Physics and Imaging in Radiation Oncology 33 (2025) 100731

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Review Article

A systematic review of the role of artificial intelligence in automating computed tomography-based adaptive radiotherapy for head and neck cancer

Edoardo Mastella [a,1,*], Francesca Calderoni [a,1], Luigi Manco [a,b], Martina Ferioli [c], Serena Medoro [c], Alessandro Turra [a], Melchiore Giganti [d], Antonio Stefanelli [c]

[a] Medical Physics Unit, University Hospital of Ferrara I-44124 Cona, Ferrara, Italy
[b] Medical Physics Unit, Azienda USL di Ferrara I-44121 Ferrara, Italy
[c] Radiation Oncology Unit, University Hospital of Ferrara I-44124 Cona, Ferrara, Italy
[d] University Radiology Unit, University of Ferrara I-44121 Ferrara, Italy

ARTICLE INFO

Keywords:
Machine learning
Deep learning
CBCT
Image quality
Automated contouring
Predictive model

ABSTRACT

Purpose: Adaptive radiotherapy (ART) may improve treatment quality by monitoring variations in patient anatomy and incorporating them into the treatment plan. This systematic review investigated the role of artificial intelligence (AI) in computed tomography (CT)-based ART for head and neck (H&N) cancer.
Methods: A comprehensive search of main electronic databases was conducted until April 2024. Titles and abstracts were reviewed to evaluate the compliance with inclusion criteria: CT-based imaging for photon ART of H&N patients and AI applications. 17 original retrospective studies with samples sizes ranging from 37 to 239 patients were included. The quality of the studies was evaluated with the Quality Assessment of Diagnostic Accuracy Studies-2 and the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) tools. Key metrics were examined to evaluate the performances of the proposed AI-methods.
Results: Overall, the risk of bias was low. The average CLAIM score was 70%. A major finding was that generated synthetic CTs improved similarity metrics with planning CT compared to original cone-beam CTs, with average mean absolute error up to 39 HU and maximum improvement of 80%. Auto-segmentation provided an efficient and accurate option for organ-at-risk delineation, with average Dice similarity coefficient ranging from 80 to 87%. Finally, AI models could be trained using clinical and radiomic features to predict the effectiveness of ART with accuracy above 80%.
Conclusions: Automation of processes in ART for H&N cancer is very promising throughout the entire chain, from the generation of synthetic CTs and auto-segmentation to predict the effectiveness of ART.

1. Introduction

In the last decades, several technical innovations have been introduced into the clinical practice of radiotherapy (RT), increasing treatment efficacy while at the same time reducing patient toxicity [1]. The development of highly conformal techniques, e.g. intensity modulated (IM)RT, volumetric-modulated arc therapy (VMAT), stereotactic radiosurgery (SRS) or stereotactic body (SB)RT, together with advances in treatment planning, on-board imaging, motion management, etc. allowed for accurate delivery of dose distributions with tight margins and steep dose gradients [2,3], thus improving the therapeutic window of RT treatments.

A drawback of the increased complexity of these RT techniques is a greater sensitivity to treatment uncertainties. Therefore, rigorous procedures need to be implemented to improve patient setup accuracy and to monitor possible variations in patient geometry and anatomy. Indeed, such variations may lead to severe degradations of the planned dose distributions, with suboptimal target coverage and/or unintended dose to organs at risk (OARs). Nowadays, image-guided (IG)RT is widely integrated in several steps of the clinical workflow, from treatment

* Corresponding author at: Struttura Complessa di Fisica Medica, Azienda Ospedaliero Universitaria di Ferrara, via A. Moro 8, I-44124 Cona, Italy.
E-mail address: edoardo.mastella@ospfe.it (E. Mastella).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.phro.2025.100731
Received 7 October 2024; Received in revised form 10 January 2025; Accepted 12 February 2025
Available online 14 February 2025
2405-6316/© 2025 The Author(s). Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

simulation to planning and delivery. This integration has dramatically increased the ability to detect the above-mentioned variations during the course of treatment in different anatomical regions, e.g. head and neck (H&N), lung, pelvis, etc. [4]. Although the most advanced technologies have already been introduced into clinical practice, e.g. hybrid magnetic resonance (MR)- or positron emission tomography (PET)-guided linear accelerators (linacs) [5,6], currently computed tomography (CT) guidance is the most commonly used technology. Several CT-based technologies have been commercially available for several years now, including kilovoltage (kV) and megavoltage (MV) modalities with both fan- and cone-beam (CB)CT [7].

The concept of adaptive radiation therapy (ART) was introduced by Yan et al. [8] in the late 1990 s. The main purpose of ART is to improve treatment quality "by systematically monitoring treatment variations and incorporating them to re-optimize the treatment plan early on during the course of treatment". To this purpose, frequent imaging is used to evaluate anatomical differences that may occur during the course of treatment. Re-planning is considered once significant changes in patient anatomy and/or pathology are observed, leading to suboptimal dose distributions. Adapting the treatment can further improve the therapeutic ratio [4].

The treatment of H&N cancer remains one of the most challenging in RT for various reasons: the high prescription doses required for tumor control [9,10]; the intricate anatomy of the H&N district, with critical structures often in close proximity to target volumes and so irradiated with doses close to tolerance, thus requiring individual prescriptions to balance between the risk of complications and of local recurrence [10–12]; the high risk of malnutrition and weight loss during the treatment course [13]. Modern RT techniques reduced the risk of severe normal tissue toxicities, such as dysphagia, xerostomia or neurological complications [12,14,15], thus improving patients' quality of life. Several studies also reported the benefits of adapting treatment plans to variations of patient anatomy for H&N cancer [16–23].

However, the success of ART depends on a crucial factor: fast plan adaptation. The ART process requires time-consuming procedures throughout the entire RT chain, resulting in additional workload for the staff. Therefore, in real-world clinical practice it may not be feasible for all patients who would benefit from it. In this context, the impact of artificial intelligence (AI) has the potential to mitigate the pressure on the staff by automating repetitive tasks.

Analyzing recent literature, several reviews explored the potential of ART and/or AI in RT with a particular concern on H&N cancer [24–27], with ART being included as one of five fundamental themes within a comprehensive review of AI applications in H&N cancer [28]. Although these papers provide valuable insights into different aspects of the workflow, such as the benefits and pitfalls of ART [26] or the potential role of AI in H&N ART [27], they do not specifically focus on the role of AI in automating the entire ART chain, nor do they concurrently investigate all aspects of the AI-based ART workflow.

Thus, the purpose of this review was to systematically evaluate the current literature landscape to concurrently investigate various parts of the AI-based ART workflow. We aimed to summarize the most relevant research findings related to the different tasks. Therefore, we explored the various CT-based imaging techniques used in ART and the different AI strategies that have been developed to address the challenges associated with plan adaptation. Finally, we discussed the potential future directions of AI-based ART research. Hence, this review bridges the gap in our understanding of the role of AI in the development of automated ART workflows for H&N cancer.

2. Materials and methods

2.1. PICOS framework

The PICOS framework was used to answer the key questions for this study "What is the potential of AI in predict the need of plan adaptation and in automating repetitive tasks for CT-based ART in H&N cancer?" "How effective are the proposed AI-methods compared to conventional and/or state-of-the-art AI methods?": (1) Population – patients with H&N cancer undergoing conventional CT-based photon RT, (2) Interventions – AI-driven automation in ART including prediction of the need for plan adaptation, (3) Comparisons – conventional methods, e.g. manual planning, state-of-the-art AI methods, (4) Outcomes – (primary) accuracy and efficiency of the proposed AI methods, (secondary) improvements with respect to other methods (5) Study design – retrospective or prospective studies in humans.

2.2. Literature search strategy

A search on the four authoritative databases and online sources (PubMed/Medline, Web of Science, Embase, Scopus) was performed running the following multiple queries: "adaptive AND (radiation therapy OR radiotherapy) AND (cone OR fan beam computed tomography OR CBCT OR megavoltage CT OR MVCT) AND (automation OR radiomic OR predictive OR prediction model OR texture analysis OR deep learning OR machine learning OR artificial intelligence) AND head and neck NOT review".

Research was performed on 30th April 2024 and articles published from 2018 were considered.

The review was carried out using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29], nevertheless it was not registered and a protocol was not published.

2.3. Study selection, data collection and quality assessment

Results were filtered through Excel spreadsheets after conversion from database outputs. For the subsequent analysis, data were summarized with the following fields: first author, journal, year, title, exclusion/issues, imaging modality, dataset type, anatomical site, purpose of the study. To evaluate study inclusion, titles and abstracts were independently reviewed by three medical physicists with at least 5 years clinical experience (EM, LM and FC). The following papers were considered ineligible: articles not in English language and studies not within the field of interest.

