Magphan Phantoms for MR for radiation therapy and quantitative imaging applications.

Smári

Sub-voxel geometric distortion measurements and critical image quality metrics engineered for ease of use in the clinical workflow.

The Phantom Laboratory manufactures high-precision phantoms coupled with Smári image analysis service and innovative custom solutions for the medical imaging and radiation therapy fields.

Click to see our latest phantoms and schedule a demo of our Smári image analysis service.

Received: 9 February 2021 | Revised: 6 June 2021 | Accepted: 13 July 2021

DOI: 10.1002/mp.15150

REVIEW ARTICLE

MEDICAL PHYSICS

Deep learning based synthetic-CT generation in radiotherapy and PET: A review

Maria Francesca Spadea[1,*] | **Matteo Maspero**[2,3,*] | **Paolo Zaffino**[1] | **Joao Seco**[4,5]

[1] Department Experimental and Clinical Medicine, University "Magna Graecia" of Catanzaro, Catanzaro 88100, Italy

[2] Division of Imaging & Oncology, Department of Radiotherapy, University Medical Center Utrecht, Heidelberglaan, Utrecht, The Netherlands

[3] Computational Imaging Group for MR Diagnostics & Therapy, Center for Image Sciences, University Medical Center Utrecht, Heidelberglaan, Utrecht, The Netherlands

[4] Division of Biomedical Physics in Radiation Oncology, DKFZ German Cancer Research Center, Heidelberg, Germany

[5] Department of Physics and Astronomy, Heidelberg University, Heidelberg, Germany

Correspondence
Joao Seco, Division of Biomedical Physics in Radiation Oncology, DKFZ German Cancer Research Center, 69120 Heidelberg, Germany.
Email: j.seco@dkfz.de

[*] Maria Francesca Spadea and Matteo Maspero contributed equally to this study.

Abstract

Recently,deep learning (DL)-based methods for the generation of synthetic computed tomography (sCT) have received significant research attention as an alternative to classical ones. We present here a systematic review of these methods by grouping them into three categories, according to their clinical applications: (i) to replace computed tomography in magnetic resonance (MR) based treatment planning, (ii) facilitate cone-beam computed tomography based image-guided adaptive radiotherapy, and (iii) derive attenuation maps for the correction of positron emission tomography. Appropriate database searching was performed on journal articles published between January 2014 and December 2020. The DL methods' key characteristics were extracted from each eligible study, and a comprehensive comparison among network architectures and metrics was reported. A detailed review of each category was given, highlighting essential contributions, identifying specific challenges, and summarizing the achievements. Lastly, the statistics of all the cited works from various aspects were analyzed, revealing the popularity and future trends and the potential of DL-based sCT generation. The current status of DL-based sCT generation was evaluated, assessing the clinical readiness of the presented methods.

KEYWORDS
artificial intelligence, convolutional neural networks, deep learning, image synthesis, machine learning, pseudo-CT, radiotherapy

1 | INTRODUCTION

Medical imaging's impact on oncological patients' diagnosis and therapy has grown significantly over the last decades.[1] Especially in radiotherapy (RT),[2] imaging plays a crucial role in the entire workflow, from treatment simulation to patient positioning and monitoring.[3–6]

Traditionally, computed tomography (CT) is considered the primary imaging modality in RT. It provides accurate and high-resolution patient's geometry, enabling direct electron density conversion needed for dose calculations.[7] X-ray based imaging, including planar imaging and cone-beam computed tomography (CBCT), are widely adopted for patient positioning and monitoring before, during or after the dose delivery.[4] Along with CT, functional and metabolic information, mainly derived from positron emission tomography

ACRONYMS & ABBREVIATIONS: μ-map, attenuation map; AC, attenuation correction; aff, affine; AG, attention gate; CBCT, cone-beam computed tomography; CC, cross-correlation; CNNs, convolutional neural networks; cor, coronal; CT, computed tomography; cycle-GAN, cycle-consistent generative adversarial networks; DD, dose difference; def, deformable; DL, deep learning; DPR, dose pass rate; DSC, Dice similarity coefficient; DVH, dose-volume histogram; FLAIR, fluid-attenuated inversion recovery; FOV, field-of-view; GANs, generative adversarial networks; Gd, gadolinium; GPR, gamma pass rate; GRE, gradient recalled-echo; H&N, head-and-neck; IGART, image-guided adaptive radiation therapy; MAE, mean absolute error; MR, magnetic resonance; MRAC, magnetic resonance attenuation correction; MRI, magnetic resonance imaging; MSE, mean squared error; mUTE, multiple echo ultra-short echo time; NCC, normalised cross correlation; OAR, organ-at-risk; *p*, proton; paed, paediatric; PET, positron emission tomography; $PET_{|err|}$, absolute error positron emission tomography; PET_{err}, relative error positron emission tomography; PSNR, peak signal-to-noise ratio; rig, rigid; RMSE, root mean squared error; ROI, region of interest; RS, range shift; RT, radiotherapy; sag, sagittal; sCT, synthetic computed tomography; SSIM, structural similarity index measure; SUV, standard uptake values; tra, transverse; TSE, turbo spin-echo; UTE, ultra-short echo time; VOI, volume of interest; x, photon

(PET), is commonly acquired, allowing tumor staging and improving tumor contouring.[8] Magnetic resonance imaging (MRI) has also proved its added value for tumors and organs-at-risk (OARs) delineation, due to its superb soft tissue contrast.[9,10]

To benefit from the complementary advantages offered by different imaging modalities, MRI is generally registered to CT.[11] However, residual misregistration and differences in patient setup may introduce systematic errors that would affect the accuracy of the whole treatment.[12,13]

Recently, MR-only based RT has been proposed[14–16] to eliminate residual registration errors. Furthermore, it can simplify and speed-up the workflow, decreasing patient's exposure to ionizing radiation, which is particularly relevant for repeated simulations[17] or fragile populations, for example, children. Also, MR-only RT may reduce overall treatment costs[18] and workload.[19] Additionally, the development of MR-only techniques can be beneficial for MR-guided RT.[20]

The main obstacle regarding the introduction of MR-only RT is the lack of tissue attenuation information required for accurate dose calculations.[12,21] Many methods have been proposed to convert MR to CT-equivalent representations, often known as synthetic CT (sCT), for treatment planning and dose calculation. These approaches are summarised in two specific reviews on this topic,[22–24] in site-specific reviews,[18,25,26] or in broader review on MR-guided[27] or proton therapy.[28]

Additionally, similar techniques to derive sCT from a different imaging modality have been envisioned to improve the quality of CBCT.[29] CBCT plays a vital role in image-guided adaptive radiation therapy (IGART) for photon and proton therapy. However, due to the severe scatter noise and truncated projections, image reconstruction is affected by several artifacts, such as shading, streaking, and cupping.[30,31] For these reasons, daily CBCT has not commonly been used for online plan adaptation. The conversion of CBCT-to-CT would allow accurate dose computation and improve the quality of IGART provided to the patients.

Finally, sCT estimation is also crucial for PET attenuation correction (AC). Accurate PET quantification requires a reliable photon AC map, usually derived from CT. In the new PET/MRI hybrid scanners, this step is not immediate, and MRI-to-sCT translation has been proposed to solve the MR attenuation correction (MRAC) issue. Besides, stand-alone PET scanners can benefit from the derivation of sCT from uncorrected PET.[32–34]

In the last years, the derivation of sCT from MRI, PET, or CBCT has raised increasing interest based on artificial intelligence algorithms such as machine learning or DL.[35] This paper aims to perform a systematic review and summarise the latest developments, challenges, and trends in DL-based sCT generation methods. DL is a branch of machine learning, a field of artificial intelligence that involves using neural networks to generate hierarchical representations of the input data to learn a specific task without hand-engineered features.[36] Recent reviews have discussed the application of DL in RT,[37–43] and in PET AC.[34] Convolutional neural networks (CNNs), which are the most successful type of models for image processing,[44,45] have been proposed for sCT generation since 2016,[46] with a rapidly increasing number of published papers on the topic. However, DL-based sCT generation has not been reviewed in details, except for applications in PET.[47] With this survey, we aim at summarizing the latest developments in DL-based sCT generation, highlighting the contributions based on the applications and providing detailed statistics discussing trends in terms of imaging protocols, DL architectures, and performance achieved. Finally, the clinical readiness of the reviewed methods will be discussed.

2 | MATERIALS AND METHODS

A systematic review of techniques was carried out using the PRISMA guidelines. PubMed, Scopus, and Web of Science databases were searched from January 2014 to December 2020 using defined criteria (for more details, see Appendix). Studies related to radiation therapy, either with photons or protons and AC for PET, were included when dealing with sCT generation from MRI, CBCT, or PET. This review considered external beam radiation therapy, excluding, therefore, investigations that are focusing on brachytherapy. Conversion methods based on fundamental machine learning techniques were not considered in this review, preferring only DL-based approaches. Also, the generation of dual-energy CT was not considered along with the direct estimation of corrected attenuation maps from PET. Finally, conference proceedings were excluded: proceedings can contain valid methodologies; however, a large number of relevant abstracts and incomplete report of information were considered not suitable for this review. After the database search, duplicate articles were removed and records screened for eligibility. A citation search of the identified articles was performed.

Each included study was assigned to a clinical application category. The selected categories were

I. MR-only RT;
II. CBCT-to-CT for image-guided (adaptive) radiotherapy;
III. PET attenuation correction.

For each category, an overview of the methods was constructed in the form of tables.[†] The tables were composed by capturing salient information of DL-based sCT generation approaches, which has been schematically depicted in Figure 1.

[†] The tables presented in this review have been made publicly accessible at https://matteomaspero.github.io/overview_sct.

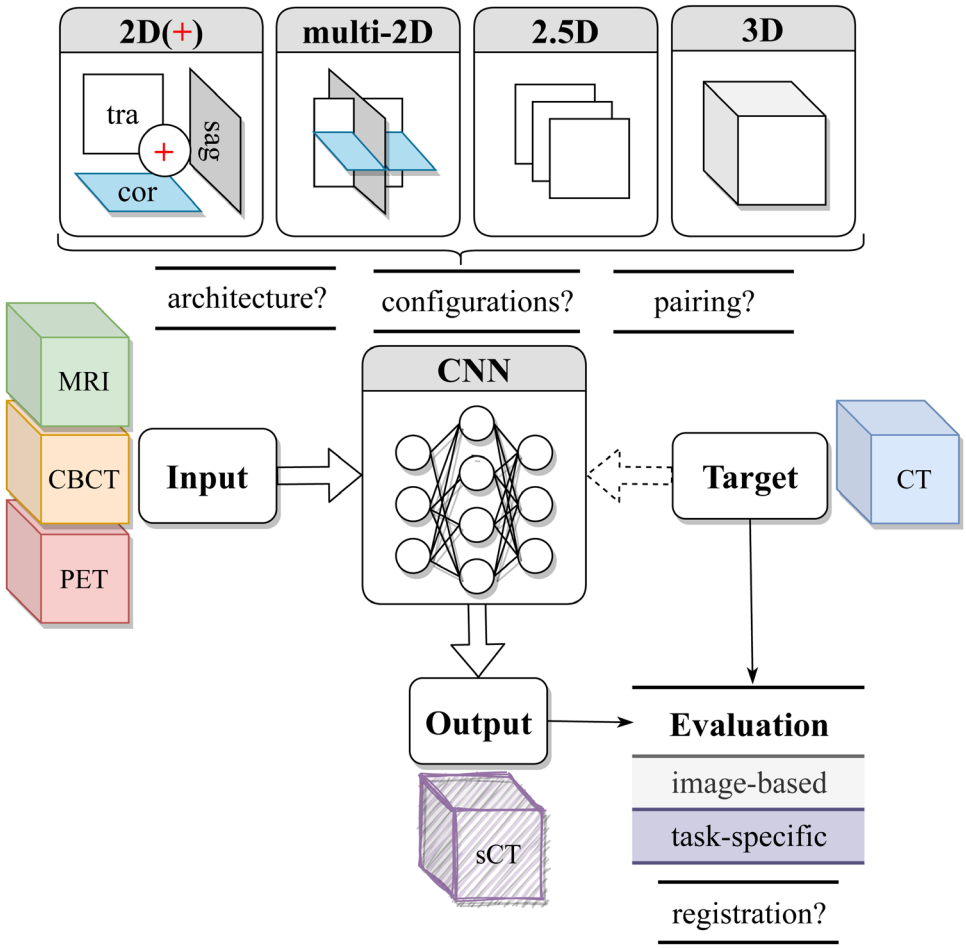

FIGURE 1 Schematic of deep learning-based sCT generation study. The input images/volumes, either being MRI (green), CBCT (yellow), or PET (red), are converted by a convolutional neural network (CNN) into sCT. The CNN is trained to generate sCT similar to the target CT (blue). Several choices can be made in terms of network architecture, configuration, data pairing. After the sCT generation, the output image/volume is evaluated with image- and task-specific metrics.

Independent of the input image, that is, MRI, CBCT, or PET, the chosen architecture (CNN) can be trained with paired or unpaired input data and different configurations. In this review, we define the following configurations: 2D (single slice, 2D, or patch) when training was performed considering transverse (tra), sagittal (sag), or coronal (cor) images; 2D+ when independently trained 2D networks in different views were combined during or after inference; multi-2D (m2D, also known as multiplane) when slices from different views, for example, transverse, sagittal, and coronal, were provided to the same network; 2.5D when training was performed with neighboring slices that were provided to multiple input channels of one network; 3D when volumes were considered as input (the whole volume, 3D, or patches). The architectures generally considered are introduced in the next section (Section 2.1). The sCTs are generated inferring on an independent test set the trained network or combining an ensemble of trained networks. Finally, the quality of the sCT can be evaluated with image-based or task-specific metrics (2.2).

For each of the sCT generation category, we compiled tables providing a summary of the published techniques, including the key findings of each study and other pertinent factors, here indicated the anatomic site investigated; the number of patients included; relevant information about the imaging protocol; DL architecture, the configuration chosen to sample the patient volume (2D or 2D+ or m-2D, 2.5D, or 3D); using paired/unpaired data during the network training; the radiation treatment adopted, where appropriate, along with the most popular metrics used to evaluate the quality of sCT (see 2.2).

The year of publication for each category was noted according to the date of the first online appearance. Statistics in terms of popularity of the mentioned fields were calculated with pie charts for each category. Specifically, we subdivided the papers according to the anatomical region they dealt with: abdomen, brain, head and neck (H&N), thorax, pelvis, and whole body; where available, tumor site was also reported. A discussion of the clinical feasibility of each methodology and observed trends follows.

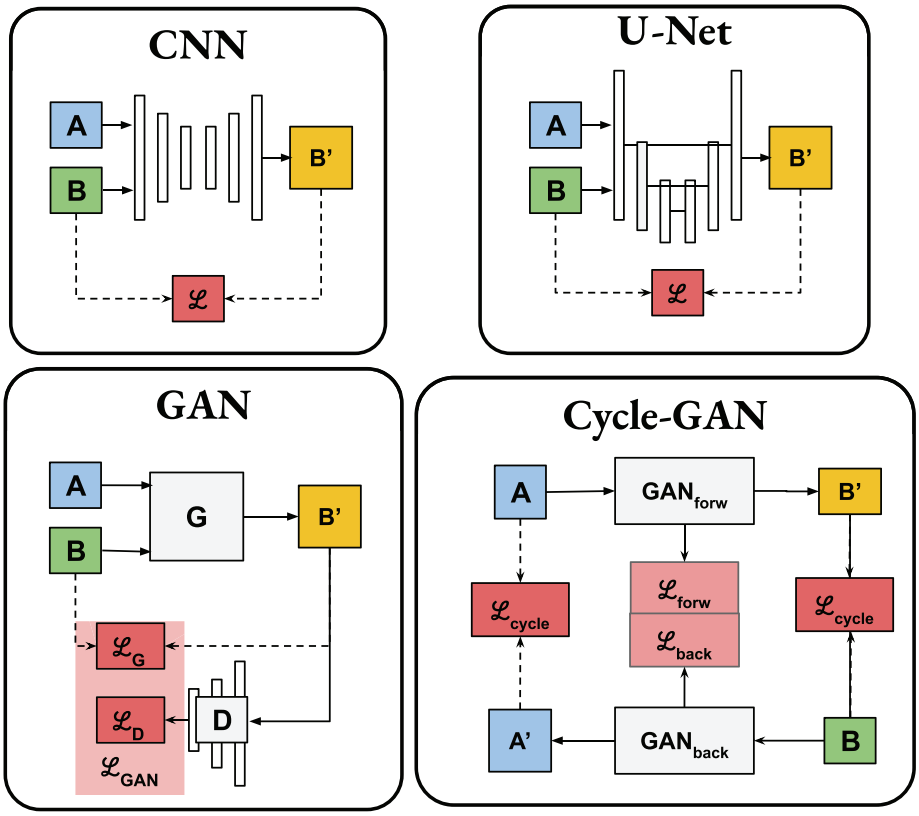

FIGURE 2 Deep learning architectures used for image-to-image translation. In the most straightforward configurations (CNN and U-net, top left and right, respectively), a single loss function between input and output images is computed. GANs (bottom) use more than one CNN and loss to train the generator's performance (G). Cycle-GANs enable unsupervised learning by employing multiple GANs and cycle-consistency losses ($\mathcal{L}_{cycle}$).

