S.; Pieper, D. Single screening versus conventional double screening for study selection in systematic reviews: A methodological systematic review. *BMC Med. Res. Methodol.* **2019**, *19*, 132. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.