The literature search identified a total of 270 studies, reduced to 241 after excluding duplicates. 137 records were removed as non-articles: case report, abstracts, review or *meta*-analysis.

Full articles were retrieved when the abstract was considered relevant according to the PICOS criteria. Any discrepancies were collegially discussed to reach a consensus. An experienced radiation oncologist (MF) was consulted if necessary. Fig. 1 shows the PRISMA flowchart of literature search steps with the inclusion/exclusion criteria. All the included studies were retrospective.

The papers were then divided into three main categories according to the context explored and/or the type of study as follows: synthetic CT (6/17), auto-segmentation (5/17) and radiomics and predictive models (6/17).

In Supplementary Fig. 1 the main steps of the ART chain are depicted highlighting the integration of these three categories of AI applications.

A careful assessment of the quality of each included study was performed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [30] and the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [31] tools, similarly to [32]. The assessments were conducted independently by EM, LM, FC and MF. Supervisors were consulted in case of discrepancies or different interpretations.

Moreover, the radiomic quality score "RQS 2.0" [33,34] was blindly computed by two of the authors (LM, MF) with at least 3 years of experience in radiomics.

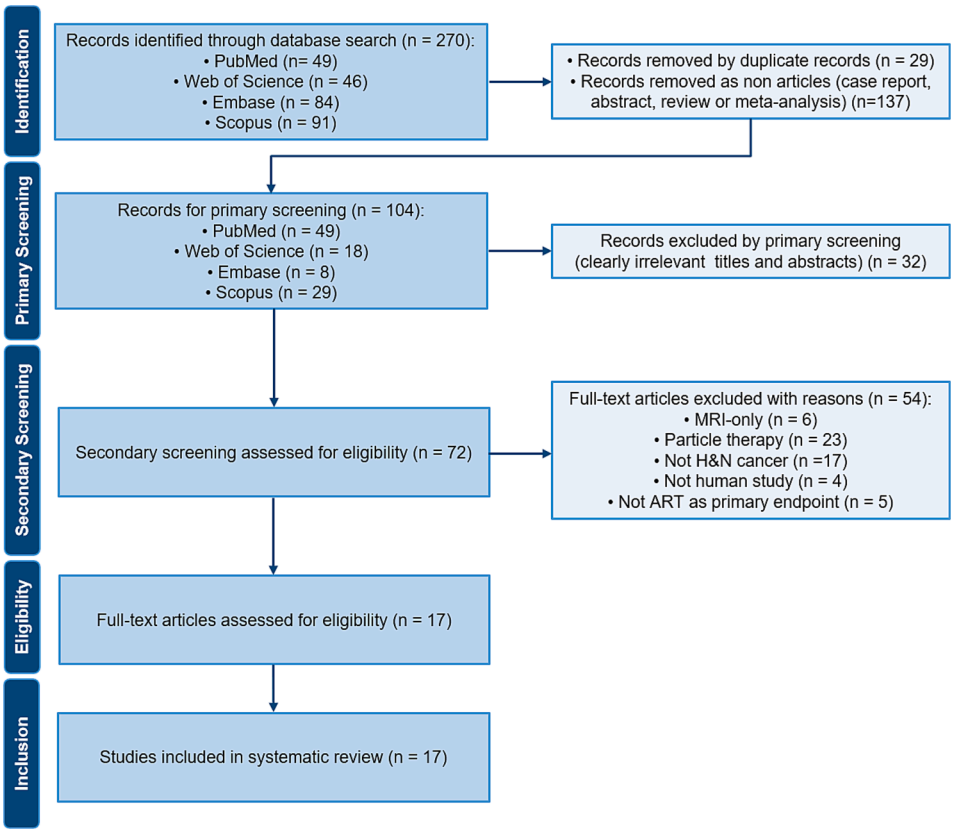

Fig. 1. PRISMA flowchart of study selection with inclusion/exclusion criteria of this systematic review.

3. Results

3.1. Overall quality assessment

Overall, the risk of bias was low except for the patient selection domain, as most of the studies did not describe how they selected the dataset, e.g. consecutive or random sample of patients, resulting in an unclear risk of bias. The mean CLAIM score (±1 standard deviation) was (70 ± 5)%, ranging from 61 % to 78 %. The results section achieved a low score (38 %) as the clinical/technical characteristics of the dataset were missing in most of the papers, while the performance of the proposed model against current standards was almost always reported. The

Table 1
Summary of 6 studies included in the synthetic CT section. The main results of the most commonly used evaluation metrics are reported.

Author	Journal (Year)	DL Model	AI SW	Imaging	No. of patients (training/ validation/test)	Evaluation metric				
						Image similarity			Dose similarity	
						MAE [HU]	PSNR [dB]	SSIM	DVH [%]	GPR (2 %/2mm) [%]
Lv T et al.[41]	Med Phys (2024)	cycleGAN	TensorFlow	MVCT, pCT, skVCT	41 (25/-/16)	60.6	34.0	0.97	no significant differences ($p > 0.05$)	99.7
Yoganathan SA et al.[37]	Phys Imaging RadiatOncol (2023)	ResUNet, UNet	MATLAB	CBCT, pCT, sCT	93 (58/10/25)	46.0	66.6	0.81	0.2	99.8
Li Z et al.[40]	BMC Cancer (2023)	RegGAN	PyTorch	CBCT, pCT, sCT	140 (80/20/40)	45.6	26.8	0.85	< 3.0	–
Aouadi S et al. [38]	Biomed Phys EngExpres (2023)	UNet, cGAN	Keras and TensorFlow	CBCT, pCT, sCT	93 (58/10/25)	67.9	36.9	0.95	< 0.6 (PTVs)< 1.2 (OARs)	99.5
O'Hara CJ et al. [39]	J Appl Clin Med Phys (2022)	cycleGAN	RayStation	CBCT, pCT, rCT, dCBCT, sCT	60 (35/10/15)	79.4	–	–	≤ 0.1 (PTVs)≤ 0.7 (OARs)	100.0
Yuan N et al. [36]	Phys Med Biol (2020)	UNet	Keras and TensorFlow	CBCT, pCT, eCBCT	55 (40/5/10)	49.3	14.3*	0.85	–	–

Abbreviations: AI SW = artificial intelligence software; CBCT = cone-beam computed tomography; cGAN = conditional generative adversarial network; cycleGAN = cycle general adversarial network; dCBCT = deformably registered CBCT; DL = deep learning; DVH = dose volume histogram; eCBCT = enhanced CBCT; GPR = gamma passing rate; MAE = mean absolute error; MVCT = megavoltage computed tomography; pCT = planning CT; PSNR = peak-signal-to-noise ratio; rCT = rescan CT; RegGAN = registration generative adversarial network; ResUNet = self-attention-residual-UNet; sCT = synthetic CT; skVCT = synthetic kilovoltage computed tomography; SSIM = structural similarity. Note: *Signal-to-noise ratio.

other information section also scored low (31 %) because none of the model weights and data were deposited into a publicly accessible repository. Supplementary Fig. 2 shows the overall quality assessment of the included studies conducted using the QUADAS-2 and the CLAIM tools.

3.2. Synthetic CT

Different DL-based methods were applied to image synthesis from CBCTs/MVCTs highlighting the great improvement of similarity with planning CTs (pCTs). Table 1 summarizes the main characteristics of 6 studies concerning synthetic CT (sCT) and the results relative to the most common evaluation metrics. In regards to image similarity, mean absolute error (MAE), (peak) signal-to-noise ratio ((p)SNR) and structural similarity (SSIM)[35] are reported. The upper panel of Fig. 2 analyzes MAE metric, which is present in all studies, depicting mean values among the considered networks, dimensionality of models or accelerators. The mean improvements relative to CBCT or MVCT are also represented. Moreover, dose volume histogram (DVH) values, possibly divided among planning target volumes (PTVs) and OARs, and gamma passing rate (GPR) were used for dose accuracy evaluation. GPRs with tolerance criteria of 2 %/2 mm and their relative improvement from CBCT or MVCT are displayed in the lower panel of Fig. 2.

Enhanced CBCTs (eCBCTs) showed significant improvement in image similarity metrics, nevertheless different sizes of the datasets can strongly affect the results. Convolutional neural network (CNN) methodology with UNet architecture [36] allowed improvement in MAE, SNR and SSIM, up to 49.3 HU, 14.3 dB and 0.85 respectively, compared to original values of 172.7 HU, 8.3 dB and 0.42, but suboptimal results were obtained with 30 datasets.