The most common network architectures and metrics will be introduced in the following sections to facilitate the tables' interpretation.

2.1 | Deep learning for image synthesis

Medical image synthesis can be formulated as an image-to-image translation problem, where a model that maps input image (A) to a target image (B) has to be found.[48] Among all the possible strategies, DL methods have dramatically improved state of the art.[49] DL approaches mainly used to synthesise sCT belong to the class of CNNs, where convolutional filters are combined through weights (also called parameters) learned during training. The depth is provided using multiple layers of filters.[50] The training is regulated by finding the "optimal" model parameters according to the search criterion defined by a loss function ($\mathcal{L}$). Many CNN-based architectures have been proposed for image synthesis, with the most popular being the U-nets[51] and generative adversarial networks (GANs)[52] (see Figure 2). U-net presents an encoding and a decoding path with additional skip connections to extract and reconstruct image features, thus learning to go from domains A to B. In the most simple GAN architecture, two networks are competing. A generator (G) that is trained to obtain synthetic images (B′) similar to the input set ($\mathcal{L}_G$), and a discriminator (D) that is trained to classify whether B′ is real or fake ($\mathcal{L}_D$), improving G's performances.

GANs learn a loss that combines both tasks resulting in realistic images.[53] Given these premises, many variants of GANs can be arranged, with U-nets being employed as a possible generator in the GAN framework. We will not detail all possible configurations since it is not the scope of this review, and we address the interested reader to References 54–56. A particular derivation of GAN, called cycle-consistent GAN (cycle-GAN), is worth mentioning. Cycle-GANs opened the era of unpaired image-to-image translation.[57] Here, two GANs are trained, one going from A to B′, called forward pass (forw), and the second going from B′ to A, called backward pass (back), are adopted with their related loss terms (Figure 2 bottom right). Two consistency losses $\mathcal{L}_c$ are introduced, aiming at minimizing differences between A and A′ and B and B′, enabling unpaired training.

2.2 | Metrics

An overview of the metrics used to assess and compare the reviewed publications' performances is summarised in Table 1.

TABLE 1 Overview of the most popular metrics reported in the literature subdivided into image similarity, geometric accuracy, task-specific metrics, and category

Category		Metric
Image similarity		M(A)E = $\frac{\sum_1^n \lvert CT_i - sCT_i \rvert}{n}$, with n=voxel number in ROI;
		(R)MSE = $(\sqrt{})\left(\frac{\sum_1^n (CT_i - sCT_i)^2}{n}\right)$
		PSNR = $10\log\left(\frac{MAX_{CT}^2}{MSE}\right)$
		SSIM = $\frac{(2\mu_{sCT}\mu_{CT}+c_1)(2\sigma_{sCT,rCT}+c_2)}{(\mu_{sCT}^2+\mu_{CT}^2+c_1)(\sigma_{sCT}^2+\sigma_{CT}^2+c_2)}$
		with
		$c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$
		μ = Mean, σ = variance/covariance
		L = Dynamic range, $k_1 = 0.01$ and $k_2 = 0.03$
Geometry accuracy		DSC(Seg_{CT}, Seg_{sCT}) = $2\frac{Seg_{sCT} \cap Seg_{CT}}{Seg_{sCT}+Seg_{CT}}$
Task-specific	MR-only and CBCT-to-CT	DD = $100 \cdot \frac{D_{sCT}-D_{CT}}{D_{CT}}$%, with D = dose;
		DPR = % of voxel with DD < x% in an ROI
		GPR = % of voxel with $\gamma < 1$ in an ROI
		DVH = difference of specific points in dose-volume histogram plot
	PET reconstruction	$PET_{\lvert err \rvert} = 100 \cdot \frac{\lvert PET_{sCT} - PET_{CT} \rvert}{PET_{CT}}$%

Abbreviations: CBCT, cone-beam computed tomography; CT, computed tomography; DD, dose difference; DPR: dose pass rate; DSC: Dice similarity coefficient; DVH, dose-volume histogram; GPR: gamma pass rate; ROI, region of interest; M(A)E, mean absolute error; MR, magnetic resonance; PET, positron emission tomography; PSNR, peak signal-to-noise ratio; (R)MSE, (root) mean squared error; SSIM, structural similarity index measure.

Image similarity: The most straightforward way to evaluate the quality of the sCT is to calculate the similarity of the sCT to the ground truth/target CT, on a voxel-wise basis. The calculation of voxel-based image similarity metrics implies that sCT and CT are aligned by translation, rigid (rig), affine or deformable (def) registrations. Widespread similarity metrics for this task are reported in Table 1 and include mean (absolute) error (M(A)E), sometimes referred to as mean absolute prediction error, peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). Other less common metrics are cross-correlation and normalised cross-correlation, along with the (root) mean squared error ((R)MSE).

M(A)E, (R)MSE are relatively easy to compute, and together with PSNR, are the most widely used fidelity measures. SSIM is a more sophisticated metric developed to take advantage of the known characteristics of the human visual system[58] perceiving the loss of image structure due to variations in lighting.

Geometric accuracy: Along with voxel-based metrics, the geometric accuracy of the generated sCT can also be assessed by comparing the volume and morphology of delineated structures or their corresponding binary masks. For example, the Dice similarity coefficient (DSC) is a common metric that assesses the accuracy of depicting specific tissue classes/structures, for example, bones, fat, muscle, air, and body. In this context, DSC is calculated after having applied a threshold to CT and sCT and, if necessary, morphological operations on the binary masks. Additionally, metrics generally used to estimate segmentation's accuracy can also be adopted as Hausdorff distance or mean absolute surface distance, which measures the maximum and average distance of two sets of contours.

Other image-based metrics can be subdivided according to the application and presented in the following sections' appropriate subcategory.

Task-specific metrics: In MR-only RT and CBCT-to-CT for adaptive RT, the accuracy of dose calculation on sCT is generally compared to CT-based in specific region of interests (ROIs) for dose calculations performed either for photon (x) and proton (p) RT.

The most common voxel-wise based metric is the dose difference (DD), calculated as the average dose (D_{CT} D_{sCT}) in ROIs as well as in the whole body, target or other structures of interest. The DD can be expressed as an absolute value (Gy) or relative (%), either to the prescribed dose, the maximum dose or the voxel-wise reference dose. The dose pass rate is directly correlated to DD, and it is calculated as the percentage of voxels with DD less than a set threshold.

Gamma (γ) analysis allows combining dose and spatial criteria[59] and it can be performed either in 2D or 3D. Several parameters need to be set to perform γ-analyses, including dose criteria, distance-to-agreement criteria, local or global analysis, and dose threshold. Interpretation and comparison between studies of gamma index results are challenging since they depend on the chosen parameters, dose grid size, and voxel resolution.[60,61] Results of γ-analysis are generally

expressed as gamma pass rate (GPR), counting the percentage of voxels with $\gamma < 1$ or the mean γ in an ROI generally defined based on a threshold of the reference dose distribution.

Dose-volume histograms (DVHs) are one of the most diffused tools in the clinical routine.[62] For the evaluation of sCT, generally, the differences among clinically relevant DVH points is reported.

In proton RT, range shift (RS) analysis is also performed. Here, the ideal range (known as prescribed range) is defined as the depth at which the dose has decreased to 80% of the maximum dose, on the distal dose fall-off (R_{80}).[63] RS error (RSe) can be defined both as the absolute difference between the prescribed and the actual range ($RSe = R_{80CT} - R_{80sCT}$) and as relative RS (%RS) error, expressed as the shift in % relative to the prescribed range, along the beam direction[64]

$$\%RS = \left| \frac{R_{80CT} - R_{80sCT}}{R_{80CT}} \right| \times 100. \quad (1)$$

For sCT for PET AC, the absolute and relative error of the PET reconstruction ($PET_{|err|}$ and PET_{err}, respectively) are usually reported along with the difference in standard uptake values (SUV).

Please note that differences could occur in the ROI where the metrics are calculated. For example, mean absolute error (MAE) can be computed on the whole predicted volume, in a volume of interest (VOI) or a cropped volume. In addition to that, the implementation of the metric computation can change. For example, $\gamma_{3\%,3\ mm}$ (γ_3), $\gamma_{2\%,2\ mm}$ (γ_2), and $\gamma_{1\%,1\ mm}$ (γ_1) can be calculated on ROI obtained from different dose thresholds and with 2D or 3D algorithms. In the following sections, we will highlight the possible differences speculating on the impact.

3 | RESULTS

Database searching led to 91 records on PubMed, 98 on Scopus, and 218 on Web of Science. After duplicates removal and content check, 83 eligible papers were found.

Figure 3 summarises the number of articles published by year, grouped in 51 (61.4%), 15 (18.1%), and 17 (20.5%) for MR-only RT (category I), CBCT-to-CT for adaptive RT (category II), and sCT for PET AC (category III), respectively. The first conference paper appeared in 2016.[46] Given that we excluded conference papers from our search, we found that the first work was published in 2017. In general, the number of articles increased over the years, except for CBCT-to-CT and sCT for PET AC, which was stable in the last years. Figure 3 shows that the brain, pelvis, and H&N were the most popular anatomical regions investigated in DL-based sCT for MR-only RT, covering ~80% of the studies. For CBCT-to-CT, H&N and pelvic regions were the most explored sites, being present in >75% of the works. Finally, for PET AC, H&N was investigated in the majority of the studies, followed by the pelvic region. Together, they covered >75% of the publications.

The total number of patients included in the analysis was variable, but most studies dealt with less than 50 patients for all three categories. The largest patient cohorts included 402[65] (I), 328[66] (II), and 193 patients[67] (I), while the smallest studies included 10 patients[68] and another 10 volunteers[69](I).

Most papers enrolled adult patients. Pediatric patients represent a more heterogeneous dataset for network training, and its feasibility has been investigated first for AC in PET[70] (79 patients) and more recently for photon and proton RT.[71,72]

All the models were trained to perform a regression task from the input to sCT, except for two studies where networks were trained to segment the input image into a predefined number of classes, thus performing a segmentation task.[73,74]

In most of the works, training was implemented in a paired manner, with unpaired training investigated in 13 of 83 articles. Four studies compared paired against unpaired.[67,75–77] The 2D networks were the most common over the three categories, being adopted about 61% of the times, 2D+ 6%, 2.5D 10%, and 3D configuration 24%. In some studies, multiple configurations were investigated, for example.[75,78,79] GANs were the most popular architectures (45 times), followed by U-nets (36) and other CNNs. Note that the GAN generator a U-net may be employed, but this counted as GAN.

All the investigations employed registration between sCT and CT to evaluate the quality of the sCT, except for Xu et al.[77] and Fetty et al.,[80] where metrics were defined to assess the quality of the sCT in an unpaired manner, for example, Frechet inception distance.

Main findings are reported in Table 2 for studies on sCT for MR-only RT without dosimetric evaluations, in Tables 3a and 3b for studies on sCT for MR-only RT with dosimetric evaluations, in Table 4 for studies on CBCT-to-CT for IGART, and in Table 5 for studies on PET AC. Tables are organised by anatomical site and tumor location where available. Studies investigating the independent training and testing of several anatomical regions are reported for each specific site.[66,77,81–83] Works using the same network to train or test data from different scanners and anatomy are reported at the bottom of the table.[84,85] Detailed results based on these tables are presented in the following sections subdivided for each category.

3.1 | MR-only radiotherapy

The first work ever published in this category, and in among all the categories, was by Han in 2017, where

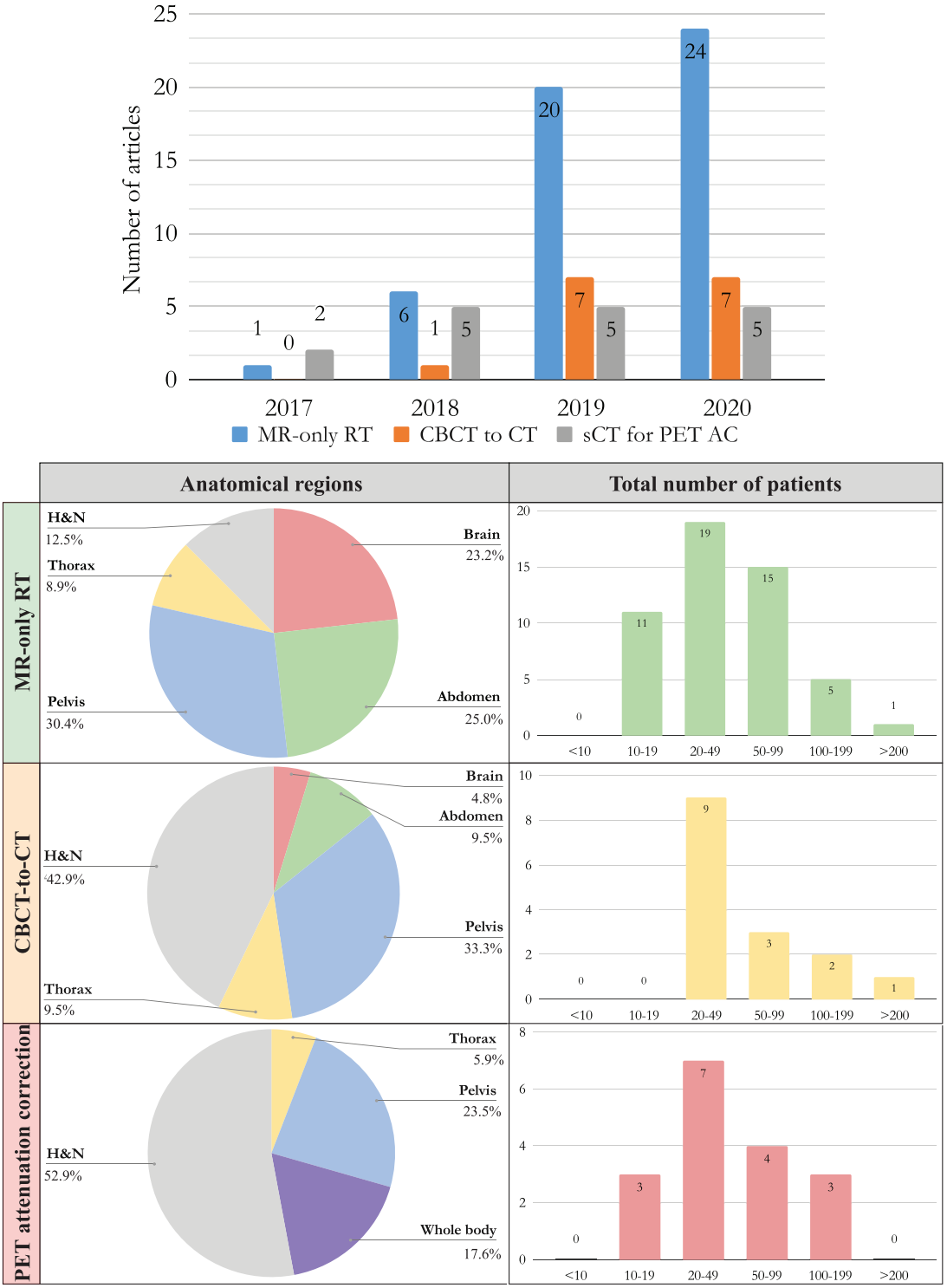

FIGURE 3 (Top) The number of published articles grouped by application and year; (b, left) pie charts of the anatomical regions investigated for each application; (bottom) bar plot of the publications binned per the total number of patients included in the study

he proposed to use a paired U-net for brain sCT generation. After 1 year, the first work published with a dosimetric evaluation was presented by Maspero et al.,[86] investigating a 2D-paired GAN trained on prostate patients and evaluated on prostate, rectal, and cervical cancer patients.

Considering the imaging protocol, we can observe that most of the MRIs were acquired at 1.5 T (51.9%), followed by 3 T (42.6%), and the remaining 6.5% at 1 T or 0.35/0.3 T. The most popular MRI sequences adopted depends on the anatomical site: T1 gradient recalled-echo (T1 GRE) for abdomen and brain; T2 turbo

TABLE 2 Overview sCT methods for MR-only radiotherapy with sole image-based evaluation

		Patients				MRI		DL method			Image similarity				
Tumor site		Train	val	Test	*x*-Fold	Field (T)	Sequence	conf	arch	Reg	MAE (HU)	PSNR (dB)	SSIM	Others	Reference
Abdomen	Abdomen	10[a]		10	LoO	n.a.	mDixon	2D pair	GAN*	def	61 ± 3			CC	Xu et al.[69]
	Abdomen	160			LoO	n.a.	n.a.	2D pair	GAN*	rig	5.1 ± 0.5		0.90 ± 0.43	(F/M)SIM IS …	Xu et al.[77]
Brain	Brain	18			6*x*	1.5	3D T1 GRE	2D pair	U-net	rig	85 ± 17			MSE, ME	Han[87]
	Brain	16			LoO	n.a.	T1	2.5Dp pair	CNN+	rig	85 ± 9	27.3 ± 1.1			Xiang et al.[81]
	Brain	15			5*x*	1.0	T1 Gd	2D pair	CNN	def	102 ± 11	25.4 ± 1.1	0.79 ± 0.03	Tissues	Emami et al.[88]
	Brain							GAN			89 ± 10	26.6 ± 1.2	0.83 ± 0.03	Tissues	Emami et al.[88]
	Brain	98CT		10		3	3D T2	2D	GAN	aff	19 ± 3	65.4 ± 0.9	0.25 ± 0.01		Jin et al.[89]
	Brain	84MR						pair/unp*							
	Brain	24			LoO	n.a.	T1	3Dp pair	GAN*	rig	56 ± 9	26.6 ± 2.3		NCC, HD body	Lei et al.[90]
	Brain	33			LoO	n.a.	T1[d]	2D unp	GAN*	No	9.0 ± 0.8		0.75 ± 0.77	(F/M)SIM IS …	Xu et al.[77]
	Brain	28[e]	2	15		1.5	n.a.	2D pair*	GAN*	aff	134 ± 12	24.0 ± 0.9	0.76 ± 0.02		Yang et al.[91]
	Brain	81		11	8*x*	1.5	3D T1 GRE	2D pair	U-net	aff	45.4 ± 8.5	43.0 ± 2.0	0.65 ± 0.05	Metrics for air	Massa et al.[92]
	Brain						3D T1 GRE Gd				44.6 ± 7.4	43.4 ± 1.2	0.63 ± 0.03	Air, bones	
	Brain						2D T2 SE				45.7 ± 8.8	43.4 ± 1.2	0.64 ± 0.03	Soft tissues	
	Brain						2D T2 FLAIR				51.2 ± 4.5	44.9 ± 1.2	0.61 ± 0.04	DSC bones	
	Brain	28		6		1.5	T2	2D pair	U-net	rig	65 ± 4	28.8 ± 0.6	0.972 ± 0.004	Same metrics for	Li et al.[76]
								2D unp	GAN		94 ± 6	26.3 ± 0.6	0.955 ± 0.007	Synthetic MRI	
Head and neck	Nasopharinx	23		10		1.5	T2	2D pair	U-net	def	131 ± 24			MAE ME tissue/bone	Wang et al.[93]
	H&N	28	4		8*x*	1.5	2D T1 ± Gd, T2	2D pair	GAN	aff	76 ± 15	29.1 ± 1.6	0.92 ± 0.02	DSC MAE bone	Tie et al.[94]
	H&N	60		30		3	T1	2D unp	GAN	n.a.	19.6 ± 0.7	62.4 ± 0.5	0.78 ± 0.2		Kearney et al.[95]
	H&N	7		8	LoO	1.5	3D T1, T2	2D pair	GAN	def	83 ± 49			ME	Largent et al.[96]
	H&N	10			LoO	1.5	3D T1, T2	2D pair	GAN*	def	42–62			RMSE, CC	Qian et al.[68]
	H&N	32		8	5*x*	3	3D UTE	2D pair	U-net	def	104 ± 21			DSC, spatial corr	Su et al.[97]

(Continues)

TABLE 2 (Continued)

Tumor site		Patients				MRI		DL method			Image similarity				Reference
		Train	val	Test	*x*-Fold	Field (T)	Sequence	conf	arch	Reg	MAE (HU)	PSNR (dB)	SSIM	Others	
Pelvis	Prostate	22			LoO	n.a.	T1	2.5Dp pair	CNN+	rig	43 ± 3	33.5 ± 0.8			Xiang et al.[81]
	Pelvis	20			LoO	n.a.	3D T2	3Dp pair	GAN*	rig	51 ± 16	24.5 ± 2.6		NCC, HD body	Lei et al.[90]
	Prostate	20			5*x*	1.5	2D T1 TSE	2D pair	U-net	def	41 ± 5			DSC bone	Fu et al.[78]
	Pelvis human							3Dp pair			38 ± 5				
	Pelvis canine	27			3*x*	3	3D T1 GRE	3Dp	U-net	def	32 ± 8	36.5 ± 1.6		MAE/DSC bone	Florkow et al.[98]
	Pelvis	18				1.5	mDixon[c]	pair			36 ± 4	36.1 ± 1.7		Surf dist <0.5 mm	
	Pelvis	15		4	5*x*	3	3D T2	2D pair	CNN	def	38 ± 6	29.5 ± 1.2	0.96 ± 0.01	ME, PCC	Bahrami et al.[99]
	Pelvis								U-net		43 ± 9	28.2 ± 1.6	0.95 ± 0.01		
	Pelvis	100				3	2D T2 FSE	2D unp	GAN	No				FID	Fetty et al.[80]
Thorax	Breast	14		2	LoO	n.a.	n.a.	2D pair	U-net[b]	def				DSC 0.74–0.76	Jeon et al.[73]

Abbreviations: arch, architecture; CC, cross-correlation; CNN, convolutional neural network; conf, configuration; CT, computed tomography; DL, deep learning; DSC, Dice score coefficient; FID, Frechet inception distance; FLAIR: fluid-attenuated inversion; (F/M)SIM, (feature/multi-scale structural) similarity; GAN: generative adversarial network; Gd: gadolinium; GRE: gradient recalled-echo; HD, Hausdorff distance; H&N, head and neck; IS, inception score; LoO, leave one out; mDixon, multicontrast Dixon reconstruction; MAE, mean absolute error; ME, mean error; MR, magnetic resonance; MRI, magnetic resonance imaging; (N)CC, normalised cross-correlation; PCC, Pearson correlation coefficient; PSNR, peak signal-to-noise ratio; (R)MSE, (root) mean squared error; SE, spin-echo; SSIM, structural similarity index; (T)SE, (turbo) spin-echo; val, validation; *x*-Fold, cross-fold.

aVolunteers, not patients.

bTo segment CT into five classes.

cMultiple combinations of Dixon images was investigated but omitted here.

dDataset from http://www.med.harvard.edu/AANLIB/.

eRobustness to training size was investigated.

* Multiple networks or architectures have been compared.

TABLE 3a Overview sCT methods for MR-only radiotherapy with image-based and dose evaluation

		Patients				MRI		DL method			Image similarity				Dose				
Tumor site		**Train**	**val**	**Test**	***x*-Fold**	**Field (T)**	**Sequence**	**conf**	**arch**	**Reg**	**MAE (HU)**	**PSNR (dB)**	**Others**	**Plan**	**DD (%)**	**GPR (%)**	**DVH**	**Others**	**Reference**
Abdomen	Liver	21			LoO	3	3D T1 GRE	3Dp pair	GAN	def	73 ± 18	22.7 ± 3.6	NCC	*p*		99.4 ± 1.0[b]	<1%	Range γ_2 γ_1	Liu et al.[100]
	Abdomen	12			4*x*	0.3	GRE	2D pair	GAN[a]	def	90 ± 19	27.4 ± 1.6		*x*	<±0.6	98.7 ± 1.5[c]	<±0.15	γ_3	Fu et al.[75]
						1.5		2D unp			94 ± 30	27.2 ± 2.2		$+B_0$	<±0.6	98.5 ± 1.6[c]			
	Abdomen	46		31	3*x*	3	3D T1 GRE	2.5D pair	U-net	syn rig	79 ± 18		MAE ME organs	*x*			<2 Gy		Liu et al.[101]
	Abdomen	39		19		0.35	GRE	2D pair	U-net	def	79 ± 18		ME	*x* + B_0	<0.1	98.7 ± 1.1[c]	<2.5%	γ_3 γ_1	Cusumano et al.[82]
	Abdomen	54	18	12	3*x*	1.5	3D T1 GRE	3Dp pair	U-net	def	62 ± 13	30.0 ± 1.8	ME, DSC tissues	*x*	<0.1	99.7 ± 0.3[c]	<2%	Beam	Florkow et al.[72]
	paed					3	GRE, T2 TSE							*p*	<0.5	96.2 ± 4.0[c]	<3%	Depth	
Brain	Brain	26			2*x*	1.5	3D T1 GRE	m2D[e] pair	CNN	rig	67 ± 11		ME tissues DSC dist body	*x*	−0.1 ± 0.3	99.8 ± 0.7[c]		Beam γ_3 Depth γ_1	Dinkla et al.[102]
	Brain	40		10		1.5	3D T1 GRE Gd	2D pair	CNN	def	75 ± 23		DSC	*x*	<0.2 ± 0.5	99.2[b]			LiuF et al.[103]
	Brain	54	9	14	5*x*	1.5	2D T1 SE Gd	2D pair	GAN	rig	47 ± 11		Each fold	*x*	−0.7 ± 0.5	99.2 ± 0.8[c]	<1%	2D/3D γ_3 γ_1	Kazemifar et al.[104]
	Brain	55	28	4		1.5	3D T1 GRE	2D pair 3Dp pair	U-net	rig	116 ± 26 137 ± 32		ME	*x* *p*		>98[c], 98 ± 2[c] >98[c], 97 ± 3[c]		Range γ_1	Neppl et al.[79]
	Brain	25	2	25		1.5	3D T1 GRE	3Dp pair	GAN	rig	55 ± 7		ME DSC	*x*	<2	98.4 ± 3.5[c]	<1.65%	Range γ_3 γ_1	Shafai et al.[105]
	Brain	47		13	5*x*	3	T1	2D pair	U-net	rig	81 ± 15		ME air, tissues	*x*	2.3 ± 0.1			Align CBCT	Gupta et al.[106]
	Brain	12	2	1	LoO	3	3D T1	2D+ pair	U-net	rig	54 ± 7		ME, DSC tissues	*p*	0.00 ± 0.01			Range	Spadea et al.[107]

(Continues)

TABLE 3a (Continued)

	Patients				MRI		DL method			Image similarity				Dose				
Tumor site	**Train**	**val**	**Test**	***x*-Fold**	**Field (T)**	**Sequence**	**conf**	**arch**	**Reg**	**MAE (HU)**	**PSNR (dB)**	**Others**	**Plan**	**DD (%)**	**GPR (%)**	**DVH**	**Others**	**Reference**
						GRE												
Brain	15			5*x*	n.a.	T1, T2	2Dp pair	GAN	def	108 ± 24		Tissues	*x*	0.7	99.2 ± 1.0[c]	<1%	Beam γ_3	Koike et al.[108]
						FLAIR[g]											Depth γ_1	
Brain	30[f,h]	10	20	3*x*	1.5	3D T1	2D+[a] pair	GAN[a]	rig	61 ± 14	26.7 ± 1.9	ME DSC	*x*	−0.1 ± 0.3	99.5 ± 0.8[c]	<1%	Beam	Maspero et al.[71]
paed					3	GRE ± Gd						SSIM	*p*	0.1 ± 0.4	99.6 ± 1.1[c]	<3%	Depth γ_3	
Brain	66		11	5*x*	1.5	2D T1	2D unp	GAN	rig	78 ± 11			*p*	0.3 ± 0.3	99.2 ± 1.0[c]	<3%	Beam γ_3	Kazemifar et al.[109]
						SE Gd											Depth γ_1	
Brain	242[f,h]	81	79		3	3D T1	3Dp pair	CNN	def	81 ± 22		Tissues	*x*	0.13 ± 0.13	99.6 ± 0.3[c]	<±0.15	γ_3	Andres et al.[65]
					1.5	GRE ± Gd		U-net		90 ± 21				0.31 ± 0.18	99.4 ± 0.5[c]			

Abbreviations: arch, architecture; CNN, convolutional neural network; conf, configuration; DL, deep learning; DRR, digitally reconstructed radiograph; DSC, Dice score coefficient; FLAIR: fluid-attenuated inversion; GAN: generative adversarial network; Gd: gadolinium; GRE: gradient recalled-echo; LoO, leave one out; LtO, MAE, mean absolute error; ME, mean error; MR, magnetic resonance; MRI, magnetic resonance imaging; NCC, normalised cross-correlation; p, proton plan; paed, pediatric; PSNR, peak signal-to-noise ratio; sCT, synthetic computed tomography; SE, spin-echo; SSIM, structural similarity index; TSE, turbo spin-echo; val, validation; *x*, photon plan; *x*-Fold, cross-fold.

aComparison with other architecture has been provided.

b$\gamma_{3\%,3\ mm} = \gamma_3$.

c$\gamma_{2\%,2\ mm} = \gamma_2$.

eTrained in 2D on multiple view and aggregated after inference.

fRobustness to training size was investigated.

gMultiple combinations (also ± Dixon reconstruction, where present) of the sequences were investigated but omitted.

hData from multiple centers.

TABLE 3b Overview sCT methods for MR-only radiotherapy with image-based and dose evaluation

		Patients				MRI		DL method			Image similarity				Dose				
Tumor site		**Train**	**val**	**Test**	***x*-fold**	**Field (T)**	**Sequence**	**conf**	**arch**	**Reg**	**MAE (HU)**	**PSNR (dB)**	**Others**	**Plan**	**DD (%)**	**GPR (%)**	**DVH**	**Others**	**Reference**
Pelvis	Prostate	36		15		3	T2 TSE	2D pair	U-net	def	30 ± 5		ME tissues	*x*	0.16 ± 0.09	99.4[c]	<0.2 Gy	γ_3 γ_1	Chen et al.[110]
	Prostate	39			4*x*	3	3D T2	2D pair	U-net	def	33 ± 8		ME DSC dist body	*x*	−0.01 ± 0.64	98.5 ± 0.7[c]	< 3%	γ_3 γ_1	Arabi et al.[111]
	Prostate	17			LoO	1.5	T2	3Dp unp	GAN[a]	rig	51 ± 17	24.2 ± 2.5	NCC, bone: dist, uniform	*p*	−0.07 ± 0.07	98 ± 6[c]	<1%	Range, γ_3 Peak, γ_1	Liu et al.[112]
	Prostate	25		14	3*x*	3	3D T2	2D pair	U-net[a]	def	34 ± 8		Tissues	*x*	<1%	99.2 ± 1[d]	<1%		Largent et al.[113]
							TSE		GAN[a]		34 ± 8		ME		<1%	99.1 ± 1[d]			
	Pelvis	11[h]		8		3	T2	2D pair	GAN[a]	def	49 ± 6		ME	*x*	0.7 ± 0.4	99.2 ± 1.0[c]	<1.5%		Boni et al.[114]
						1.5	TSE						organs						
	Pelvis	26	15	10+19[h]		0.35	3D T2	2.5D pair	GAN[a]	def	41 ± 4	31.4 ± 1	ME MSE	*x*	<±1		<1.5%		Fetty et al.[115]
						1.5/3							bone						
	Pelvis	39		14		0.35	GRE	2D pair	U-net	def	54 ± 12		Tissues	*x* + B_0	<0.5	99.0 ± 0.7[c]	<1%	γ_3 γ_1	Cusumano et al.[82]
	Rectum	46[h]		44		1.5	3D T2	2D pair	GAN	def	35 ± 7		ME	*x*	<±0.8	99.8 ± 0.1[c]	<1%	γ_3 γ_1	Bird et al.[116]
													bone						
Head & neck	H&N	34			3*x*	1.5	3D T2	3Dp	U-net	def	75 ± 9		ME	*x*	−0.07 ± 0.22	95.6 ± 2.9[c]		γ_3	Dinkla et al.[117]
							TSE	pair					DSC bone						
	H&N	15		12		3	T1	2Dp[a]	GAN[a]	def		68 ± 2	SSIM	*p*	<0.5	<98[c]	<0.5		Klages et al.[118]
							GRE	pair					RMSE						
	H&N	30		15		3	T1 ± Gd	2D pair	GAN[a]	rig	70 ± 12	29.4 ± 1.3	SSIM	*p*	−0.3 ± 0.2	97.8 ± 0.9[c]			Qi et al.[119]
							T2 TSE[g]		U-net		71 ± 12	29.2 ± 1.3	DSC, DRR		−0.2 ± 0.2	97.6 ± 1.3[c]			
	H&N	135[f]	10	28		3	3D T1	2D pair	GAN[a]	def	70 ± 9		ME, DSC	*x*	−0.1 ± 0.3	98.7 ± 1.0[c]	<1.5%	Beam	Peng et al.[67]
							GRE	2D unp			101 ± 8		tissues		0.1 ± 0.4	98.5 ± 1.1[c]	<1.5%	Depth	
	H&N	27			3*x*	3	3D T1	2D+	GAN	def	65 ± 4		ME	*p*	<±0.2	93.5 ± 3.4[c]	<1.5%	NTCP	Thummerer et al.[120]
							GRE	pair					DSC					RS γ_3	

(Continues)

TABLE 3b (Continued)

		Patients				MRI		DL method			Image similarity				Dose				
Tumor site		**Train**	**val**	**Test**	***x*-fold**	**Field (T)**	**Sequence**	**conf**	**arch**	**Reg**	**MAE (HU)**	**PSNR (dB)**	**Others**	**Plan**	**DD (%)**	**GPR (%)**	**DVH**	**Others**	**Reference**
Thorax	Breast	12[f]		18	LtO	1.5	3D GRE	2Dp[e]	GAN[a]	def	94 ± 11		NCC	*p*	<0.5	98.4 ± 3.5[c]		DRR	Olberg et al.[121]
							mDixon	pair			103 ± 15							dist bone	
		Multiple sites with one network																	
	Prostate	32		27		3	3D T1	2D pair	GAN	rig	60 ± 6		ME	*x*	−0.3 ± 0.4	99.4 ± 0.6[b]	<1%	γ_2	Maspero et al.[86]
	Rectum			18		1.5	GRE				56 ± 5				−0.3 ± 0.5	98.5 ± 1.1[b]			
	Cervix			14		1.5/3	mDixon				59 ± 6				−0.1 ± 0.3[a]	99.6 ± 1.9[b]			

Abbreviations: arch, architecture; CC, cross-correlation; CNN, convolutional neural network; conf, configuration; CT, computed tomography; DD: dose difference; DL, deep learning; DRR, digitally reconstructed radiograph; DSC, Dice score coefficient; DVH, dose-volume histogram; FID, Frechet inception distance; FLAIR: fluid-attenuated inversion; GAN: generative adversarial network; Gd: gadolinium; GRE: gradient recalled-echo; GRE, gradient echo; HD, Hausdorff distance; H&N, head and neck; LoO, leave-one-out; LtO, leave-two-out; mDixon, multicontrast Dixon reconstruction; ME, mean error; MR, magnetic resonance; MRI, magnetic resonance imaging; NCC, normalised cross-correlation; NTCP, normal tissue complication probability; PSNR, peak signal-to-noise ratio; RMSE, root mean squared error; SSIM, structural similarity index; (T)SE, (turbo) spin-echo; val, validation; *x*-fold, cross-fold; *x*, photon plan; *p*, proton plan.

aComparison with other architecture has been provided.

b$\gamma_{3\%,3\ mm} = \gamma_3$.

c$\gamma_{2\%,2\ mm} = \gamma_2$.

d$\gamma_{1\%,1\ mm} = \gamma_1$.

eTrained in 2D on multiple view and aggregated after inference.

fRobustness to training size was investigated.

gMultiple combinations (also ±Dixon reconstruction, where present) of the sequences were investigated but omitted.

hData from multiple centers.