Also a novel DL model, called “self-attention-residual-UNet” (ResUNet) [37] demonstrated both a remarkable 11 % enhancement in the image quality over standard UNet for non-adversarial training with 2.5D model and a pronounced superiority when dealing with limited training data, from 100 % to 25 %. On the other hand, DVH deviations remained below 0.5 %.

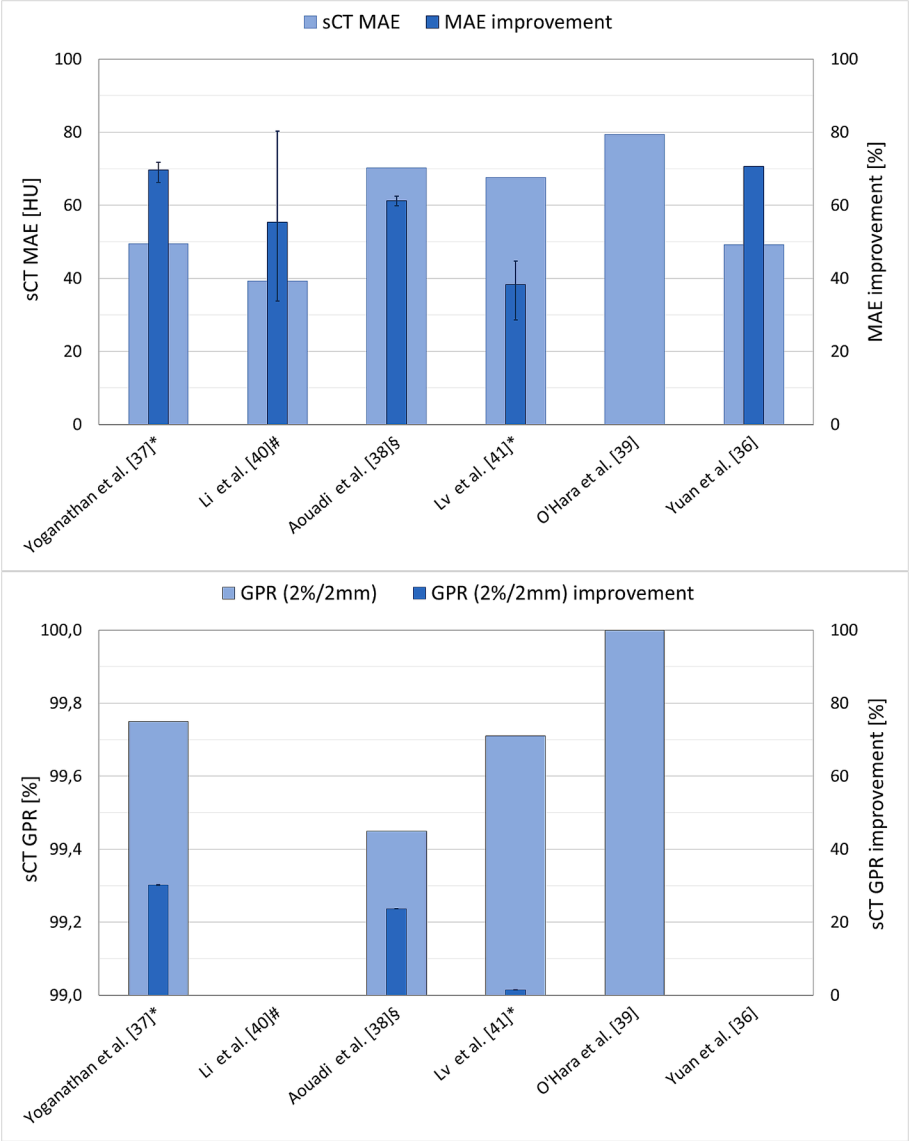

Fig. 2. Mean values of the main image and dose similarity metrics reported in the studies of the synthetic CT category. The average improvements relative to CBCT or MVCT are also represented together with the overall minimum and maximum achieved (represented by the error bars). Upper panel: Mean absolute error (MAE). Lower panel: Gamma passing rate with tolerance criteria of 2 % and 2 mm (GPR (2 %/2mm)). (*)For Yoganathan et al. [37] and Lv et al. [41] mean values are computed among the considered networks and dimensionality of models, (#)for Li et al. [40] among different accelerators, (§)for Aouadi et al. [38] among two networks.

DL approaches, such as deep residual UNet (DRUnet) and conditional generative adversarial network (cGAN), were compared to non-DL methods, such as atlas-based ones [38]. Although all methods improved the CBCT image quality, the results of DRUnet and cGAN were very similar and outperformed the atlas-based one. Dosimetric evaluation of DRUnet sCTs presented percentages of DVH metric deviations below 0.6 % for PTVs and 1.2 % for OARs with a GPR of 99.5 %.

Considering other non-DL approaches for sCT generation together with a cycleGAN model [39], all of them, excluding mass density assignment, showed mean MAE lower than 80 HU, GPR higher than 99.9 % and dose differences below 0.3 Gy. Overall, the cycleGAN method was found to be most convenient for clinical use providing both dosimetric accuracy and short generation time.

Another important feature is generalizability across different accelerators. A registration GAN (RegGAN) model proved to be suitable for generalization across different accelerators [40], the assessment of image quality highlighted a great improvement in sCT over CBCT across all of them, with the best result observed on the Synergy (Elekta Corporation, Sweden) accelerator: MAE decreased from 231.3 to 45.6, while PSNR increased from 19.4 to 26.8 and SSIM from 0.82 to 0.85. Dose differences were always within 3 % with no significant differences in dose discrepancies between linacs ($p < 0.05$).

MVCT images acquired on tomotherapy (Accuray, Sunnyvale, USA) were also used to generate high-quality synthetic kilovoltage CT (skVCT) [41] through a 3D cycleGAN with residual blocks (cycleGAN-Resnet) with attention gate. Quality evaluation metrics included MAE, PSNR and SSIM relative to kVCT with improvements of 44.7 %, 23.6 %, and 5.0 % compared to MVCT. The GPR 2 %/2mm also improved from 98.3 % to 99.7 %, while no significant differences ($p > 0.05$) were found comparing DVH parameters.

3.3. Auto-segmentation

ART offers potential improvements in cancer treatment by adjusting the treatment plan based on changes in the patient's anatomy. In this section we have collected the papers regarding auto-segmentation. Table 2 summarizes the main characteristics of 5 studies included in this section. The average values of the most commonly used evaluation metrics, investigated by at least two studies, are reported.

Fig. 3 shows the Dice similarity coefficient (DSC, upper panel) and various surface-based metrics (lower panel) together with their average improvement achieved over the other methods compared in the authors' work.

For offline ART, a self-channel-and-spatial-attention neural network (SCSA-Net) was proposed for OAR segmentation on CT images and compared with state-of-the-art segmentation methods [42]. SCSA-Net outperformed the other methods, achieving statistically significant improvements in terms of DSC (on average + 4 %) on all OARs.

Efficient auto-segmentation is even more challenging in CBCT-based ART due to the poor image quality of the CBCTs. To generate enhanced (e)CBCT, UNet shape architecture can be trained using planning CTs and first-fraction CBCTs [43]. Mean DSC of eCBCT-to-repeat simulation CT (rCT) was higher than original (o)CBCT-to-rCT (83 % vs 70 %), with statistically significant improvements in all OARs except the brainstem. Statistically significant improvements of the Hausdorff distance (HD) were found for all investigated OARs (mean HD 4.2 mm vs 7.2 mm).

Rigid body (RB) and deformable image registration (DIR) algorithms can be integrated with DL segmentation models in a registration-guided (Rg)DL framework [44]. RB-RgDL achieved a mean DSC of 84.5 %, outperforming RB contour propagation and DL segmentation without guidance by 4.5 % and 4.7 %, respectively. DIR-RgDL achieved a mean DSC of 86.5 % outperforming DIR and DL alone by 2.4 % and 6.7 %, respectively.

Pre-trained DL-based DIR population models can be refined with test-time optimization (TTO) models [45]. Using TTO, average improvements of the DSC can be up to 5 % with the individualized models compared to the population models using different DL networks, ranging from very simple CNN to state-of-the-art neural networks, like

Table 2
Summary of 5 studies included in the auto-segmentation section. The main results of the most commonly used evaluation metrics are reported.