TABLE 4 Overview sCT methods for adaptive radiotherapy with CBCT

		Patients				DL method			Image similarity					Dose					
Tumor site		**Train**	**val**	**Test**	***x*-Fold**	**conf**	**arch**	**Reg**	**MAE (HU)**	**PSNR (dB)**	**SSIM**	**Others**	**Plan**	**DD (%)**	**DPR (%)**	**GPR (%)**	**DVH**	**Others**	**Reference**
Abdomen	Pancreas	30			LoO	3Dp pair	GAN[a]	def	56.9 ± 13.8	28.8 ± 2.5	0.71 ± 0.03	NCC SNU	*x*				<1Gy		Liu et al.[131]
Thorax	Thorax	53		15		2D pair	GAN	def	94 ± 32			ME DSC HD tis	*x*		76.7 ± 17.3[c]	93.8 ± 5.9[c]	<2.6	γ_3	Eckl et al.[66]
	Brain	24			LoO	3Dp	GAN	rig	13 ± 2	37.5 ± 2.3		NCC	No						Harms et al.[83]
Pelvis	Pelvis	20				pair			16 ± 5	30.7 ± 3.7		SNU							
	Prostate	16		4	5*x*	2D pair	U-net	def		50.9	0.967	SNU RMSE	No						Kida et al.[124]
	Prostate	27	7	8		2D pair	U-net[a]	def	58			ME	*x*		>98.4[1]	99.5[c]		γ_1 γ_3 DPR_2	Landry et al.[132]
													p		88.5[b]	>96.5[c]		DPR_2 RS	
	Prostate	18		8	4*x*	2D ens unp	GAN[e]	rig	87 ± 5			ME	*x*		99.9 ± 0.3[c]		<±1.5%	DPR^1 γ_3	Kurz et al.[133]
													p		80.5 ± 5[c]	95.9 ± 2.0[c]	<1%	DPR^b RS	
	Prostate	16		4		2D pair	GAN[a]	rig				SSIM diffROI	No						Kida et al.[126]
	Pelvis	205		15		2D pair	GAN	def	42 ± 5			ME DSC HD tis	*x*		88.9 ± 9.3[c]	98.5 ± 1.7[c]	<1	γ_3	Eckl et al.[66]
	H&N	81	9	20		2D unp	GAN[a]	def	29.9 ± 4.9	30.7 ± 1.4	0.85 ± 0.03	RMSE phantom	*x*			98.4 ± 1.7[c] 96.3 ± 3.6[1]			Liang et al.[128]
	Nasophar	50	10	10		2D pair	U-net	rig	6-27			ME organs	*x*	0.2 ± 0.1		95.5 ± 1.6[1]	<1%		Li et al.[129]
	H&N	30	7	7		2D pair	U-net[a]	rig	18.98	33.26	0.8911	RMSE	No						Chen et al.[125]
	H&N	50[f]		10		2.5D pair	U-net	rig	49.28	14.25	0.85	SNR	No						Yuan et al.[127]
	H&N	22		11	3*x*	2D[d] pair	U-net	def	36 ± 6			ME DSC SNU	*p*	−0.1 ± 0.3		98.1 ± 1.2[c]		RS γ_3	Thummerer et al.[134]

(Continues)

TABLE 4 (Continued)

Tumor site		Patients				DL method			Image similarity				Plan	Dose					Reference
		Train	val	Test	*x*-Fold	conf	arch	Reg	MAE (HU)	PSNR (dB)	SSIM	Others		DD (%)	DPR (%)	GPR (%)	DVH	Others	
	H&N	30		14		2D pair	GAN	def	82.4 ± 10.6			ME	*x*	91.0 ± 5.3[c]			<1 Gy		Barateau et al.[130]
												tissues					<1%		
	H&N	25		15		2D pair	GAN	def	77.2 ± 16.6			ME DSC HD tis	*x*		91.5 ± 4.3[c]	95.0 ± 2.4[c]	<2.4	γ_3	Eckl et al.[66]
									Multiple sites with one network										
	H&N	15	8	10		2D unp[a]	GAN[a]	rig	53 ± 12	30.5 ± 2.2	0.81 ± 0.04	ME	*x*	0.1 ± 0.5		97.8 ± 1[c]	<2%	γ_3	Maspero et al.[84]
	Lung	15	8	10					83 ± 10	28.5 ± 1.6	0.78 ± 0.04			0.2 ± 0.9		94.9 ± 3[c]			
	Breast	15	8	10					66 ± 18	29.0 ± 2.1	0.76 ± 0.02			0.1 ± 0.4		92 ± 8[c]			
	Pelvis	135	15	15	10*x*	2.5D pair	GAN[a]	def	24 ± 5	20.1 ± 3.4			*x*	<1%					Zhang et al.[85]
	H&N			10					24 ± 4	22.8 ± 3.4			*p*					RS	Zhang et al.[85]

Abbreviations: arch, architecture; conf, configuration; CT, computed tomography; DD: dose difference; DL, deep learning; DSC, Dice score coefficient; DVH, dose-volume histogram; ens, ensemble; GAN: generative adversarial network; HD, Hausdorff distance; H&N, head and neck; LoO, leave one out; ME, mean error; NCC, normalised cross-correlation; *p*, proton plan; PSNR, peak signal-to-noise ratio; RMSE, root mean squared error; RS, range shift; SSIM, structural similarity index; val, validation; *x*-fold, cross-fold; *x*, photon plan.

aComparison with other architecture has been provided.

bDose pass rate (DPR) 1% or $\gamma_{1\%,1\ mm} = \gamma_1$.

cDPR 2% or $\gamma_{2\%,2\ mm} = \gamma_2$; DPR 3% or $\gamma_{3\%,3\ mm} = \gamma_3$.

dTrained in 2D on multiple view and aggregated after inference.

eDifferent nets were trained and the different outputs were weighted to obtain final sCT.

fRobustness to training size was investigated.

spin-echo (TSE) for pelvis and H&N. Unfortunately, for more than 10 studies, either sequence or magnetic field were not adequately reported.

Generally, a single MRI sequence is used as input. However, eight studies investigated using multiple input sequences or Dixon reconstructions [69,72,86,94,95,98,108,121] based on the assumption that more input contrast may facilitate sCT generation. A relevant aspect related to MRI is which kind on preprocessing is applied to the data before being fed to the network. Generally, intensity normalization techniques such as *z*-score,[122] percentile,[71,86] or range-based normalization, histogram matching,[75,78,81,94] or linear rescaling were applied.[107,123] However, techniques such as bias field,[65,75,78,80,81,87,90,91,94,96,100,101,105,108,111,118] intensity homogeneity[65,75,78,80,81,87,90,91,94,96,100,101,105,108,111] were also applied to minimise interpatient intensity variations.

Some studies compared the performance of sCT generation depending on the sequence acquired. For example, Massa et al. [92] compared sCT from the most adopted MRI sequences in the brain, for example, T1 GRE with (+Gd) and without gadolinium (−Gd), T2 SE and T2 fluid-attenuated inversion recovery, obtaining the lowest MAE and highest PSNR for T1 GRE sequences with Gd administration. Florkow et al.[98] investigated how the performance of a 3D patch based paired U-net was impacted by different combinations of T1 GRE images along with its Dixon reconstructions, finding that using multiple Dixon images is beneficial in the human and canine pelvis. Qi et al.[119] studied the impact of combining T1 (±Gd) and T2 TSE, obtaining that their 2D-paired GAN model trained on multiple sequences outperformed any model on a single sequence.

When focusing on the DL model configuration, we found that 2D models were the most popular ones, followed by 3D patch based and 2.5D models. Only one study adopted a multi-2D (m2D) configuration.[102] Three studies also investigated whether the impact of combining sCTs from multiple 2D models after inference (2D+) shows that 2D+ is beneficial compared to single 2D view.[71,107,118] When comparing the performances of 2D against 3D models, Fu et al.[78] found that a modified 3D U-net outperformed a 2D U-net; while Neppl et al.[79] 1 month later published that their 3D U-net underperformed a 2D U-net not only on image similarity metrics but also considering photon and proton DDs. These contradicting results will be discussed later. Paired models were the most adopted, with only ten studies investigating unpaired training.[67,75–77,80,89,91,95,109,112] Interestingly, Li et al. [76] compared a 2D U-net trained in a paired manner against a cycle-GAN trained in an unpaired manner, finding that image similarity was higher with the U-net. Similarly, two other studies compared 2D-paired against unpaired GANs, achieving slightly better similarity and lower DD with paired training in the abdomen[75] and H&N.[67] Mixed paired/unpaired training was proposed by Jin et al.[89] who found such a technique beneficial against either paired or unpaired training. Yang et al.[91] found that structure-constrained loss functions and spectral normalization ameliorated unpaired training performances in the pelvic and abdominal regions.

An interesting study on the impact of the directions of patch-based 2D slices, patch size and GAN architecture was conducted by Klages et al.[118] who reported that 2D+ is beneficial against a single view (2D) training, overlapping/nonoverlapping patches is not a crucial point, and that upon good registration training of paired GANs outperforms unpaired training (cycle-GANs).

If we now turn to the architectures employed, we can observe that GAN covers the majority of the studies (∼55%), followed by U-net (∼35%) and other CNNs (∼10%). A detailed examination of different 2D-paired GANs against U-net with different loss functions by Largent et al.[113] showed that U-net and GANs could achieve similar image- and dose-base performances. Fetty et al.[115] focused on comparing different generators of a 2D-paired GAN against the performance of an ensemble of models, finding that the ensemble was overall better than single models being more robust to generalization on data from different scanners/centers. When considering CNNs architectures, it is worth mentioning using 2.5D-dilated CNNs by Dinkla et al.,[102] where the m2D training was claimed to increase the robustness of inference in a 2D+ manner, maintaining a big receptive field and a low number of weights.

An exciting aspect investigated by four studies is the impact of the training size,[65,67,71,91,121] which will be further reviewed in the discussion section.

Finally, when considering the metric performances, we found that 21 studies reported only image similarity metrics, and 30 also investigated the accuracy of sCT-based dose calculation on photon RT (19), proton RT (8), or both (3). Two studies performed treatment planning, considering the contribution of magnetic fields,[75,82] which is crucial for MR-guided RT. Also, only four publications studied the robustness of sCT generation in a multicentric setting.[65,71,114,116]

Overall, DL-based sCT resulted in DD on average <1% and $\gamma_{2\%,2\ mm}$ GPR > 95%, except for one study.[120] For each anatomical site, the metrics on image similarity and dose were not always calculated consistently. Such aspect will be detailed in the next section.

3.2 | CBCT-to-CT generation

CBCT-to-CT conversion via DL is the most recent CT synthesis application, with the first paper published in 2018.[124] Some of the works (5 out of 15) focused only on improving CBCT image quality for

better IGRT.[83,124–127] The remaining 10 proved the validity of the transformation with dosimetric studies for photons,[66,71,101,128–131] protons,[120] and for both photons and protons.[85,132,133]

Only three studies investigated unpaired training[84,128,133]; in 11 cases, paired training was implemented by matching the CBCT and ground truth CT by rigid or deformable registration. In Eck et al.,[66] however, CBCT and CT were not registered for the training phase, as the authors claimed the first fraction CBCT was geometrically close enough to the planning CT for the network. Deformable registration was then performed for image similarity analyzing. In this work, the quality of contours propagated to sCT from CT was compared to manual contours drawn on the CT to assess each step of the IGART workflow: image similarity, anatomical segmentation and dosimetric accuracy. The network, a 2D cycle-GAN implemented on a vendor's provided research software, was independently trained and tested on different sites, H&N, thorax, and pelvis, leading to best results for the pelvic region.

Other authors studied training a single network for different anatomical regions. In Maspero et al.,[84] authors compared the performances of three cycle-GANs trained independently on three anatomical sites (H&N, breast, and lung) versus a single trained with all the anatomical sites together, finding similar results in terms of image similarity.

Zhang et al.[85] trained a 2.5D conditional GAN[57] with feature matching on a large cohort of 135 pelvic patients. Then, they tested the network on additional 15 pelvic patients acquired with a different CT scanner and 10 H&N patients. The network predicted sCT with similar MAE for both testing groups, demonstrating the potentialities to transfer pretrained models to different anatomical regions. They also compared different GAN flavors and U-net finding the latter statistically worse than any GAN configuration.

Three works tested unpaired training with cycle-GANs.[84,128,133] In particular, Liang et al.[128] compared unsupervised training among cycle-GAN, DCGAN,[135] and PGGAN[136] on the same dataset, finding the first to perform better both in terms of image similarity and dose agreement.

As regards the body region, most of the studies were focused on H&N and pelvic region. Liu et al.[131] investigated CBCT-to-CT in the framework of breath-hold stereotactic pancreatic RT, where they trained a 3D patch cycle-GAN introducing an attention gate (AG)[137] to deal with moving organs. They found that the cycle-GAN with AG performed better than U-net and cycle-GAN without AG. Moreover, the DL approach led to a statistically significant improvement in sCT versus CBCT, although some residual discrepancies were still present for this particular anatomical site.

3.3 | PET attenuation correction

DL methods for deriving sCT for PET AC have been published since 2017.[138] Two possible image translations are available in this category: (i) MR-to-CT for MRAC, where 14 papers were found; (ii) uncorrected PET-to-CT, with three published articles.

In the first case, most methods have been tested with paired data in H&N (nine papers) and the pelvic region (four papers) except Baydoun et al.[139] who investigated the thorax district. The number of patients used for training ranged between 10 and 60. Most of the MR images employed in these studies have been acquired directly through 3T PET/MRI hybrid scanners, where specific MR sequences, such as ultra-short echo time (UTE) and zero time echo (ZTE) are used to enhance short T_2 tissues, such as in the cortical bone and Dixon reconstruction is employed to derive fat and water images.

Leynes et al.[138] compared the Dixon-based sCT versus sCT predicted by U-net receiving both Dixon and ZTE. Results showed that DL prediction reduced the root mean squared error (RMSE) in corrected PET SUV by a factor of 4 for bone lesions and 1.5 for soft tissue lesions. Following this first work, other authors showed the improvement of DL-based AC over the traditional atlas-based MRAC proposed by the vendors,[70,139–144] also comparing several network configurations.[145,146]

Torrado et al.[142] pretrained their U-net on 19 healthy brains acquired with T_1 GRE MRI and, subsequently, they trained the network using Dixon images of colorectal and prostate cancer patients. They showed that pretraining led to faster training with a slightly smaller residual error than U-net weights' random initialization.

Pozaruk et al.[145] proposed data augmentation over 18 prostate cancer patients by perturbing the deformation field used to match the MR/CT pair for feeding the network. They compared the performance of GAN with augmentation versus (1) Dixon based and (2) Dixon + bone segmentation from the vendor, (3) U-net with and (4) without augmentation. They found significant differences between the 3 DL methods and classic MRAC routines. GAN with augmentation performed slightly better than the U-net with/without augmentation, although the differences were not statistically relevant.