Author	Journal (Year)	DL Model	AI SW	Imaging	OARs	No. of patients (training/ validation/ test)	Evaluation Metric				
							DSC [%]	HD [mm]	ASD [mm]	COMD [mm]	SEN [%]
Kakkos I et al. [46]	Bioengineering (2024)	AttentionUNet	PyTorch	pCT, CBCT	L/R parotid gland	113* (n.d.)	82.7	6.2	–	–	82.7
Liang X et al. [45]	Med Phys (2022)	CycleGAN	TensorFlow	pCT, CBCT, sCT	L/R brachial plexus, brainstem, oral cavity, constrictor, esophagus, nodal GTV, larynx, mandible, L/R masseter, posterior arytenoid-cricoid space, L/R parotid gland, L/R SMG, spinal cord	239 (200/-/ 39)	84.5	2.3§	–	1.7	–
Ma L et al. [44]	Med Phys (2022)	UNet	Pytorch	pCT, CBCT	mandible, L/R parotid gland, L/R SMG, L/R masseters	37 (22/7/8)	86.5	2.5§	0.6	–	–
Chen W et al. [43]	Front Artif Intell (2021)	UNet	Keras and TensorFlow	pCT; rCT; oCBCT; eCBCT; sCT	L/R parotid gland, L/R SMG, larynx, brainstem, spinal cord	55 (40/15/-)	83.0	4.2	–	2.8	–
Gou S et al. [42]	Phys Med Biol (2020)	SCSA-Net	Keras and TensorFlow	pCT; rCT	brainstem, optic chiasm, L/R optic nerve, L/R parotid gland, mandible, L/R SMG	58** (38/ 10/10)	79.7	–	1.2	–	81.4

Notes: *90H&N CT images from two open-access information repositories: 48 from The Cancer Imaging Archive (TCIA) [47] and 42 from The Head and Neck Organ-at-Risk CT & MR Segmentation Dataset (HaN-Seg) [48]; **48H&N CT images from the TCIA[47]; §HD95 = 95th percentile of HD.
Abbreviations: AI SW = artificial intelligence software; ASD = average surface distance; CBCT = cone-beam computed tomography; COMD = center-of-mass distance; DCNN = deep-learning convolutional neural network; DL = deep learning; DSC = Dice similarity coefficient; eCBCT = enhanced CBCT; GTV = gross tumor volume HD = Hausdorff distance; L/R = left/right; n.d. = not defined; oCBCT = CBCT scan taken on the same day; pCT = planning CT; rCT = repeat simulation CT; SCSA-Net = self-channel-and-spatial-attention neural network; sCT = synthetic CT; SEN = sensitivity; SMG = submandibular gland.

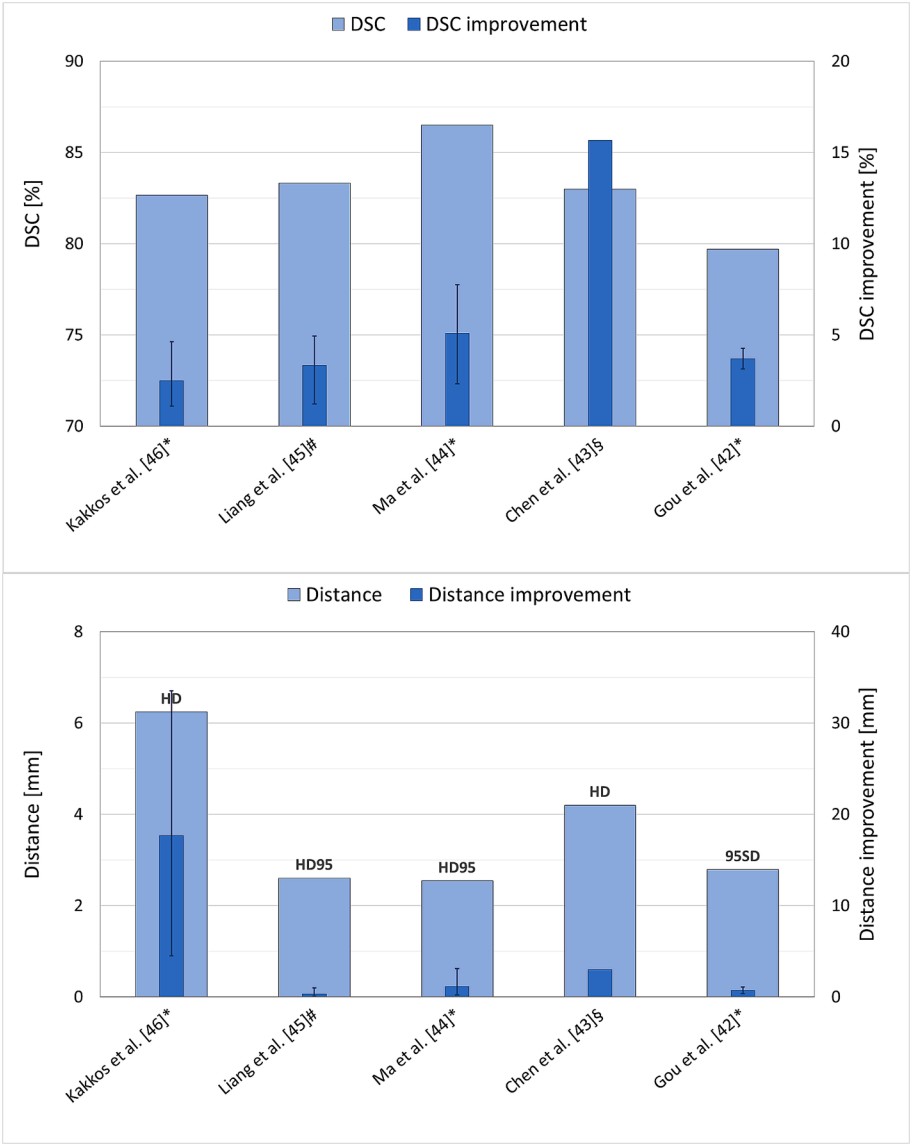

Fig. 3. Mean values of the key performance metrics reported in the studies included in this review. The average improvements achieved over the other methods compared in the authors' work are also reported, together with the minimum and maximum improvements (represented by the error bars). Upper panel: Dice similarity coefficient (DSC). Lower panel: Key surface-based metrics (HD = Hausdorff distance; HD95 = 95th percentile HD; 95SD = 95 % maximum surface distance). (*)For Kakkos et al. [46], Ma et al. [44] and Gou et al. [42] the improvement values were calculated with respect to other DL models, (#)for Liang et al. [45] before and after TTO application on different DL architectures, (§)for Chen et al. [43] to the original CBCTs.

VoxelMorph and cascaded VTN.

Parotid glands can undergo significant volume reductions during RT. An AttentionUNet-based DL method can be used for parotid gland auto-segmentation with excellent results [46]. Indeed, AttentionUNet outperformed against other state-of-the-art DL segmentation approaches with DSC (HD) improvements ranging from + 1.1 to 4.6 % (from 4.5 to 33.5 mm).

As shown in Fig. 3, the maximum DSC improvement (+15.7 %) was reported by Chen et al.[43]. The other studies reported mean improved performances ranging from 2.5 %[46] to 5.1 %[44] in comparison to other DL methods. Different surface-based metrics and/or percentiles were reported in the various studies. The maximum distance improvement was reported by Kakkos et al.[46] in the delineation of parotid glands, with a mean HD improvement of 17.7 mm. The other studies reported mean improved performances ranging from 0.3 mm ([45], 95th percentile of HD) to 3.0 mm ([43], HD).

3.4. Radiomics and predictive models

Several studies in recent years have evaluated the possibility of training AI models trained by radiomics and clinical features. Analysis of radiomic features can be a tool to define patients in need a re-planning based on H&N volumetric changes. In detail, the H&N volumes were evaluated with CBCT and classified into a group with a large volume reduction and another group with no reduction (threshold for the classification was defined as reduction rate being at least 10 % in three consecutive fractions). The quantitative evaluation consisted in the extraction of features from the primary gross tumor volume (GTV) and nodal (n)GTV, feature selection using four different methods and two machine learning (ML) algorithms were tested to develop the prediction model. The best predictive performance was random forest model with AUC, accuracy, sensitivity and specificity of 0.662, 0.692, 0.700 and 0.816, respectively [49].

Another potentiality investigated in literature was the role of delta-radiomics features extracted from CBCT images for the purpose of

identifying patients eligible for ART. The ML framework, trained on delta-radiomics features of CTV and parotid glands, achieved 0.90 classification performance accuracy while detecting a small subset of discriminative characteristics from the 1st week of RT [50].

The ML models can also be able to predict local failure (LF) in patients who underwent H&N radiotherapy. Separate models of primaries or nodes were created using the explainable boosting machine (EBM) classifier with 5-fold cross-validation for clinical only, radiomic only, and fused models (clinical + radiomic). The latter achieved an AUC of 0.871 with a sensitivity of 78.3 % and specificity of 90.9 % [51].

Instead, a deep neural network could predict the anatomical 3D displacement field (DF) changes of the parotid glands during RT through longitudinal pattern analysis of deformation between CBCT and planning CT images. The model consisted in a CNN and a recurrent neural network (RNN) for the feature extraction of weekly volumetric DFs, to restore the predicted DF and to analyze the longitudinal sequential feature. The results were then used as an input in their NN to predict the DF of the following weeks. The accuracy of the results was evaluated through the measurement of DSC between the manual and predicted contours [52].

Volumetric changes on CBCTs images could be the starting point that combined with other clinical data (including the mean oral cavity dose calculated from DVH, age, body mass index (BMI), the number of irradiations (fr), DSC of CTV, and rate of change in DSC) in an ML model, such as support vector machine (SVM) classifier and k-nearest neighbor (KNN) classifier can achieve a prediction accuracy > 90 % in the identification of patients eligible for ART [53].

A different application to those described above concerns the application of AI to automate the re-planning process. The performance of a commercial knowledge-based planning (KBP) tool to help to automate the process of re-planning were investigated. In particular, their model was developed to estimate the potential gain of plan adaptation in OAR sparing during the radiotherapy treatment course. A dataset of 100H&N patients was selected for the study. Using the training set, KBP DVH prediction models were created for 8 OARs. The gain from re-planning was quantified by comparing DVHs of the original plan recalculated on two CBCT acquired during the treatment course with the KBP DVH predictions and with the KBP DVH obtained with an adapted plan on the same image set. Overall, the predicted and final gains were correlated ($p < 0.01$) with a KBP prediction of 48 % of 310 cases who may benefit from plan adaptation confirmed in 47 % of final DVHs [54].

Table 3 summarize the main characteristics of these studies included in this section.

Table 3
Summary of 3 radiomics studies (above) and 3 non-radiomic studies (below) included in the radiomics and predictive models section.

Author	Journal (Year)	Radiomics (SW)	Algorithm (SW)	Prediction	ROIs	No. of patients (training/ validation/ test)	FTs types (No.)	RQS 2.0
Ishizawa M et al. [49]	J Radiat Res (2023)	Handcrafted (Pyradiomics)	RF, SVM (Scikit-learn)	The need of ART	GTV and nodal GTV	172 (104/ 34/34)	Shape, first-order, GLCM, GLRLM, GLSZM, GLDM, NGTDM (107)	16 (24.4 %)
Iliadou V et al. [50]	Cancers (2022)	Handcrafted (Pyradiomics)	SVM, KNN (MATLAB)	The need of ART	Clinical Target Volume (CTV) and Parotid Glands (PG)	40 (n.d.)	Shape, first-order, GLCM, GLDM, GLRLM, GLSZM, NGTDM (104)	24 (36.4 %)
Morgan HE et al. [51]	Quant Imaging Med Surg (2021)	Handcrafted (Pyradiomics)	EBM (InterpretML)	Treatment Outcome	All primary structures and nodal structures receiving 70 Gy	90 (72/-/18)	Shape, first-order, GLCM, GLDM, GLRLM, GLSZM (102)	25 (37.9 %)
Author	**Journal (Year)**	**AI area**	**Algorithm (SW)**	**Prediction**	**Training data**		**Evaluation metric**	
Rachi T et al. [53]	Adv Radiat Oncol (2023)	ML	SVM, KNN (n.d.)	The need of ART	Number of irradiations, amount of change in CTV, rate of change in CTV, mean oral cavity dose, age, body mass index	68 (n.d.)	AS = 0.96 for SVM AS = 0.96 for KNN	
Cagni E et al. [54]	Phys Med Biol (2021)	ML	MCO-KBP (Eclipse TPS)	Amount of sparing in OARs in re-planning strategy for ART	DVH of original plan and re-calculated plan on daily CBCT	100 (80/-/ 20)	AUC = 0.83 SEN = 0.69 SPEC = 0.78	
Lee D et al. [52]	Phys Med Biol (2020)	Deep Learning (n.d.)	CNN,RNN (n.d.)	The need of ART	DFs between the planning pCT and CBCT	63 (53/-/10)	DICE (B-Spline)*= 0.81 ± 0.07/ 0.81 ± 0.04 (week 4), 0.79 ± 0.06/0.81 ± 0.05 (week 5) and 0.78 ± 0.06/0.82 ± 0.02 (week 6)DICE (Demons) * = 0.78 ± 0.08/0.82 ± 0.03 (week 4), 0.77 ± 0.07/0.82 ± 0.04 (week 5) and 0.75 ± 0.07/0.82 ± 0.02 (week 6)	

Note: *DICE of ipsilateral/contralateral parotid glands at week 4 to 6.

Abbreviations: AI = artificial intelligence; ART = adaptive radiotherapy; AS = accuracy score; AUC = area under the curve; CNN = convolution neural network; DICE = dice similarity; DFs = displacement fields; DL = deep learning; DVH = dose volume histogram; EBM = explainable boosting machine; FT = feature; KBP = knowledge-based planning; KNN = k-nearest neighbor; GLCM = gray-level co-occurrence matrix; GLDM = gray-level dependence matrix; GLRLM = gray-level run length matrix; GLSZM = gray-level size zone matrix, MCO = multicriteria optimization; ML = machine learning; n.d. = not defined; NGTDM = neighboring gray tone difference matrix;; OAR = organ at risk; pCT = planning CT; RF = random forest; RNN = recurrent neural networks; RQS = radiomic quality score; SEN = sensitivity; SPEC = specificity; SVM = support vector machine; SVM = support vector machine; SW = software; TPS = treatment planning system.

4. Discussion

This systematic review explored how AI can streamline various processes of the ART chain by automating key tasks. In particular, the most relevant research findings related to three different tasks of the ART workflow, i.e. generation of synthetic CTs, auto-segmentation and predictive models, were summarized. The results reported here clearly demonstrated that automation plays a critical role in facilitating the efficient implementation of ART, particularly for online approaches that require faster workflows.

This review focused on H&N cancer patients who can significantly benefit from ART strategies [55] since they are frequently subject to changes in anatomy during radiotherapy, such as target shrinkage, OARs modifications or weight loss. This anatomical site is also one of the most challenging in RT due to the large number of OARs involved in the treatment fields, some of which are serial-like structures, e.g. optic pathways and brainstem [12]. In addition, treatments are generally delivered with simultaneous integrated boost technique [56] with 2 to 4 dose levels [10], making planning even more complex.

Overall, quality assessments highlighted areas similar to those reported by Wang et al. [32] where AI-based ART research could be improved to minimize the risk of bias and increase the generalizability of the results.

Currently, CT-based online ART workflows require dedicated equipment, e.g. dedicated CBCT-guided machine [57] and many efforts are being invested to bring online ART to standard C-arm linac and to integrate AI into RT workflows [58,59].

The use of AI algorithms to generate sCTs from CBCT, MRI and PET raised increasing interest, however studies on MRI and PET mainly focus on MRI-only RT and PET attenuation correction [60,61]. Regarding CBCT, early works focused only on DL capabilities of improving CBCT image quality for IGRT, while recently many studies have explored the use of these methods to convert CBCT/MVCT to sCT images able to resemble pCTs for ART [62].

Dose similarity was evaluated with DVH-based metrics in 5 papers, nevertheless, differences in reported information (PTV and/or OARs, DVH points, etc.) make them unsuitable for comparisons. On the other hand, gamma analysis assessment [63] with tight criteria of 2 %/2mm [64] was included in 4 articles. Overall, GPR values relative to the considered ground-truth are greater than 99.5 %. Only Yoganathan et al. [37] specified GPR value for several network architecture and dimensionality of model, obtaining very similar results with a mean improvement of 30.2 %.

Both Aouadi et al. [38] and O'Hara et al. [39] concluded that DL strategies for sCT generation outperformed other methods with higher dosimetric accuracy and shorter generation time, making them most suitable for clinical use.

ART also requires rapid and precise segmentation of organs at risk and treatment volumes to adapt the treatment plan to the new anatomy. Manual segmentation is generally slow, tedious and requires time and human resources. Therefore, in clinical practice it may not be feasible for all patients needing plan adaptation.

Five out of 17 studies investigated fast and accurate auto-segmentation of OARs in different imaging CT-based modalities, for both online (4/5) and offline (1/5) ART.

All the studies investigated auto-segmentation of the parotid glands, which are among the most sensitive OARs as they can undergo a volume reduction of about 30 % during RT [65,66].