Gong et al.[146] used unregistered MR/CT pair for a 3D patch cycle-GAN, comparing the results versus atlas-based MRAC and CNN with registered pair. Both DL methods performed better than atlas MRAC in DSC, MAE, and $PET_{(err)}$. No significant difference was found between CNN and cycle-GAN. They concluded that cycle-GAN has the potentiality to skip the limit of using a perfectly aligned dataset for training. However, it requires more input data to improve output.

Baydoun et al.[139] tried different network configurations (VGG16,[153] VGG19,[153] and ResNet[154]) as a benchmark with a 2D conditional GAN receiving either

two Dixon input (water and fat) or four (water, fat, in-phase, and opposed-phase). The GAN always performed better than VGG19 and ResNet, with more accurate results obtained with four inputs.

In the effort to reduce the time for image acquisition and patient discomfort, some authors proposed to obtain the sCT directly from diagnostic images, T_1- or T_2-weighted, both using images from stand-alone MRI scanners[111,147,149] or hybrid machines.[74] In particular, Bradshaw et al.[74] trained a combination of three CNNs with T_1 GRE and T_2 TSE MRI (single sequence or both) to derive an sCT stratified in classes (air, water, fat, and bone), which was compared with the scanner default MRAC output. The RMSE on PET reconstruction computed on SUV and was significantly lower with the DL method and T_1/T_2 input. However, recently, Gong et al.[144] tested on a brain patient cohort a CNN with either T_1 or Dixon and multiple echo UTE (mUTE) as input. The latter overperformed the others. Liu et al.[147] trained a CNN to predict CT tissue classes from diagnostic 1.5 T T_1 GRE of 30 patients. They tested on ten independent patients of the same cohort, whose results are reported in Table 5 in terms of DSC. Then, they predicted sCT for five patients acquired prospectively with a 3 T MRI/PET scanner (T_1 GRE), and they computed the PET_{err}, resulting <1%. They concluded that DL approaches are flexible and promising to be applied to heterogeneous datasets acquired with different scanners and settings.

DL methods have also been proposed to estimate sCT from uncorrected PET. Due to the more considerable number of single PET exams, these methods have been tested on the full-body acquisitions and larger patient populations (up to 100 for training and 39 for testing). Although the global MAE is higher than site-specific MR-to-CT studies (about 110 HU vs. 10–15 HU), PET_{err} is below 1% on average, demonstrating the validity of the approach for the scope of PET AC.

4 | DISCUSSION

This review encompassed DL-based approaches to generate sCT from other RT-imaging modalities, focusing on published journal articles. The research topic was earlier introduced at conferences from 2016.[46] Since 2016, we have observed increasing interest in using DL for sCT generation. DL methods' success is probably related to the growth of available computational resources in the last decade that allowed training large volume datasets[50] achieving fast image translation, i.e. in the order of a few seconds.[155] Fast image-to-image translation facilitates applying DL in clinical cases and demonstrates its feasibility for clinical scenarios. In this review, we considered three clinical purposes for deriving sCT from other image modality, which are discussed in the following:

I. *MR-only RT*. The generation of sCT for MR-only RT with DL is the most populated category. Its 51 papers demonstrate the potential of using DL for sCT generation from MRI. Several training techniques and configurations have been proposed. For anatomical regions, as pelvis and brain/H&N, high image similarity and dosimetric accuracy, that is, DDs < 1%, can be achieved for photon RT and proton therapy. In region strongly affected by motion,[156,157] for example, abdomen and thorax, the first feasibility studies seem to be promising.[72,75,82,112,121] However, no study proposed the generation of DL-based 4D sCT yet, as from classical methods.[158] An exciting application is the DL-based sCT generation for the pediatric population,[71,72] which is considered more radiation-sensitive than an adult population[159] and could enormously benefit from MR-only, especially when patients' simulations are repeated.[19]
The geometric accuracy of sCT needs to be thoroughly tested to enable the clinical adoption of sCT for treatment planning purposes, primarily when MRI or sCT are used to substitute CT for position verification purposes. So far, the number of studies that investigated such an aspect from DL-based sCT is still scarce. Only Gupta et al.,[106] for brain, and Olberg et al.,[121] for breast cancer, have investigated this aspect assessing the accuracy of alignment based on CBCT and digitally reconstructed radiography, respectively. Future studies are required to strengthen the clinical use of sCT, especially considering that geometric accuracy has been already extensively investigated for sCT generated with classical methods for 3 T and below.[160–162]
DL-based sCT generation in the context of MR-guided RT[20,163–167] may reduce the treatment time, facilitating for daily image guidance and plan adaptation based on sole MRI.[168,169] For this application, the accuracy of dose calculation in the magnetic field's presence must be assessed before clinical implementation. So far, the studies investigating this aspect are still few, for example, for abdominal[75] and pelvic tumors[82] and only considered low magnetic fields. Recently, Groot Koerkamp et al.[170] published the first dosimetric evaluation of DL-based sCT for high magnetic field MR-guided RT achieving DDs < 1% for breast cases. The results are promising, but we advocate for further studies on additional anatomical sites and magnetic field strengths.

II. *CBCT-to-CT for image-guided (adaptive) radiotherapy*. In-room CBCT imaging is widespread in photon and proton RT for daily patient setup.[171] However, CBCT is not commonly exploited for daily plan adaptation and dose recalculation due to the artifacts associated with scatter and reconstruction algorithms that affect the quality of the electron density predicted by CBCT.[172] Traditional methods to cope with this issue have been based on image

TABLE 5 Overview methods on sCT for PET AC

Region	Patients				MRI		DL method			Image similarity		PET-related				Reference
	Train	val	Test	*x*-Fold	Field	Contrast	conf	arch	Reg	MAE (HU)	DSC	Tracer	PET_{err} (%)	Others		
Pelvis	10		16		3[d]	Dixon ± ZTE	3Dp pair	U-net	def			^{18}F-FDG		RMSE		Leynes et al.[138]
												^{68}Ga-PSMA		SUV diff		
Pelvis	15		4	4	3[d]	T1 GRE[b]	2D pair	U-net	def			^{18}F-FDG	1.8 ± 2.4	μ-map diff		Torrado et al.[142]
						Dixon							1.7 ± 2.0[j]			
													1.8 ± 2.4[i]			
													3.8 ± 3.9[h]			
Pelvis	12		6		3[d]	T1 GRE[n]	3Dp pair	CNN[l]	def		0.99 ± 0.00[i]	^{18}F-FDG		RMSE		Bradshaw et al.[74]
						T2 TSE					0.48 ± 0.21[g]					
											0.94 ± 0.01[j]					
											0.88 ± 0.03[k]					
											0.98 ± 0.01[i]					
Prostate	18		10		3[d]	Dixon	2D pair	GAN[a]	def			^{68}Ga-PSMA		SSIM		Pozaruk et al.[145]
													0.75 ± 0.64[e]	μ-map diff		
													0.52 ± 0.62[f]			
Head	30		10		1.5	T1 GRE	2D pair	CNN[l]	def		0.971 ± 0.005[g]	n.a.	-0.7 ± 1.1^{pet}			Liu et al.[147]
			5^{pet}			Gd					0.936 ± 0.011[i]					
											0.803 ± 0.021[h]					
Head	30[b] + 6		8		1.5[b]+3[d]	UTE	2D pair	U-net[l]	def		0.76 ± 0.03[g]	^{18}F-FDG	<1			Jang et al.[141]
											0.96 ± 0.01[i]					
											0.88 ± 0.01[h]					
H&N	32		8	5	3[d]	Dixon	2D pair	U-net	rig	13.8 ± 1.4	0.76 ± 0.04[h]	^{18}F-FDG	<3			Gong et al.[140]
	12		2	7		± ZTE				12.6 ± 1.5	0.80 ± 0.04[h]					

(Continues)

TABLE 5 (Continued)

Region	Patients				MRI		DL method			Image similarity		PET-related		Others	Reference
	Train	val	Test	*x*-Fold	Field	Contrast	conf	arch	Reg	MAE (HU)	DSC	Tracer	PET_{err} (%)		
Head	60		19	4	3[d]	mDixon	3Dp	U-net	rig		0.90 ± 0.07[m]	^{18}F-FET		biol tumor	Ladefoged et al.[70]
paed						+UTE	pair							vol, SUV	
Head	40			2	3	T1 GRE	3Dp	GAN	def	101 ± 40	0.80 ± 0.07[h]	^{18}F-FDG	3.2 ± 3.4	rel vol dif	Arabi et al.[148]
							pair			302 ± 79[h]			1.2 ± 13.8[h]	surf dist ME	
										407 ± 228[g]			3.2 ± 13.6[i]	RMSE PSNR	
										10 ± 5[i]			3.2 ± 13.6[g]	SSIM SUV	
Head	44	11	11		1.5	T1 GRE	2.5D pair	U-net	rig			^{11}C-WAY	−0.49 ± 1.7	synt μ-map,	Spuhler et al.[149]
												^{11}C-DASB	−1.52 ± 0.73	kin anal	
Head	23		47		3[d]	ZTE	3Dp pair	U-net	def		0.81 ± 0.03[h]	^{18}F-FDG	−0.2 ± 5.6	Jac	Blanc-Durand et al. [143]
Head	32			4	3[d]	Dixon[n]	3Dp pair	GAN[a]	def	15.8 ± 2.4%	0.74 ± 0.05[h]	^{18}F-FDG	−1.0 ± 13	SUV	Gong et al. [146]
Head	35			5	3	mDixon	2.5D pair	U-net	rig	10.94 ± 0.01%	0.87 ± 0.03[h]	^{11}C-PiB	<2		Gong et al.[144]
						UTE[n]						^{18}F-MK_{6240}			
Thorax	14			LoO	3[d]	Dixon[n]	2D pair	GAN[a]	def	67.45 ± 9.89		^{18}F-NaF		PSNR SSIM RMSE	Baydoun et al.[139]
								Other than MR-based sCT							
Body	100		28		PET, no att corrected		2D pair	U-net	Y[o]	111 ± 16	0.94 ± 0.01[h]	^{18}F-FDG	−0.6 ± 2.0	abs err	Liu et al.[150]
Body	80		39		PET, no att corrected		3Dp pair	GAN	Y[o]	109 ± 19	0.87 ± 0.03[h]	^{18}F-FDG	<1.0	NCC PSNR ME	Dong et al.[151]
Body	100		25		PET, no att corrected		2.5D pair	GAN	Y[o]			^{18}F-FDG	−0.8 ± 8.6	SUV ME	Armanious et al.[152]

aComparison with other architecture has been provided.
bData from another MRI sequence used as pretraining.
dMRI data from hybrid PET/MRI scanner.
eIn SUV max.
fIn SUV mean.
gIn air or bowel gas.
hIn the bony structures.
iIn the soft tissue.
jIn the fatty tissue.
kIn water.
lTrained to segment the CT/sCT into classes,
mExpressed in terms of Jaccard index and not DSC,
nMultiple combinations (also ±Dixon reconstruction, where present) of the sequences were investigated but omitted.
oIntrinsically registered: PET-CT data.

registration,[173,174] on scatter correction,[175] look-up-table to rescale HU intensities,[176] and histogram matching.[177] DL's introduction for converting CBCT to sCT has substantially improved image quality leading to faster results than image registration and analytical corrections.[134] Speed is crucial for the translation of the method into the clinical routine. However, one of the problems arising in CBCT-to-CT conversion for clinical application, is the different field of view (FOV) between CBCT and CT. Usually, the training is performed by registering, cropping, and resampling the volume to the CBCT size that is smaller than the planning CT.

Nonetheless, for replanning purposes, the limited FOV may hinder calculating the plan to the sCT. Some authors have proposed to assign water equivalent density within the CT body contour for the missing information.[130] In other cases, the sCT patch has been stitched to the planning CT to cover the entire dose volume.[84] Ideally, appropriate FOV coverage should be employed when recalculating the plan for online adaptive RT. Besides the dosimetric aspect, improved image quality may increase accuracy during image guidance for patient set-up and OAR segmentation. These are necessary steps for online adaptive RT especially for anatomical sites prone to large movements, as speculated by Liu et al.[131] in the framework of pancreatic treatments.

CBCT-to-CT has been proved both for photon and proton RT. For proton RT, the setup accuracy and dose calculation are even more relevant to avoid RS errors that could jeopardise the benefit of treatment.[63] Because there is an intrinsic error in converting HU to relative proton stopping power,[178] it has been shown that DL methods can translate CBCT directly to stopping power.[179] This approach has not been covered in this review, but it is an exciting approach that will probably lead to further investigations.

Interestingly, increasing the quality of CBCT can be tackled not only as an image-to-image translation problem but also as an inverse problem, that is, from a reconstruction perspective. Specifically, by having the raw data measurements (projections), DL could be used to improve tomography. In this sense, many investigations have been proposed, but they are considered out of the scope this review. For the interested reader, we suggest the following resources.[180–184] Currently, it is unclear whether formulating (CB)CT quality enhancement as a synthesis or reconstruction problem would be beneficial. First attempts showed that training convolutional networks for reconstruction enhanced their generalization capability to other anatomy[185]; however, research on such aspects is still ongoing.

III. *PET attenuation correction*. The sCT in this category is obtained either from MRI or from uncorrected PET. In the first case, the work's motivation is to overcome the current limitations in generating attenuation maps (μ-maps) from MR images in MRI/PET hybrid acquisitions that miscalculated the bone contribution.[186] In the second case, the limits to overcome are different: (i) to avoid extra-radiation dose when the sole PET exam is required, (ii) to avoid misregistration errors when stand-alone CT and PET machines are used, and (iii) to be independent of the MR contrast in MRI/PET acquisitions. Besides the network configuration, MRI used for the input, or the number of patients included in the studies, DL-based sCT have consistently outperformed current MRAC methods available on commercial software. The results of this review support the idea that DL-based sCT will substitute current AC methods, being also able to overcome most of the limitations mentioned above. These aspects seem to contradict the stable number of papers in this category in the last three years. Nonetheless, we have to consider that the recent trend has been to directly derive the μ-map from uncorrected PET via DL. Because this review considered only image-to-CT translation, these works were not included, but they can be found in a recent review by Lee.[47] However, it is worth mentioning a recent study from Shiri et al.,[187] where the largest patient cohort ever (1150 patients split in 900 for training, 100 for validation and 150 for test) was used for the scope. Direct μ-map prediction via DL is an auspicious opportunity that may direct future research efforts in this context.

4.1 | Deep-learning considerations and trends

The number of patients used for training the networks is quite variable, ranging from a minimum of 7 (in I)[68] to a maximum of 205 (in II),[66] and 242[65] (in I). In most of the cases, the patient number is limited to the availability of training pairs. Data augmentation is performed as linear and nonlinear transformation[188] to increase the training accuracy, as demonstrated in Pozaruk et al.[145] However, few publications investigated the impact of increasing the training size,[65,67,71,121,127] finding that image similarity increases when training up to 50 patients. This investigation can indicate the minimum amount of patients necessary to include in the training to achieve the state of the art performances. The optimal patient number may also depend on the anatomical site and its inter- and intrafraction variability. Besides, attention should be dedicated to balancing the training set, as performed in Reference 65, 71. Otherwise, the network may overfit, as previously demonstrated for segmentation tasks.[189]

GANs were the most popular architecture, but we cannot conclude that it is the best network scheme for

sCT. Indeed, some studies compared U-net or other CNN versus GAN finding GAN performing statistically better[85,139]; others found similar results[145,146] or even worse performances.[76,144] We can speculate that, as demonstrated by Largent et al.,[113] a vital role is played by the loss function, which, despite being the effective driver for network learning, has been investigated less than the network architecture, as highlighted for image restoration.[190] Another important aspect is the growing trend, except category III, in unpaired training (five and seven papers in 2019 and 2020, respectively). The quality of the registration when training in a paired manner influences the quality of DL-based sCT generation.[122] In this sense, unpaired training offers an option to alleviate the need for well-matched training pairs. When comparing paired versus unpaired training, we observed that paired training leads to slightly better performances. However, the differences were not always statistically significant.[67,76,91] As proposed by Yang et al.,[91] unsupervised training decreases the semantic information from one domain to another.[91] Such an issue may be solved by introducing a structure-consistency loss, which extracts structural features from the image defining the loss in the feature space. Yang et al.'s results showed improvements in this sense relative to other unsupervised methods. They also showed that preregistering unpaired MR-CT further improves unsupervised training results, which can be an option when input and target images are available, but perfect alignment is not achievable. In some cases, unpaired training even demonstrated to be superior to paired training.[191] A trend lately emerged is the use of architecture initially thought for unpaired training, for example, cycle-GAN to be used for paired training.[83,90]

Focusing on the body sites, we observed that most of the investigations were conducted in the brain, H&N, and pelvic regions. Simultaneously, fewer studies are available for the thorax and the abdomen, representing a more challenging patient population due to the organ motion.[192]

In MR-only RT results, we found contradicting results regarding the best performing spatial configuration for the papers that directly compared 2D versus 3D training.[78,79] It is undoubtedly clear that 2D+ increases the sCT quality compared to a single 2D views, as demonstrated in Spadea et al.[107] and Maspero et al.[71]; however, when comparing 2D against 3D training, patch size is a vital aspect.[118] 3D deep networks require a more significant number of training parameters than 2D networks.[193] For sCT generation, the approaches adopted have chosen to use patch size much smaller than the whole volume, probably hindering the contextual information considered. Generally, downsampling approaches have been proposed to increase the network's perceptive field, for example, for segmentation tasks,[194] but they have not been applied to sCT generation. We believe this will be an exciting area of research.