All the studies investigated and compared at least three assessment metrics. The most commonly used metrics were geometric as recently reported in literature [67]. The most investigated metric was the DSC (5/5), an overlap-based metric between an auto-contour and a ground-truth structure. However, none of the studies reported information on the measurement of inter- and intra- observer variability and/or ways to mitigate such variabilities. Comparing the studies included in this review with different OARs analyzed, we observed a small variation in the average DSC, ranging from 80 % [42] to 87 % [44], with a variation for the parotid glands from 83 % [46] to 88 % [44]. In general, a DSC greater than 80 % is considered to be a high similarity [68]. In their review, Liu et al. [68] analyzed the performance of different DL algorithms in contouring H&N OARs. They reported a mean DSC of 85 % for the parotid glands, while the worst performances were found for the optic pathways, with DSC of 62 %(73 %) for the optic chiasm(nerves). In our review, the model by Gou et al. [42] achieved the worst performance in terms of average DSC (80 %). However it was the only study which included the optic pathways, with DSC of 61 %(72 %) for the optic chiasm(nerves), in agreement with the data summarized in [68]. Excluding optical pathways from the analysis, the average DSC would increase to 86 %. Anyway, all the works included in our review reported improved performances in the DSC (improvement range 3 %-16 %) in comparison to other DL methods, from simple CNNs to state-of-the-art architectures. Therefore, all the studies achieved very good results in terms of DSC.

Although a more pressing clinical need would be target segmentation, the only work which evaluated the nGTV was the study by Liang et al. [45], who obtained a mean DSC of 79 % with a mean improvement of 5 %. They also reported significant improvements of 7 % in outlier patients with structures of large anatomical changes, such as parotid glands and nGTV.

Despite the recent challenge demonstrated that OAR auto-segmentation tools in H&N can achieve performance comparable to or even surpassing inter-observer agreement [69], automated target segmentation requires further investigation.

Even though in other disease sites, e.g. lungs, much has already been done for detection, screening and diagnosis, only recently the efficacy of DL algorithms in CT target segmentation was investigated by Wang et al. [32], who highlighted that different tumor stages and clinical settings can affect delineation accuracy.

Highest variations were found in the mean improvements of surface-based metrics, ranging from 0.3 mm [45] to 17.7 mm [46]. Different metrics and/or percentiles were chosen potentially causing bias and greater variation in results [67] and making comparison between studies more difficult. Nevertheless, all the studies included in our review investigated at least two surface-based metrics, with HD being the most popular. Gou et al. [42] and Kakkos et al. [46] also evaluated classification accuracy metrics (e.g. sensitivity of their method).

The only study which performed a deep time–cost analysis was the one by Liang et al. [45] which found the number of iterations might be a trade-off between accuracy and efficiency for clinical implementation.

In principle, the presence of data from different fractions can be used as priors for model building. However, none of the included studies investigated the influence of multiple CBCTs acquired at different timepoints. None of the studies used dosimetric assessment metrics or evaluated the potential of radiomic features as a useful metric for comparing segmentation accuracy [70].

The last 6/17 papers investigated strategies to predict the need for ART, such as handcrafted radiomic features for early prediction of planning adaptation requirement due to volumetric alterations. Only 3 studies performed radiomic analysis and among them only one used the methodology of radiomic deep learning. Their qualities were assessed by the RQS2.0 score. They achieved a low RQS, none above 50 %. The main reasons for this are related to: lack of prospective studies, validation on external datasets, use of textural data harmonization and statistical optimization. All these factors then represent limitations to the repeatability of the results. Despite this, however, the scores evaluated are in line with reviews of radiomic studies also on other anatomical sites such as breast and lung [7172].

Among the remaining studies included in the predictive models section, in the work of Rachi et al. [53] ML algorithms were used to predict patients undergoing H&N radiotherapy treatment who are candidates for ART. Again, a validation of the model on external cohorts could lead to positive implications on clinical activity.

And finally, a study with the objective of evaluating the performance of a commercial, knowledge-based tool for radiotherapy planning could be used to estimate the amount of sparing in organs at risk (OARs) in the re-planning strategy for adaptive radiotherapy (ART) [54].

A limitation of this review could be related to the focus, centered only on conventional CT-based photon radiotherapy. However, despite the growing applications of MR-linacs and the use of other radiating particles, it remains clinically more widespread on a much larger scale due to the lower cost per patient of this technology [7374]. In addition, the implementation of MR-linac for treating H&N cancer is still developing [7576] and even if the variations in patient anatomy in particle therapy may cause even higher deviations due to the finite range [77], only in recent years in-room CT and CBCT were introduced in clinical practice [78] and daily 3D image guidance is currently used in a small percentage of facilities [79]. Furthermore, in particle therapy MR imaging remains at its infancy and is far from its clinical implementation [80].

Based on the findings of this review, we sincerely believe that there are currently unmet needs in AI-based ART research. Among the potential future directions, we believe that the most relevant ones include target auto-segmentation and time-efficient patient-specific quality assurance and outcome prediction, as also reported in [81]. These tasks could be integrated with AI-models developed for decision-making to identify patients who can really benefit from plan adaptation. We also believe that prospective studies are needed to validate and standardize these models before realistic clinical implementation.

In conclusion, automation of processes in adaptive radiation therapy for H&N cancer is very promising throughout the entire chain, starting from the generation of high-quality synthetic CT to the prediction of the need for re-planning. The integration of AI tools in clinical practice can significantly streamline the ART workflow by automating repetitive tasks, thus reducing the workload and mitigating the pressure on the staff.

Funding statement

The authors received no specific funding for this work.

CRediT authorship contribution statement

Edoardo Mastella: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Francesca Calderoni:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Luigi Manco:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Martina Ferioli:** Methodology, Validation, Formal analysis, Investigation, Writing – review & editing. **Serena Medoro:** Supervision, Writing – review & editing. **Alessandro Turra:** Supervision, Writing – review & editing. **Melchiore Giganti:** Supervision, Writing – review & editing. **Antonio Stefanelli:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2025.100731.

References

[1] Fiorino C, Guckenberger M, Schwarz M, van der Heide UA, Heijmen B. Technology-driven research for radiotherapy innovation. Mol Oncol 2020;14:1500–13. https://doi.org/10.1002/1878-0261.12659.
[2] Verellen D, De RM, Linthout N, Tournel K, Soete G, Storme G. Innovations in image-guided radiotherapy. Nat Rev Cancer 2007;7:949–60. https://doi.org/10.1038/nrc2288.
[3] Rong Y, Ding X, Daly ME. Hypofractionation and SABR: 25 years of evolution in medical physics and a glimpse of the future. Med Phys 2023;50:95–103. https://doi.org/10.1002/mp.16270.
[4] Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. Nat Rev Clin Oncol 2012;9:688–99. https://doi.org/10.1038/nrclinonc.2012.194.
[5] Cuccia F, Alongi F, Belka C, Boldrini L, Hörner-Rieber J, McNair H, et al. Patient positioning and immobilization procedures for hybrid MR-Linac systems. Radiat Oncol 2021;16:183. https://doi.org/10.1186/s13014-021-01910-6.
[6] Oderinde OM, Shirvani SM, Olcott PD, Kuduvalli G, Mazin S, Larkin D. The technical design and concept of a PET/CT linac for biology-guided radiotherapy. Clin Transl Radiat Oncol 2021;29:106–12. https://doi.org/10.1016/j.ctro.2021.04.003.
[7] Bissonnette J, Balter PA, Dong L, Langen KM, Lovelock DM, Miften M, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a report of the AAPM TG-179. Med Phys 2012;39:1946–63. https://doi.org/10.1118/1.3690466.
[8] Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. Phys Med Biol 1997;42:123–32. https://doi.org/10.1088/0031-9155/42/1/008.
[9] Merlotti A, Alterio D, Vigna-Taglianti R, Muraglia A, Lastrucci L, Manzo R, et al. Technical guidelines for head and neck cancer IMRT on behalf of the Italian association of radiation oncology - head and neck working group. Radiat Oncol 2014;9:264. https://doi.org/10.1186/s13014-014-0264-9.
[10] Lee AW, Ng WT, Pan JJ, Chiang C-L, Poh SS, Choi HC, et al. International guideline on dose prioritization and acceptance criteria in radiation therapy planning for nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2019;105:567–80. https://doi.org/10.1016/j.ijrobp.2019.06.2540.
[11] Brodin NP, Kabarriti R, Garg MK, Guha C, Tomé WA. Systematic review of normal tissue complication models relevant to standard fractionation radiation therapy of the head and neck region published after the QUANTEC reports. Int J Radiat Oncol Biol Phys 2018;100:391–407. https://doi.org/10.1016/j.ijrobp.2017.09.041.
[12] Chow JCH, Ho JCS, Cheung KM, Johnson D, Ip BYM, Beitler JJ, et al. Neurological complications of modern radiotherapy for head and neck cancer. Radiother Oncol 2024;194:110200. https://doi.org/10.1016/j.radonc.2024.110200.
[13] Berg M, Hansson C, Silander E, Bove M, Johansson L, Haugen Cange H, et al. A randomized study comparing the nutritional effects of radiotherapy with cetuximab versus cisplatin in patients with advanced head and neck cancer. Head Neck 2024;46:760–71. https://doi.org/10.1002/hed.27619.
[14] Hawkins PG, Kadam AS, Jackson WC, Eisbruch A. Organ-sparing in radiotherapy for head-and-neck cancer: improving quality of life. Semin Radiat Oncol 2018;28:46–52. https://doi.org/10.1016/j.semradonc.2017.08.002.
[15] Hutchison AR, Cartmill B, Wall LR, Ward EC. Dysphagia optimized radiotherapy to reduce swallowing dysfunction severity in patients undergoing treatment for head and neck cancer: a systematized scoping review. Head Neck 2019;41:2024–33. https://doi.org/10.1002/hed.25688.
[16] Geets X, Tomsej M, Lee JA, Duprez T, Coche E, Cosnard G, et al. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: impact on target volume delineation and dose distribution using helical tomotherapy. Radiother Oncol 2007;85:105–15. https://doi.org/10.1016/j.radonc.2007.05.010.
[17] Ahn PH, Chen C-C, Ahn AI, Hong L, Scripes PG, Shen J, et al. Adaptive planning in intensity-modulated radiation therapy for head and neck cancers: single-institution experience and clinical implications. Int J Radiat Oncol Biol Phys 2011;80:677–85. https://doi.org/10.1016/j.ijrobp.2010.03.014.
[18] Zhao L, Wan Q, Zhou Y, Deng X, Xie C, Wu S. The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma. Radiother Oncol 2011;98:23–7. https://doi.org/10.1016/j.radonc.2010.10.009.
[19] Schwartz DL, Garden AS, Thomas J, Chen Y, Zhang Y, Lewin J, et al. Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. Int J Radiat Oncol Biol Phys 2012;83:986–93. https://doi.org/10.1016/j.ijrobp.2011.08.017.
[20] Yang H, Hu W, Wang W, Chen P, Ding W, Luo W. Replanning during intensity modulated radiation therapy improved quality of life in patients with nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2013;85:e47–54. https://doi.org/10.1016/j.ijrobp.2012.09.033.
[21] Chen AM, Daly ME, Cui J, Mathai M, Benedict S, Purdy JA. Clinical outcomes among patients with head and neck cancer treated by intensity-modulated radiotherapy with and without adaptive replanning. Head Neck 2014;36:1541–6. https://doi.org/10.1002/hed.23477.
[22] Morgan HE, Sher DJ. Adaptive radiotherapy for head and neck cancer. Cancers Head Neck 2020;5:1. https://doi.org/10.1186/s41199-019-0046-z.
[23] Castelli J, Thariat J, Benezery K, Hasbini A, Gery B, Berger A, et al. Weekly adaptive radiotherapy vs standard intensity-modulated radiotherapy for improving salivary function in patients with head and neck cancer. JAMA Oncol 2023;9:1056. https://doi.org/10.1001/jamaoncol.2023.1352.
[24] Volpe S, Pepa M, Zaffaroni M, Bellerba F, Santamaria R, Marvaso G, et al. Machine learning for head and neck cancer: a safe bet?—a clinically oriented systematic