For what concerns the latest development from the DL perspective, in 2018, Oktay et al.[137] proposed a new mechanism, called AG, to focus on target structures that can vary in shape and size. Liu et al.[131] incorporated the AG in the generator of a cycle-GAN to learn organ variation from CBCT-CT pairs in the context of pancreas adaptive RT, showing that its contribution significantly improved the predictions compared to a network without AG. Other papers also adopted attention.[91,95] Embedding has also been proposed to increase the network's expressivity of the network and applied by Xiang et al.[81] (I). As AG's mechanism is a way to focus the *attention* on specific portions of the image, it can potentially open the path for new research topics. In 2019, Schlemper and colleagues[195] evaluated the AG for different tasks in medical image processing: classification, object detection, segmentation. So, we can envision that in the online IGART, such a mechanism could lead to multitask applications, such as deriving sCT, while delineating the structure of interests.

4.2 | Benefits and challenges for clinical implementations

DL-based sCT generations may reduce the need for additional or nonstandard MRI sequences, for example, UTE or ZTE, which it could turn in shorten the total acquisition time and speed up the workflow or increase patient throughput. As already mentioned, speed is particularly interesting for MR-guided RT as well as for adaptive RT in II, where it is considered crucial for on line correction. For what concern categories II and III, the generation of DL-based sCT possibly enables dose decreasing during imaging by reducing the need for CT in case of anatomical changes (in II) or by possibly diminishing the amount of radioactive material injected (in III).

Finally, it is worth commenting on the current status of the clinical adoption of DL-based sCT. We could not find that any of the methods considered are now clinically implemented and used. We speculate that this is probably related to the fact that the field is still relatively young, with the first publications only from 2017 and that time for clinical implementations generally last years, if not decades.[196,197] Additionally, as already mentioned, for categories I/II, the impact of sCT for position verification still needs to be thoroughly investigated. The implementation may also be more comfortable for category III if the methods would be directly integrated into scanners. In general, the involvement of vendors may streamline the clinical adoption of DL-based sCT. In this sense, we can report that vendors are currently active in evaluating their methods in research settings, e.g. for brain,[65] pelvis[116] in I, and for H&N, thorax, and pelvis in II.[66] In the last month, Palmer et al.[198] also reported using a prereleased version of a DL-based sCT

generation approach for H&N in MR-only RT. Another essential aspect that needs to be satisfied is the compliance to the currently adopted regulations,[199] where vendors can offer vital support.[200,201]

A key aspect of clinical implementation is the precise definition of a DL-based solution's requirements before being accepted. If we consider the reported metrics, we cannot find uniform criteria for reporting. Multiple metrics have been defined, and it is not clear on which ROIs they should be computed. For example, the image-based similarity was reported on the body contour or in tissues generally defined by different thresholds; for task-specific metrics, the methods employed are even more heterogeneous. For example, in I and II, GPRs can be performed in 2D, 3D, and different dose thresholds level have been employed, for example, 10%, 30%, 50%, or 90% of the prescribed or maximum dose. In III, the $PET_{|err|}$ can be computed either on SUV, max SUV, or in larger VOI, making it difficult to compare different network configurations' performances. We think that this lack of standardization in reporting the results is also detrimental to clinical adoption. A first attempt at revising the metrics currently adopted has been performed by Liesbeth et al.[202] However, this is still insufficient, considering the differences in how such metrics can be calculated and reported. In this sense, we advocate for consensus-based requirements that may facilitate reporting in future clinical trials.[203] Also, no public datasets arranged in the form of grand challenges (https://grand-challenge.org/) are available to enable a fair and open evaluation of different approaches.[204]

To date, four scientific studies have already investigated the performance of DL-based sCT in a multicenter setting.[71,114–116] These studies have been reported only for MR-only RT. Future work should focus on assessing the performance of DL-based sCT generation for II and III. On the contrary, investigations on sCT generation with classical methods using multicenter data are more diffuse for all the three categories.[26,205–209] Of particular relevance when considering the generalization of a DL model for sCT generation may be the application of transfer learning.[210,211] Mainly, transfer learning may be exploited to facilitate fine-tuning a model pretrained on a specific MRI contrast or CBCT image protocols; or generalise among multiple anatomies. No paper was found up to December 2020 investigating this aspect, but it could be an exciting research area. More recently, Li et al.[212] showed that transfer learning facilitated training a DL model on different MRI contrasts for sCT generation.

The quality of sCT cannot be judged by a user, except when its quality is inferior. Therefore, software-based quality assurance (QA) procedures should be put in place. It could be pretty interesting to have a phantom to allow regular QA procedures, for example, for CT.[213] This would be relatively straightforward for II; however, in MR-based sCT, phantoms' manufacturing is quite challenging due to the need for contrast in MRI and CT. Recently, the first phantoms have been proposed for such task[214–217] showing the potential of additive manufacturing.

Alternatively, it would be relevant if a CNN could automatically generate a metric to assess the quality of sCTs, for example, already presented for automatic segmentation.[218] In this sense, Bragman et al.[219] introduce the use of uncertainty for such a task, by adopting a multitask network and a Bayesian probabilistic framework. More recently, two other works proposed to use uncertainty either from the combination of independently trained networks[71] or via dropout-based variational inference.[220] So far, the field of uncertainty estimation with DL[221] has been superficially touched for sCT generation. It would be interesting to see future work focusing on developing criteria for automatically identifying of failure cases using uncertainty prediction. Patients with inaccurate sCTs will be flagged for CT rescan or manual adjustment of the sCT if deemed feasible.

4.3 | Beyond sCT for radiotherapy

We found other possible applications of DL-based image generation during the database search, which are beyond the categories mentioned so far or the RT application. For example, Kawahara et al.[222] proposed to generate synthetic dual-energy CT from CT to assess the body material composition using 2D-paired GANs. Also, commercial solutions start to be evaluated for the generation of DL-based sCT from MRI for lesion detection of suspected sacroiliitis[223] or to facilitate surgical planning of the spine.[224] An exciting application is also the generation of sCT to facilitate multimodal image registration, as proposed by Mckenzie et al.[225]

All the techniques of category I could be directly applied to MR-guided high-intensity focused ultrasound, where otherwise an additional CT would be required to properly plan the treatment.[226]

Additionally, the methods here reviewed to generate sCT can be applied to translating other image modalities. Interesting examples in the RT realm are provided by Jiang et al.[227] who investigated using MRI-to-CT translation to increase the segmentation's robustness. Kieselmann et al.[228] generated synthetic MRI from CT to train segmentation networks that exploit the wealth of delineation on another modality. A detailed review of other image-to-image translation applications in RT has been recently compiled by Wang et al.[49]

5 | CONCLUSION

DL-based generation of sCT has a bright future, with extensive amount of research work being done on the topic. DL methods for sCT generation have been

reviewed in the context of (I) MR to replace CT in RT treatment planning, (II) CBCT-based adaptive RT, and (III) in generating attenuation maps for PET.

A detailed review of each category was presented, providing a comprehensive comparison among DL-based methods in terms of the most popular metrics reported. We found that DL-based sCT generation is an active and growing area of research. For several anatomical sites, for example, H&N/brain and pelvis, sCT seems feasible, with DL achieving DD to CT-based planning <1% in the RT context and better performance for PET AC compared to the standard MRAC methods.

While DL-based sCT generation techniques are up and upcoming, comprehensive commissioning and QA of DL-based sCT techniques are critical prior and essential during clinical deployment, in order to ensure patient safety. The key to further diffusion of DL-based sCT techniques is evaluating their generalization capability in a multicenter setting.

ACKNOWLEDGMENTS

Matteo Maspero is grateful to prof.dr.ir. Cornelis (Nico) AT van den Berg, head of the Computational Imaging Group for MR diagnostics and therapy, Center for Image Sciences, UMC Utrecht, the Netherlands, for the general support provided during this manuscript's compilation.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Maria Francesca Spadea
https://orcid.org/0000-0002-5339-9583
Matteo Maspero
https://orcid.org/0000-0003-0347-3375
Paolo Zaffino
https://orcid.org/0000-0002-0219-0157
Joao Seco https://orcid.org/0000-0002-9458-2202

REFERENCES

1. Husband J, Reznek RH, Husband JE. *Imaging in Oncology*. CRC Press; 2016. https://doi.org/10.1201/9780203732656
2. Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. *Br J Cancer*. 2019;120:779-790.
3. Verellen D, De Ridder M, Linthout N, Tournel K, Soete G, Storme G. Innovations in image-guided radiotherapy. *Nat Rev Cancer*. 2007;7:949-960.
4. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. *Nat Rev Clin Oncol*. 2012;9:688.
5. Seco J, Spadea MF. Imaging in particle therapy: state of the art and future perspective. *Acta Oncol*. 2015;54:1254-1258.
6. IAEA. *Radiotherapy in Cancer Care: Facing the Global Challenge*. International Atomic Energy Agency; 2017. http://www-pub.iaea.org/MTCD/Publications/PDF/P1638_web.pdf
7. Seco J, Evans PM. Assessing the effect of electron density in photon dose calculations. *Med Phys*. 2006;33:540-552.
8. Unterrainer M, Eze C, Ilhan H, et al. Recent advances of PET imaging in clinical radiation oncology. *Radiat Oncol*. 2020;15:1:15.
9. Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. *Semin Radiat Oncol*. 2014;24:151-159.
10. Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol*. 2015;60:R323.
11. Devic S. MRI simulation for radiotherapy treatment planning. *Med Phys*. 2012;39:6701.
12. Nyholm T, Nyberg M, Karlsson MG, Karlsson M. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiat Oncol*. 2009;4:1-9.
13. Ulin K, Urie MM, Cherlow JM. Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol Biol Phys*. 2010;77:1584-1589.
14. Fraass BA, McShan DL, Diaz RF, et al. Integration of magnetic resonance imaging into radiation therapy treatment planning: I. Technical considerations. *Int J Radiat Oncol Biol Phys*. 1987;13:1897-908.
15. Lee YK, Bollet M, Charles-Edwards G, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiother Oncol*. 2003;66:203-216.
16. Nyholm T, Jonsson J. Counterpoint: opportunities and challenges of a magnetic resonance imaging-only radiotherapy work flow. *Semin Radiat Oncol*. 2014;24:175-80.
17. Kapanen M, Collan J, Beule A, Seppälä T, Saarilahti K, Tenhunen M. Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. *Magn Reson Med*. 2013;70:127-35.
18. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. *Phys Med Biol*. 2018;63:05TR01.
19. Karlsson M, Karlsson MG, Nyholm T, Amies C, Zackrisson B. Dedicated magnetic resonance imaging in the radiotherapy clinic. *Int. J. Radiat. Oncol. Biol. Phys.* 2009;74:644-651.
20. Lagendijk JJW, Raaymakers BW, Berg CAT, Moerland MA, Philippens ME, Van Vulpen M. MR guidance in radiotherapy. *Phys Med Biol*. 2014;59:R349.
21. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol*. 2010;5:62.
22. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12:28.
23. Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100:199-217.
24. Wafa B, Moussaoui A. A review on methods to estimate a CT from MRI data in the context of MRI-alone RT. *Med Tech J*. 2018;2:150-178.
25. Kerkmeijer LGW, Maspero M, Meijer GJ, van der Voort van Zyp JRN, de Boer HCJ, van den Berg CAT. Magnetic Resonance Imaging only Workflow for Radiotherapy Simulation and Planning in Prostate Cancer. *Clin Oncol*. 2018;30:(11):692-701. https://doi.org/10.1016/j.clon.2018.08.009
26. Bird D, Henry AM, Sebag-Montefiore D, Buckley DL, Al-Qaisieh B, Speight R. A systematic review of the clinical implementation of pelvic magnetic resonance imaging-only planning for external beam radiation therapy. *Int J Radiat Oncol Biol Phys*. 2019;105:479-492.
27. Thorwarth D, Low DA. Technical challenges of real-time adaptive MR-guided radiotherapy. *Front Oncol*. 2021;11.
28. Hoffmann A, Oborn B, Moteabbed M, et al. MR-guided proton therapy: a review and a preview. *Radiat Oncol*. 2020;15.
29. Taasti VT, Klages P, Parodi K, Muren LP. Developments in deep learning based corrections of cone beam computed tomography to enable dose calculations for adaptive radiotherapy. *Physics and Imaging in Radiat Oncol*. 2020;15:77-79.

30. Zhu L, Wang J, Xing L. Noise suppression in scatter correction for cone-beam CT. *Med Phys*. 2009;36:(3):741-752. https://doi.org/10.1118/1.3063001
31. Zhu L, Xie Y, Wang J, Xing L. Scatter correction for cone-beam CT in radiation therapy. *Med Phys*. 2009;36:(6Part1):2258-2268. https://doi.org/10.1118/1.3130047
32. Mehranian A, Arabi H, Zaidi H. Vision 20/20: magnetic resonance imaging-guided attenuation correction in PET/MRI: challenges, solutions, and opportunities. *Med Phys*. 2016;43:1130-1155.
33. Mecheter I, Alic L, Abbod M, Amira A, Ji J. MR image-based attenuation correction of brain PET imaging: review of literature on machine learning approaches for segmentation. *Journal of Digital Imaging*. 2020;1-18.
34. Catana C. Attenuation correction for human PET/MRI studies. *Phys Med Biol*. 2020;65:TR02.
35. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
36. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*. Vol. 2 in Adaptive Computation and Machine Learning. MIT Press; 2016.
37. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. *Comp Biol Med*. 2018;98:126-146.
38. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2018;46:e1-e36.
39. Boon I, Yong TA, Boon C. Assessing the role of artificial intelligence (AI) in clinical oncology: utility of machine learning in radiotherapy target volume delineation. *Medicines*. 2018;5:131.
40. Wang C, Zhu X, Hong JC, Zheng D. Artificial intelligence in radiotherapy treatment planning: present and future. *Tech Cancer Res Treat*. 2019;18:153303381987392.
41. Boldrini L, Bibault JE, Masciocchi C, Shen Y, Bittner M-I. Deep learning: a review for the radiation oncologist. *Front Oncol*. 2019;9.
42. Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. *Brit J Radiol*. 2019;92:20190001.
43. Kiser KJ, Fuller CD, Reed VK. Artificial intelligence in radiation oncology treatment planning: a brief overview. *J Med Art Intell*. 2019;2:9.
44. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neur Inf Proc Syst*. 2012;25:1097-1105.
45. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
46. Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT image from MRI data using 3D fully convolutional networks. 2016;2016:170-178.
47. Lee JS. A Review of Deep-Learning-Based Approaches for Attenuation Correction in Positron Emission Tomography. *IEEE Trans Radiat Plasma Med Sci*. 2021;5:(2):160-184. https://doi.org/10.1109/trpms.2020.3009269
48. Yu B, Wang Y, Wang L, Shen D, Zhou L. *Medical Image Synthesis via Deep Learning*. Springer International Publishing; 2020:23-44.
49. Wang T, Lei Y, Fu Y, et al. A review on medical imaging synthesis using deep learning and its clinical applications. *J Appl Clin Med Phys*. 2021;22:11-36.
50. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
51. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015:234-241.
52. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inform Process Syst*. 2014;27:2672-2680.
53. Isola P, Zhu J-Y, Zhou T & Efros AA. Image-to-image translation with conditional adversarial networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; Honolulu, HI:1125-1134. https://doi.org/10.1109/CVPR.2017.632
54. Wu X, Xu K, Hall P. A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Sci Technol*. 2017;22:(6):660-674. https://doi.org/10.23919/tst.2017.8195348
55. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative Adversarial Networks: An Overview. *IEEE Signal Process Mag*. 2018;35:(1):53-65. https://doi.org/10.1109/msp.2017.2765202
56. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. 2019;58:101552.
57. Zhu J-Y, Park T, Isola P & Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision; 2017:2223-2232. https://doi.org/10.1109/ICCV.2017.244
58. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Imag Proc*. 2004;13:600-612.
59. Low DA. Gamma dose distribution evaluation tool. *J Phys*. 2010;250:012071.
60. Clasie BM, Sharp GC, Seco J, Flanz JB, Kooy HM. Numerical solutions of the γ-index in two and three dimensions. *Phys Med Biol*. 2012;57:6981.
61. Hussein M, Clark CH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy—towards good practice. *Phys Med*. 2017;36:1-11.
62. Drzymala RE, Mohan R., Brewster L., et al. Dose-volume histograms. *Int J Radiat Oncol Biol Phys*. 1991;21:71-78.
63. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol*. 2012;57:R99.
64. Pileggi G, Speier C, Sharp GC, et al. Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR. *Acta Oncol*. 2018;57:1521-1531.
65. Andres EA, Fidon L, Vakalopoulou M, et al. Dosimetry-driven quality measure of brain pseudo computed tomography generated from deep learning for MRI-only radiotherapy treatment planning. *Int J Radiat Oncol Biol Phys*. 2020;108:813-823.
66. Eckl M, Hoppen L, Sarria GR, et al. Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy. *Phys Med*. 2020;80:308-316.
67. Peng Y, Chen S, Qin A, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. *Radiother Oncol*. 2020;150:217-224.
68. Qian P, Xu K, Wang T, Zheng Q, Yang H, Baydoun A, Zhu J, Traughber B, Muzic RF. Estimating CT from MR Abdominal Images Using Novel Generative Adversarial Networks. *J Grid Comput*. 2020;18:(2):211-226. https://doi.org/10.1007/s10723-020-09513-3
69. Xu K, Cao J, Xia K, et al. Multichannel residual conditional GAN-leveraged abdominal pseudo-CT generation via Dixon MR images. *IEEE Access*. 2019;7:163823-163830.
70. Ladefoged CN, Marner L, Hindsholm A, Law I, Højgaard L, Andersen FL. Deep learning based attenuation correction of PET/MRI in pediatric brain tumor patients: evaluation in a clinical setting. *Front Neurosci*. 2019;12:1005.
71. Maspero M, Bentvelzen LG, Savenije MHF, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. *Radiother Oncol*. 2020;153:197-204. https://doi.org/10.1016/j.radonc.2020.09.029
72. Florkow MC, Guerreiro F, Zijlstra F, et al. Deep learning-enabled MRI-only photon and proton therapy treatment planning for paediatric abdominal tumours. *Radiother Oncol*. 2020;153:220-227.
73. Jeon W, An HJ, Kim J-i, et al. Preliminary application of synthetic computed tomography image generation from magnetic