review for the radiation oncologist. Front Oncol 2021;11:772663. https://doi.org/10.3389/fonc.2021.772663.
[25] Lindegaard AM, Håkansson K, Bernsdorf M, Gothelf AB, Kristensen CA, Specht L, et al. A systematic review on clinical adaptive radiotherapy for head and neck cancer. Acta Oncol 2023;62:1360–8. https://doi.org/10.1080/0284186X.2023.2245555.
[26] Nuyts S, Bollen H, Eisbruch A, Strojan P, Mendenhall WM, Ng SP, et al. Adaptive radiotherapy for head and neck cancer: pitfalls and possibilities from the radiation oncologist's point of view. Cancer Med 2024;13:e7192.
[27] Dona Lemus OM, Cao M, Cai B, Cummings M, Zheng D. Adaptive radiotherapy: next-generation radiotherapy. Cancers 2024;16:1206. https://doi.org/10.3390/cancers16061206.
[28] Mäkitie AA, Alabi RO, Ng SP, Takes RP, Robbins KT, Ronen O, et al. Artificial intelligence in head and neck cancer: a systematic review of systematic reviews. Adv Ther 2023;40:3360–80. https://doi.org/10.1007/s12325-023-02527-9.
[29] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. statement: an updated guideline for reporting systematic reviews. BMJ 2020;2021:n71. https://doi.org/10.1136/bmj.n71.
[30] Whiting PF. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
[31] Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020;2:e200029. https://doi.org/10.1148/ryai.2020200029.
[32] Wang T-W, Hong J-S, Huang J-W, Liao C-Y, Lu C-F, Wu Y-T. Systematic review and meta-analysis of deep learning applications in computed tomography lung cancer segmentation. Radiother Oncol 2024;197:110344. https://doi.org/10.1016/j.radonc.2024.110344.
[33] https://www.radiomics.world/rqs2 n.d. https://www.radiomics.world/rqs2 (accessed October 7, 2024).
[34] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62. https://doi.org/10.1038/nrclinonc.2017.141.
[35] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13:600–12. https://doi.org/10.1109/tip.2003.819861.
[36] Yuan N, Dyer B, Rao S, Chen Q, Benedict S, Shang L, et al. Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy. Phys Med Biol 2020;65:035003. https://doi.org/10.1088/1361-6560/ab6240.
[37] Yoganathan SA, Aouadi S, Ahmed S, Paloor S, Torfeh T, Al-Hammadi N, et al. Generating synthetic images from cone beam computed tomography using self-attention residual UNet for head and neck radiotherapy. Phys Imaging Radiat Oncol 2023;28:100512. https://doi.org/10.1016/j.phro.2023.100512.
[38] Aouadi S, Yoganathan S, Torfeh T, Paloor S, Caparrotti P, Hammoud R, et al. Generation of synthetic CT from CBCT using deep learning approaches for head and neck cancer patients. Biomed Phys Eng Express 2023;9:055020. https://doi.org/10.1088/2057-1976/acea27.
[39] O'Hara CJ, Bird D, Al-Qaisieh B, Speight R. Assessment of CBCT–based synthetic CT generation accuracy for adaptive radiotherapy planning. J Appl Clin Med Phys 2022;23:e13737. https://doi.org/10.1002/acm2.13737.
[40] Li Z, Zhang Q, Li H, Kong L, Wang H, Liang B, et al. Using RegGAN to generate synthetic CT images from CBCT images acquired with different linear accelerators. BMC Cancer 2023;23:828. https://doi.org/10.1186/s12885-023-11274-7.
[41] Lv T, Xie C, Zhang Y, Liu Y, Zhang G, Qu B, et al. A qualitative study of improving megavoltage computed tomography image quality and maintaining dose accuracy using cycleGAN-based image synthesis. Med Phys 2024;51:394–406. https://doi.org/10.1002/mp.16633.
[42] Gou S, Tong N, Qi S, Yang S, Chin R, Sheng K. Self-channel-and-spatial-attention neural network for automated multi-organ segmentation on head and neck CT images. Phys Med Biol 2020;65:245034. https://doi.org/10.1088/1361-6560/ab79c3.
[43] Chen W, Li Y, Yuan N, Qi J, Dyer BA, Sensoy L, et al. Clinical enhancement in AI-based post-processed fast-scan low-dose CBCT for head and neck adaptive radiotherapy. Front Artif Intell 2021;3:614384. https://doi.org/10.3389/frai.2020.614384.
[44] Ma L, Chi W, Morgan HE, Lin M, Chen M, Sher D, et al. Registration-guided deep learning image segmentation for cone beam CT–based online adaptive radiotherapy. Med Phys 2022;49:5304–16. https://doi.org/10.1002/mp.15677.
[45] Liang X, Chun J, Morgan H, Bai T, Nguyen D, Park J, et al. Segmentation by test-time optimization for CBCT-based adaptive radiation therapy. Med Phys 2023;50:1947–61. https://doi.org/10.1002/mp.15960.
[46] Kakkos I, Vagenas TP, Zygogianni A, Matsopoulos GK. Towards automation in radiotherapy planning: a deep learning approach for the delineation of parotid glands in head and neck cancer. Bioengineering 2024;11:214. https://doi.org/10.3390/bioengineering11030214.
[47] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26:1045–57. https://doi.org/10.1007/s10278-013-9622-7.
[48] Podobnik G, Strojan P, Peterlin P, Ibragimov B, Vrtovec T. HaN-Seg: The head and neck organ-at-risk CT and MR segmentation dataset. Med Phys 2023;50:1917–27. https://doi.org/10.1002/mp.16197.
[49] Ishizawa M, Tanaka S, Takagi H, Kadoya N, Sato K, Umezawa R, et al. Development of a prediction model for head and neck volume reduction by clinical factors, dose–volume histogram parameters and radiomics in head and neck cancer. J Radiat Res 2023;64:783–94. https://doi.org/10.1093/jrr/rrad052.
[50] Iliadou V, Kakkos I, Karaiskos P, Kouloulias V, Platoni K, Zygogianni A, et al. Early prediction of planning adaptation requirement indication due to volumetric alterations in head and neck cancer radiotherapy: a machine learning approach. Cancers 2022;14:3573. https://doi.org/10.3390/cancers14153573.
[51] Morgan HE, Wang K, Dohopolski M, Liang X, Folkert MR, Sher DJ, et al. Exploratory ensemble interpretable model for predicting local failure in head and neck cancer: the additive benefit of CT and intra-treatment cone-beam computed tomography features. Quant Imaging Med Surg 2021;11:4781–96. https://doi.org/10.21037/qims-21-274.
[52] Lee D, Zhang P, Nadeem S, Alam S, Jiang J, Caringi A, et al. Predictive dose accumulation for HN adaptive radiotherapy. Phys Med Biol 2020;65:235011. https://doi.org/10.1088/1361-6560/abdb8.
[53] Rachi T, Ariji T, Takahashi S. Development of machine-learning prediction programs for delivering adaptive radiation therapy with tumor geometry and body shape changes in head and neck volumetric modulated arc therapy. Adv Radiat Oncol 2023;8:101172. https://doi.org/10.1016/j.adro.2023.101172.
[54] Cagni E, Botti A, Chendi A, Iori M, Spezi E. Use of knowledge based DVH predictions to enhance automated re-planning strategies in head and neck adaptive radiotherapy. Phys Med Biol 2021;66:135004. https://doi.org/10.1088/1361-6560/ac08b0.
[55] Brouwer CL, Steenbakkers RJHM, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? Radiother Oncol 2015;115:285–94. https://doi.org/10.1016/j.radonc.2015.05.018.
[56] Orlandi E, Palazzi M, Pignoli E, Fallai C, Giostra A, Olmi P. Radiobiological basis and clinical results of the simultaneous integrated boost (SIB) in intensity modulated radiotherapy (IMRT) for head and neck cancer: a review. Crit Rev Oncol Hematol 2010;73:111–25. https://doi.org/10.1016/j.critrevonc.2009.03.003.
[57] Liu H, Schaal D, Curry H, Clark R, Magliari A, Kupelian P, et al. Review of cone beam computed tomography based online adaptive radiotherapy: current trend and future direction. Radiat Oncol 2023;18:144. https://doi.org/10.1186/s13014-023-02340-2.
[58] Groot Koerkamp ML, Bol GH, Kroon PS, Krikke LL, Harderwijk T, Zoetelief AJ, et al. Bringing online adaptive radiotherapy to a standard C-arm linac. Phys Imaging Radiat Oncol 2024;31:100597. https://doi.org/10.1016/j.phro.2024.100597.
[59] Ersted Rasmussen M, Dueholm Vestergaard C, Folsted Kallehauge J, Ren J, Haislund Guldberg M, Nørrevang O, et al. RadDeploy: a framework for integrating in-house developed software and artificial intelligence models seamlessly into radiotherapy workflows. Phys Imaging Radiat Oncol 2024;31:100607. https://doi.org/10.1016/j.phro.2024.100607.
[60] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. Med Phys 2021;48:6537–66. https://doi.org/10.1002/mp.15150.
[61] Sherwani MK, Gopalakrishnan S. A systematic literature review: deep learning techniques for synthetic medical image generation and their applications in radiotherapy. Front Radiol 2024;4:1385742. https://doi.org/10.3389/fradi.2024.1385742.
[62] Rusanov B, Hassan GM, Reynolds M, Sabet M, Kendrick J, Rowshanfarzad P, et al. Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: a systematic review. Med Phys 2022;49:6019–54. https://doi.org/10.1002/mp.15840.
[63] Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. Med Phys 2003;30:2455–64. https://doi.org/10.1118/1.1598711.
[64] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 2018;45:e53–83. doi: doi: 10.1002/mp.12810.
[65] Ricchetti F, Wu B, McNutt T, Wong J, Forastiere A, Marur S, et al. Volumetric change of selected organs at risk during IMRT for oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2011;80:161–8. https://doi.org/10.1016/j.ijrobp.2010.01.071.
[66] Sanguineti G, Ricchetti F, Wu B, McNutt T, Fiorino C. Parotid gland shrinkage during IMRT predicts the time to Xerostomia resolution. Radiat Oncol 2015;10:19. https://doi.org/10.1186/s13014-015-0331-x.
[67] 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;35:354–69. https://doi.org/10.1016/j.clon.2023.01.016.
[68] Liu P, Sun Y, Zhao X, Yan Y. Deep learning algorithm performance in contouring head and neck organs at risk: a systematic review and single-arm meta-analysis. Biomed Eng Online 2023;22:104. https://doi.org/10.1186/s12938-023-01159-y.
[69] Podobnik G, Ibragimov B, Tappeiner E, Lee C, Kim JS, Mesbah Z, et al. HaN-Seg: the head and neck organ-at-risk CT and MR segmentation challenge. Radiother Oncol 2024;198:110410. https://doi.org/10.1016/j.radonc.2024.110410.
[70] Maffei N, Manco L, Aluisio G, D'Angelo E, Ferrazza P, Vanoni V, et al. Radiomics classifier to quantify automatic segmentation quality of cardiac sub-structures for radiotherapy treatment planning. Phys Med 2021;83:278–86. https://doi.org/10.1016/j.ejmp.2021.05.009.
[71] Urso L, Manco L, Castello A, Evangelista L, Guidi G, Castellani M, et al. PET-derived radiomics and artificial intelligence in breast cancer: a systematic review. Int J Mol Sci 2022;23:13409. https://doi.org/10.3390/ijms232113409.

[72] Evangelista L, Fiz F, Laudicella R, Bianconi F, Castello A, Guglielmo P, et al. PET radiomics and response to immunotherapy in lung cancer: a systematic review of the literature. Cancers 2023;15:3258. https://doi.org/10.3390/cancers15123258.
[73] Giuliani J, Mandarà M, Fiorica F. Cost effectiveness of magnetic resonance imaging guidance versus computed tomography guidance for stereotactic body radiotherapy. Eur Urol Oncol 2023. https://doi.org/10.1016/j.euo.2023.09.003.
[74] Lomax A. What will the medical physics of proton therapy look like 10 yr from now? A personal view. Med Phys 2018;45. https://doi.org/10.1002/mp.13206.
[75] Boeke S, Mönnich D, van Timmeren JE, Balermpas P. MR-guided radiotherapy for head and neck cancer: current developments, perspectives, and challenges. Front Oncol 2021;11. doi: 10.3389/fonc.2021.616156.
[76] McDonald BA, Dal Bello R, Fuller CD, Balermpas P. The use of MR-Guided radiation therapy for head and neck cancer and recommended reporting guidance. Semin Radiat Oncol 2024;34:69–83. https://doi.org/10.1016/j.semradonc.2023.10.003.
[77] Engelsman M, Schwarz M, Dong L. Physics controversies in proton therapy. Semin Radiat Oncol 2013;23:88–96. https://doi.org/10.1016/j.semradonc.2012.11.003.
[78] Paganetti H, Beltran C, Both S, Dong L, Flanz J, Furutani K, et al. Roadmap: proton therapy physics and biology. Phys Med Biol 2021;66:05RM01. doi: 10.1088/1361-6560/abcd16.
[79] Trnkova P, Zhang Y, Toshito T, Heijmen B, Richter C, Aznar MC, et al. A survey of practice patterns for adaptive particle therapy for interfractional changes. Phys Imaging Radiat Oncol 2023;26:100442. https://doi.org/10.1016/j.phro.2023.100442.
[80] Hoffmann A, Oborn B, Moteabbed M, Yan S, Bortfeld T, Knopf A, et al. MR-guided proton therapy: a review and a preview. Radiat Oncol 2020;15:129. https://doi.org/10.1186/s13014-020-01571-x.
[81] Hurkmans C, Bibault J-E, Brock KK, van Elmpt W, Feng M, David Fuller C, et al. A joint ESTRO and AAPM guideline for development, clinical validation and reporting of artificial intelligence models in radiation therapy. Radiother Oncol 2024;197:110345. https://doi.org/10.1016/j.radonc.2024.110345.