resonance image using deep-learning in breast cancer patients. *J Radiat Prot Res*. 2019;44:149-155.
74. Bradshaw TJ, Zhao G, Jang H, Liu F, McMillan AB. Feasibility of deep learning–based PET/MR attenuation correction in the pelvis using only diagnostic MR images. *Tomography*. 2018;4:138.
75. Fu J, Singhrao K, Cao M, et al. Generation of abdominal synthetic CTs from 035 T MR images using generative adversarial networks for MR-only liver radiotherapy. *Biom Phys Eng Express*. 2020;6:015033.
76. Li Y, Li W, Xiong J, Xia J, Xie Y. Comparison of Supervised and Unsupervised Deep Learning Methods for Medical Image Synthesis between Computed Tomography and Magnetic Resonance Images. *Biomed Res Int*. 2020;2020:1-9. https://doi.org/10.1155/2020/5193707
77. Xu L, Zeng X, Zhang H, Li W, Lei J, Huang Z. BPGAN: bidirectional CT-to-MRI prediction using multi-generative multi-adversarial nets with spectral normalization and localization. *Neural Netw*. 2020;128:82-98.
78. Fu J, Yang Y, Singhrao K, et al. Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging. *Med Phys*. 2019;46:3788-3798.
79. Neppl S, Landry G, Kurz C, et al. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudo-CTs from T1-weighted MR head scans. *Acta Oncol*. 2019;58:1429-1434.
80. Fetty L, Bylund M, Kuess P, et al. Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Zeits Med Phys*. 2020;30.
81. Xiang L, Wang Q, Nie D, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med Imag Anal*. 2018;47:31-44.
82. Cusumano D, Lenkowicz J, Votta C, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol*. 2020;153:205-212.
83. Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. 2019;46:3998-4009.
84. Maspero M, Houweling AC, Savenije MHF, et al. A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. *Phys Imag Radiat Oncol*. 2020;14:24-31.
85. Zhang Y, Yue N, Su MY, Liu B, Ding Y, Zhou Y, Wang H, Kuang Y, Nie K. Improving CBCT quality to CT level using deep learning with generative adversarial network. *Med Phys*. 2021;48:(6):2816-2826. https://doi.org/10.1002/mp.14624
86. Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, Kerkmeijer LGW, van den Berg CAT. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63:(18):185001. https://doi.org/10.1088/1361-6560/aada6d
87. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408-1419.
88. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. 2018;45:3627-3636.
89. Jin C-B, Kim H, Liu M, et al. Deep CT to MR synthesis using paired and unpaired data. *Sensors*. 2019;19:2361.
90. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46:3565-3581.
91. Yang H, Sun J, Carass A, et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Trans Med Imag*. 2020;39:4249-4261.
92. Massa H, Johnson J, McMillan A. Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs. *Phys Med Biol*. 2020;65:NT03.
93. Wang Y, Liu C, Zhang X, Deng W. Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN). *Front Oncol*. 2019;9.
94. Tie X, Lam S-K, Zhang Y, Lee K-H, Au K-H, Cai J. Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. *Med Phys*. 2020;47:1750-1762.
95. Kearney V, Ziemer BP, Perry A, Wang Y, Chan JW, Ma L, Morin O, Yom SS, Solberg TD. Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks. *Radiol Artif Intell*. 2020;2:(2):e190027. https://doi.org/10.1148/ryai.2020190027
96. Largent A, Marage L, Gicquiau I, et al. Head-and-neck MRI-only radiotherapy treatment planning: from acquisition in treatment position to pseudo-CT generation. *Cancer Radiother*. 2020;24:288-297.
97. Su P, Guo S, Roys S, et al. Transcranial MR imaging–guided focused ultrasound interventions using deep learning synthesized CT. *Am J Neurorad*. 2020;41:1841-1848.
98. Florkow MC, Zijlstra F, Willemsen K, Maspero M, Berg CAT, Kerkmeijer LGW, Castelein RM, Weinans H, Viergever MA, Stralen M, Seevinck PR. Deep learning–based MR-to-CT synthesis: The influence of varying gradient echo–based MR images as input channels. *Magn Reson Med*. 2020;83:(4):1429-1441. https://doi.org/10.1002/mrm.28008
99. Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A new deep convolutional neural network design with efficient learning capability: application to CT image synthesis from MRI. *Med Phys*. 2020;47:5158-5171.
100. Liu Y, Lei Y, Wang Y, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. *Phys Med Biol*. 2019;64:145015.
101. Liu L, Johansson A, Cao Y, Dow J, Lawrence TS, Balter JM. Abdominal synthetic CT generation from MR Dixon images using a U-net trained with "semi-synthetic" CT data. *Phys Med Biol*. 2020;65:125001.
102. Dinkla AM, Wolterink JM, Maspero M, et al. MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network. *Int J Radiat Oncol Biol Phys*. 2018;102:801-812.
103. Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. *J Appl Clin Med Phys*. 2019;20:105-114.
104. Kazemifar S, McGuire S, Timmerman R, et al. MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother Oncol*. 2019;136:56-63.
105. Shafai-Erfani G, Lei Y, Liu Y, et al. MRI-based proton treatment planning for base of skull tumors. *Int J Part Ther*. 2019;6:12-25.
106. Gupta D, Kim M, Vineberg KA, Balter JM. Generation of synthetic CT images from MRI for treatment planning and patient positioning using a 3-channel U-net trained on sagittal images. *Front Oncol*. 2019;9:964.
107. Spadea MF, Pileggi G, Zaffino P, et al. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images–application in brain proton therapy. *Int J Radiat Oncol Biol Phys*. 2019;105:495-503.
108. Koike Y, Akino Y, Sumida I, et al. Feasibility of synthetic computed tomography generated with an adversarial network for multi-sequence magnetic resonance-based brain radiotherapy. *J Radiat Res*. 2020;61:92-103.

109. Kazemifar S, Barragán MAM, Souris K, et al. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. *J Appl Clin Med Phys*. 2020;21:76-86.
110. Chen S, Qin A, Zhou D, Yan D. U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Med Phys*. 2018;45:5659-5665.
111. Arabi H, Dowling JA, Burgos N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. *Med Phys*. 2018;45:5218-5233.
112. Liu Y, Lei Y, Wang Y, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. *Phys Med Biol*. 2019;64:205022.
113. Largent A, Barateau A, Nunes J-C, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int J Radiat Oncol Biol Phys*. 2019;105:1137-1150.
114. Boni KNB, Klein J, Vanquin L, et al. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. *Phys Med Biol*. 2020;65:075002.
115. Fetty L, Löfstedt T, Heilemann G, et al. Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion. *Phys Med Biol*. 2020;65:5004.
116. Bird D, Nix MG, McCallum H, et al. Multicentre, deep learning, synthetic-CT generation for ano-rectal MR-only radiotherapy treatment planning. *Radiother Oncol*. 2021;156:23-28.
117. Dinkla AM, Florkow MC, Maspero M, Savenije MHF, Zijlstra F, Doornaert PAH, Stralen M, Philippens MEP, Berg CAT, Seevinck PR. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. *Med Phys*. 2019;46:(9):4095-4104. https://doi.org/10.1002/mp.13663
118. Klages P, Benslimane I, Riyahi S, et al. Patch-based generative adversarial neural network models for head and neck MR-only planning. *Med Phys*. 2020;47:626-642.
119. Qi M, Li Y, Wu A, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med Phys*. 2020;47:1880-1894.
120. Thummerer A, Jong BA, Zaffino P, et al. Comparison of the suitability of CBCT-and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. *Phys Med Biol*. 2020;65:235036.
121. Olberg S, Zhang H, Kennedy WR, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. *Med Phys*. 2019;46:4135-4147.
122. Florkow MC, Zijlstra F, Kerkmeijer LGW, Maspero M, van den Berg CAT, van Stralen M, Seevinck PR, et al. The impact of MRI-CT registration errors on deep learning-based synthetic CT generation. In: *Medical Imaging 2019: Image Processing*, Vol. 10949. International Society for Optics and Photonics; 2019:1094938. https://doi.org/10.1117/12.2512747
123. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. In: Angelini ED, Landman BA, eds. *Medical Imaging 2019: Image Processing*. SPIE; 2019.
124. Kida S, Nakamoto T, Nakano M, et al. Cone beam computed tomography image quality improvement using a deep convolutional neural network. *Cureus*. 2018;10:e2548.
125. Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. *Med Phys*. 2020;47:1115-1125.
126. Kida S, Kaji S, Nawa K, et al. Visual enhancement of cone-beam CT by use of CycleGAN. *Med Phys*. 2020;47:998-1010.
127. Yuan N, Dyer B, Rao S, 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.
128. Liang X, Chen L, Nguyen D, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. *Phys Med Biol*. 2019;64:125002.
129. Li Y, Zhu J, Liu Z, et al. A preliminary study of using a deep convolution neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma. *Phys Med Biol*. 2019;64:145010.
130. Barateau A, De Crevoisier R, Largent A, et al. Comparison of CBCT-based dose calculation methods in head and neck cancer radiotherapy: from Hounsfield unit to density calibration curve to deep learning. *Med Phys*. 2020;47:4683-4693.
131. Liu Y, Lei Y, Wang T, et al. CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy. *Med Phys*. 2020;47:2472-2483.
132. Landry G, Hansen D, Kamp F, et al. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Phys Med Biol*. 2019;64. https://doi.org/10.1088/1361-6560/aaf496.
133. Kurz C, Maspero M, Savenije MHF, et al. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. *Phys Med Biol*. 2019;64:225004.
134. Thummerer A, Zaffino P, Meijers A, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. *Phys Med Biol*. 2020;65:095002.
135. Radford A, Metz L & Chintala S Unsupervised representation learning with deep convolutional generative adversarial networks. Preprint 2015. arXiv:1511.06434.
136. Karras T, Aila T, Laine S & Lehtinen J Progressive growing of gans for improved quality, stability, and variation. Preprint 2017. arXiv:1710.10196.
137. Oktay O, Schlemper J & Folgoc LL et al. Attention U-net: learning where to look for the pancreas. Preprint 2018. arXiv:1804.03999.
138. Leynes AP, Yang J, Wiesinger F, et al. Direct pseudo-CT generation for pelvis PET/MRI attenuation correction using deep convolutional neural networks with multi-parametric MRI: zero echo-time and Dixon deep pseudo-CT (ZeDD-CT). *J Nucl Med*. 2017;59:852-858.
139. Baydoun A, Xu K, Yang H, Zhou F, Heo JU, Jones RS, Avril N, Traughber MS, Traughber BJ, Qian P, Muzic RF. Dixon-based thorax synthetic CT generation using Generative Adversarial Network. *Artif Intell Med*. 2020;3-4:100010. https://doi.org/10.1016/j.ibmed.2020.100010
140. Gong K, Yang J, Kim K, El FG, Seo Y, Li Q. Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images. *Phys Med Biol*. 2018;63:125011.
141. Jang H, Liu F, Zhao G, Bradshaw T, McMillan AB. Deep learning based MRAC using rapid ultrashort echo time imaging. *Med Phys*. 2018;45:3697-3704.
142. Torrado-Carvajal A, Vera-Olmos J, Izquierdo-Garcia D, et al. Dixon-VIBE deep learning (DIVIDE) pseudo-CT synthesis for pelvis PET/MR attenuation correction. *J Nucl Med*. 2019;60:429-435.
143. Blanc-Durand P, Khalife M, Sgard B, et al. Attenuation correction using 3D deep convolutional neural network for brain 18F-FDG PET/MR: comparison with Atlas, ZTE and CT based attenuation correction. *PloS One*. 2019;14:e0223141.
144. Gong K, Han PK, Johnson K, El Fakhri G, Ma C, Li Q. Attenuation correction using deep learning and integrated UTE/multi-echo Dixon sequence: evaluation in amyloid and tau PET imaging. *Eur J Nucl Med Mol Imaging*. 2021;48:1-11.
145. Pozaruk A, Pawar K, Li S, et al. Augmented deep learning model for improved quantitative accuracy of MR-based PET

attenuation correction in PSMA PET-MRI prostate imaging. *Eur J Nucl Med Mol Imaging*. 2021;48:9-20.
146. Gong K, Yang J, Larson PEZ, et al. MR-based attenuation correction for brain PET using 3D cycle-consistent adversarial network. *IEEE Trans Radiat Plasma Med Sci*. 2021;5:185-192.
147. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging–based attenuation correction for PET/MR imaging. *Radiology*. 2018;286:676-684.
148. Arabi H, Zeng G, Zheng G, Zaidi H. Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI. *Eur J Nucl Med Mol Imaging*. 2019;46:2746-2759.
149. Spuhler KD, Gardus J, Gao Y, DeLorenzo C, Parsey R, Huang C. Synthesis of patient-specific transmission data for PET attenuation correction for PET/MRI neuroimaging using a convolutional neural network. *J Nucl Med*. 2019;60:555-560.
150. Liu F, Jang H, Kijowski R, Zhao G, Bradshaw T, McMillan AB. A deep learning approach for 18F-FDG PET attenuation correction. *EJNMMI Phys*. 2018;5:(1). https://doi.org/10.1186/s40658-018-0225-8
151. Dong X, Wang T, Lei Y, et al. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. *Phys Med Biol*. 2019;64:215016.
152. Armanious K, Hepp T, Küstner T, et al. Independent attenuation correction of whole body [^{18}F] FDG-PET using a deep learning approach with generative adversarial networks. *EJNMMI Res*. 2020;10:1-9.
153. Simonyan K & Zisserman A Very deep convolutional networks for large-scale image recognition. Preprint 2014. arXiv:1409.1556.
154. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770-778.
155. Van Dyk JJ, ed. *The Modern Technology of Radiation Oncology*. Medical Physics Publisher; 2020.
156. Stemkens B, Paulson ES, Tijssen RHN. Nuts and bolts of 4D-MRI for radiotherapy. *Phys Med Biol*. 2018;63:21TR01.
157. Paganelli C, Whelan B, Peroni M, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol*. 2018;63:22TR03.
158. Freedman JN, Bainbridge HE, Nill S, et al. Synthetic 4D-CT of the thorax for treatment plan adaptation on MR-guided radiotherapy systems. *Phys Med Biol*. 2019;64:115005.
159. Goodman TR, Mustafa A, Rowe E. Pediatric CT radiation exposure: where we were, and where we are now. *Pediatric Radiol*. 2019;49:469-478.
160. Walker A, Metcalfe P, Liney G, et al. MRI geometric distortion: impact on tangential whole-breast IMRT. *J Appl Clin Med Phys*. 2016;17:7-19.
161. Gustafsson C, Nordström F, Persson E, Brynolfsson J, Olsson LE. Assessment of dosimetric impact of system specific geometric distortion in an MRI only based radiotherapy workflow for prostate. *Phys Med Biol*. 2017;62:2976-2989.
162. Maspero M, Tyyger MD, Tijssen RHN, Seevinck PR, Intven MPW, van den Berg CAT. Feasibility of magnetic resonance imaging-only rectum radiotherapy with a commercial synthetic computed tomography generation solution. *Phys Imaging Radiat Oncol*. 2018;7:58-64. https://doi.org/10.1016/j.phro.2018.09.002
163. Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, et al. MRI/linac integration. *Radiother Oncol*. 2008;86:25-29.
164. Fallone BG. The rotating biplanar linac–magnetic resonance imaging system. 2014;24:200-202.
165. Mutic S, Dempsey JF. The ViewRay system: magnetic resonance–guided and controlled radiotherapy. 2014;24:196-199.
166. Keall PJ, Barton M, Crozier S, et al. The Australian magnetic resonance imaging–linac program. *Semin Radiat Oncol*. 2014;24:203-206.
167. Jaffray DA, Carlone MC, Milosevic MF, et al. A facility for magnetic resonance–guided radiation therapy. *Semin Radiat Oncol*. 2014;24:193-195.
168. Winkel D, Bol GH, Kroon PS, et al. Adaptive radiotherapy: the Elekta unity MR-linac concept. *Clin Transl Radiat Oncol*. 2019;18:54-59.
169. Hall WA, Paulson ES, Heide UA, et al. The transformation of radiation oncology using real-time magnetic resonance guidance: a review. *Eur J Cancer*. 2019;122:42-52.
170. Groot Koerkamp ML, de Hond YJM, Maspero M, Kontaxis C, Mandija S, Vasmel JE, Charaghvandi RK, Philippens MEP, van Asselen B, van den Bongard HJGD, Hackett SS, Houweling AC. Synthetic CT for single-fraction neoadjuvant partial breast irradiation on an MRI-linac. *Phys Med Biol*. 2021;66:(8):085010. https://doi.org/10.1088/1361-6560/abf1ba
171. Boda-Heggemann J, Lohr F, Wenz F, Flentje M, Guckenberger M. kV cone-beam CT-based IGRT. *Strahlen Onkol*. 2011;187:284-291.
172. Elstrøm UV, Muren LP, Petersen JB, Grau C. Evaluation of image quality for different kV cone-beam CT acquisition and reconstruction methods in the head and neck region. *Acta Oncol*. 2011;50:908-917.
173. Peroni M, Ciardo D, Spadea MF, et al. Automatic segmentation and online virtualCT in head-and-neck adaptive radiation therapy. *Int J Radiat Oncol Biol Phys*. 2012;84:e427-e433.
174. Veiga C, Alshaikhi J, Amos R, et al. Cone-beam computed tomography and deformable registration-based "dose of the day" calculations for adaptive proton therapy. *Int J Part Ther*. 2015;2:404-414.
175. Park Y-K, Sharp GC, Phillips J, Winey BA. Proton dose calculation on scatter-corrected CBCT image: feasibility study for adaptive proton therapy. *Med Phys*. 2015;42:4449-4459.
176. Kurz C, Nijhuis R, Reiner M, et al. Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images. *Radiat Oncol*. 2016;11:1-9.
177. Arai K, Kadoya N, Kato T, et al. Feasibility of CBCT-based proton dose calculation using a histogram-matching algorithm in proton beam therapy. *Phys Med*. 2017;33:68-76.
178. Gomà C, Almeida IP, Verhaegen F. Revisiting the single-energy CT calibration for proton therapy treatment planning: a critical look at the stoichiometric method. *Phys Med Biol*. 2018;63:235011.
179. Harms J, Lei Y, Wang T, et al. Cone-beam CT-derived relative stopping power map generation via deep learning for proton radiotherapy. *Med Phys*. 2020;47:4416-4427.
180. Hansen DC, Landry G, Kamp F, et al. ScatterNet: a convolutional neural network for cone-beam CT intensity correction. *Med Phys*. 2018;45:4916-4926.
181. Wang G, Ye JC, Mueller K, Fessler JA. Image reconstruction is a new frontier of machine learning. *IEEE Trans Med Imag*. 2018;37:1289-1296.
182. Wang G, Zhang Y, Ye X, Mou X. *Machine Learning for Tomographic Imaging*. IOP Publishing; 2019.
183. Li Y, Li K, Zhang C, Montoya J, Chen G-H. Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. *IEEE Trans Med Imag*. 2019;38:2469-2481.
184. Maier AK, Syben C, Stimpel B, et al. Learning with known operators reduces maximum error bounds. *Nat Machine Intell*. 2019;1:373-380.
185. Lønning K, Putzky P, Sonke J-J, Reneman L, Caan MWA, Welling M. Recurrent inference machines for reconstructing heterogeneous MRI data. *Med Image Anal*. 2019;53:64-78.

186. Izquierdo-Garcia D, Sawiak SJ, Knesaurek K, et al. Comparison of MR-based attenuation correction and CT-based attenuation correction of whole-body PET/MR imaging. *Eur J Nucl Med Mol Imag*. 2014;41:1574-1584.
187. Shiri I, Arabi H, Geramifar P, et al. Deep-JASC: joint attenuation and scatter correction in whole-body ^{18}F-FDG PET using a deep residual network. *Eur J Nucl Med Mol Imag*. 2020;47:2533-2548.
188. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6:(1). https://doi.org/10.1186/s40537-019-0197-0
189. Li Z, Kamnitsas K, Glocker B. Overfitting of neural nets under class imbalance: analysis and improvements for segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2019:402-410.
190. Hang Z, Orazio G, Iuri F, Jan K. Loss Functions for neural networks for image processing. *CoRR*. 2015;abs/1511.08861.
191. Wolterink JM, Dinkla AM, Savenije Mark H, Seevinck PR, Berg Cornelis AT, Išgum Ivana. *Deep MR to CT Synthesis Using Unpaired Data*. Springer; 2017:14-23.
192. Rehman A, Khan FG. A deep learning based review on abdominal images. *Multimed Tools Appl*. 2020;https://doi.org/10.1007/s11042-020-09592-0
193. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. *Sensors*. 2020;20:5097.
194. Kamnitsas K, Ferrante E, Parisot S, et al. DeepMedic for brain tumor segmentation. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer; 2016:138-149.
195. Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal*. 2019;53:197-207.
196. Keeling P, Clark J, Finucane S. Challenges in the clinical implementation of precision medicine companion diagnostics. *Expert Rev Mol Diagn*. 2020;20:593-599.
197. Bertholet J, Anastasi G, Noble D, et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part II: offline and online plan adaption for interfractional changes. *Radiother Oncol*. 2020;153:88-96.
198. Palmér E, Karlsson A, Nordström F, et al. Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy. *Phys Imag Radiat Oncol*. 2021;17:36-42.
199. Council of European Union. *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC*; 2017. http://data.europa.eu/eli/reg/2017/745/oj.
200. Fiorino C, Guckenberger M, Schwarz M, Heide UA, Heijmen B. Technology-driven research for radiotherapy innovation. *Mol Oncol*. 2020;14:1500-1513.
201. Beckers R, Kwade Z, Zanca F. The EU medical device regulation: implications for artificial intelligence-based medical device software in medical physics. *Phys Med*. 2021;83:1-8.
202. Liesbeth V, Michaël C, Anna M D, et al. Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance. *Radiother Oncol*. 2020;153:55-66.
203. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Brit Med J*. 2020;370:m3164.
204. Dowling JA, Korhonen J. MR-only methodology. In: *MRI for Radiotherapy*. Springer International Publishing; 2019:131-151. https://doi.org/10.1007/978-3-030-14442-5_9
205. Teuho J, Johansson J, Linden J, et al. Effect of attenuation correction on regional quantification between PET/MR and PET/CT: a multicenter study using a 3-dimensional brain phantom. *J Nucl Med*. 2016;57:818-824.
206. Wyatt JJ, Dowling JA, Kelly CG, et al. Investigating the generalisation of an atlas-based synthetic-CT algorithm to another centre and MR scanner for prostate MR-only radiotherapy. *Phys Med Biol*. 2017;62:N548-N560.
207. Persson E, Gustafsson C, Nordström F, et al. MR-OPERA: a multicenter/multivendor validation of magnetic resonance imaging–only prostate treatment planning using synthetic computed tomography images. *Int J Radiat Oncol Biol Phys*. 2017;99:692-700.
208. Greer P, Martin J, Sidhom M, et al. A multi-center prospective study for implementation of an MRI-only prostate treatment planning workflow. *Front Oncol*. 2019;9:826.
209. Loi G, Fusella M, Vecchi C, et al. Computed tomography to cone beam computed tomography deformable image registration for contour propagation using head and neck, patient-based computational phantoms: a multicenter study. *Pract Radiat Oncol*. 2020;10:125-132.
210. Pan SJ Jialin, Yang Q. A Survey on Transfer Learning. *IEEE Trans Knowl Data Eng*. 2010;22:(10):1345-1359. https://doi.org/10.1109/tkde.2009.191
211. Cheplygina V, Bruijne M, Pluim JPW. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal*. 2019;54:280-296.
212. Li W, Kazemifar S, Bai T, et al. Synthesizing CT images from MR images with deep learning: model generalization for different datasets through transfer learning. 2021;7:025020.
213. Mutic S, Palta JR, Butker EK, et al. Quality assurance for computed-tomography simulators and the computed-tomography-simulation process: report of the AAPM Radiation Therapy Committee Task Group No. 66. *Med Phys*. 2003;30:2762-2792.
214. Gallas RR, Hünemohr N, Runz A, Niebuhr NI, Jäkel O, Greilich S. An anthropomorphic multimodality (CT/MRI) head phantom prototype for end-to-end tests in ion radiotherapy. *Zeitsch Mediz Phys*. 2015;25:391-399.
215. Niebuhr NI, Johnen W, Echner G, et al. The ADAM-pelvis phantom–an anthropomorphic, deformable and multimodal phantom for MRgRT. *Phys Med Biol*. 2019;64:04NT05.
216. Singhrao K, Fu J, Wu HH, et al. A novel anthropomorphic multimodality phantom for MRI-based radiotherapy quality assurance testing. *Med physics*. 2020;47:1443-1451.
217. Colvill E, Krieger M, Bosshard P, et al. Anthropomorphic phantom for deformable lung and liver CT and MR imaging for radiotherapy. *Phys Med Biol*. 2020;65:07NT02.
218. Chen X, Men K, Chen B, et al. CNN-based quality assurance for automatic segmentation of breast cancer in radiotherapy. *Front Oncol*. 2020;10.
219. Bragman FJ, Tanno R & Eaton-Rosen Z et al. Uncertainty in multitask learning: joint representations for probabilistic MR-only radiotherapy planning. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2018:3-11. https://doi.org/10.1007/978-3-030-00937-3_1
220. Hemsley M, Chugh B, Ruschin M, et al. Deep generative model for synthetic-CT generation with uncertainty predictions. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2020:834-844.
221. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, Fieguth P, Cao X, Khosravi A, Acharya UR, Makarenkov V & Nahavandi S. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf Fusion*. 2021;76:243-297. https://doi.org/10.1016/j.inffus.2021.05.008

222. Kawahara D, Saito A, Ozawa S, Nagata Y. Image synthesis with deep convolutional generative adversarial networks for material decomposition in dual-energy CT from a kilovoltage CT. *Comp Biol Med*. 2020;128:104111.
223. Jans LBO, Chen M, Elewaut D, et al. MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: comparison with MRI. *Radiol*. 2020;298:343-349.
224. Staartjes VE, Seevinck PR, Vandertop WP, Stralen M, Schröder ML. Magnetic resonance imaging–based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept. *Neurosurg Focus*. 2021;50:E13.
225. McKenzie EM, Santhanam A, Ruan D, O'Connor D, Cao M, Sheng K. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. *Med Phys*. 2020;47:1094-1104.
226. Siedek F, Yeo SY, Heijman E, Grinstein O, Bratke G, Heneweer C, Puesken M, Persigehl T, Maintz D, Grüll H. Magnetic Resonance-Guided High-Intensity Focused Ultrasound (MR-HIFU): Technical Background and Overview of Current Clinical Applications (Part 1). *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. 2019;191:(06):522-530. https://doi.org/10.1055/a-0817-5645
227. Jiang J, Hu Y-C, Tyagi N, et al. Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets. *Med Phys*. 2019;46:4392-4404.
228. Kieselmann JP, Fuller CD, Gurney-Champion OJ, Oelfke U. Cross-modality deep learning: Contouring of MRI data from annotated CT data only. *Med Phys*. 2020;48:1673-1684.

How to cite this article: Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: A review. *Med. Phys.* 2021;48:6537–6566. https://doi.org/10.1002/mp.15150

APPENDIX

The query used in selected databases—PubMed, Scopus and Web of Science—in the fields (Title/Abstract/Keywords) was the following (Figure A1):

((“radiotherapy”)OR (“radiation therapy”) OR (“proton therapy”) OR (“oncology”) OR (“imaging”) OR (“radiology”) OR (“healthcare”) OR (“CBCT”) OR (“cone-beam CT”) OR (“PET”) OR (“attenuation correction”) OR (“attenuation map”)) AND ((“synthetic CT”) OR (“syntheticCT”) OR (“synthetic-CT”) OR (“pseudo CT”) OR (“pseudoCT”) OR (“pseudo-CT”) OR (“virtual CT”) OR (“virtualCT”) OR (“virtual-CT”) OR (“derived CT”) OR (“derivedCT”) OR (“derived-CT”) OR (sCT)) AND ((“deep learning”) OR (“convolutional network”) OR (“CNN”) OR (“GAN”) OR (“GANN”) OR (artificial intelligence)).

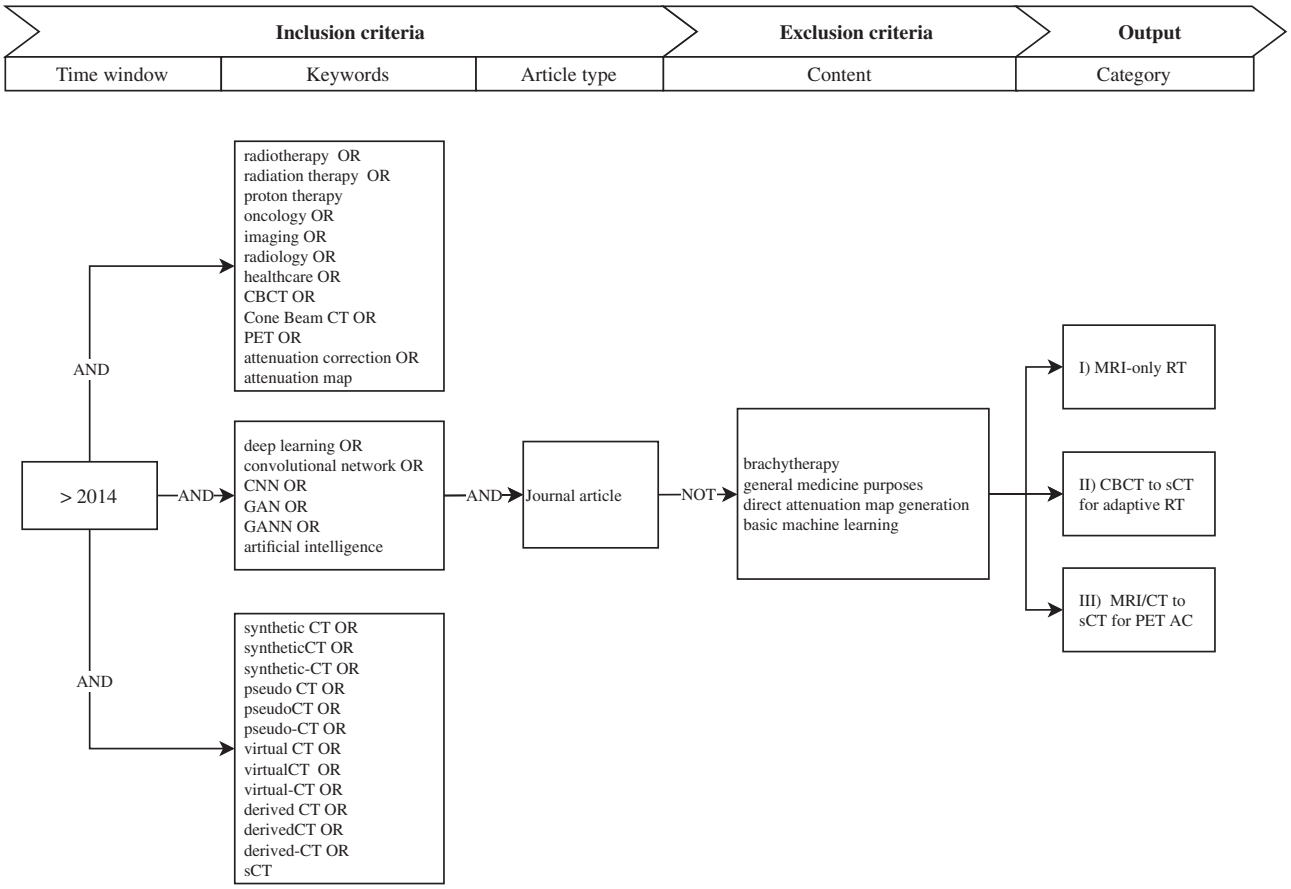

FIGURE A1 Schematic of the search inclusion/exclusion criteria adopted for this review selecting the time window, keywords, type of article, content, and the three categories defined