AI Medicine
https://www.sciltp.com/journals/aim

Review

MRI-to-CT Generation with Deep Learning: A Review and Future Directions

Ruiming Zhu [1], Xinliang Liu [2], Wei Qian [1] and Yueyang Teng [1,3,*]

[1] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China
[2] College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150006, China
[3] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Shenyang 110169, China
* Correspondence: tengyy@bmie.neu.edu.cn

How To Cite: Zhu, R.; Liu, X.; Qian, W.; et al. MRI-to-CT Generation with Deep Learning: A Review and Future Directions. *AI Medicine* **2026**, *3*(1), 4. https://doi.org/10.53941/aim.2026.100004

Received: 24 January 2026
Revised: 30 April 2026
Accepted: 25 May 2026
Published: 15 June 2026

Abstract: Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) play a vital role in the diagnosis of various pathologies and radiotherapy planning, in which CT is typically used for dose calculation. But MRI-only radiotherapy planning, with a synthetic CT (sCT) image synthesized from MRI, offers advantages in terms of time efficiency and patient safety. This method avoids the need for CT scans while retaining dose calculation information. Recently, deep learning models for image-to-image translation have shown great potential for MRI-to-CT synthesis, as they can efficiently preserve the common structure of the image data across different domains while changing the distinctive attributes of each domain. In this review, we discuss the four main deep learning methodologies for MRI-to-CT image synthesis: Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Transformer models, and Diffusion models, discussing the potential of each model and provide insights into how to improve current MRI-to-CT image synthesis approaches.

Keywords: MRI-to-CT synthesis; deep learning techniques; convolutional neural networks; generative adversarial networks; diffusion models

1. Introduction

Medical imaging plays a crucial role in diagnosing and treatment planning of diseases. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are two widely employed imaging techniques that offer valuable information about the human body, but with different imaging mechanisms. MRI mainly produces images of soft tissues, due to the interaction between hydrogen nuclei, magnetic fields, and water and fat-rich tissues such as the brain and muscles. MRI is a safe procedure for patients as the process isn't invasive and there's no radiation, although it can take a longer time (between 30 min and one hour). On the other hand, CT produces detailed images of dense tissues such as bones using X-rays. Though CT scans are quick, typically taking 5 to 10 min, they do expose patients to radiation [1]. Because MRI is commonly used for diagnosis and soft-tissue evaluation of cancer patients, while CT is required for radiation treatment planning to aid in calculation of mass and electron density, separate scans are required. This results in the need for two imaging scans, leading to additional time and radiation exposure for the patients, which may be problematic, particularly in sensitive patient groups such as children [2].

Recently, medical image synthesis has been proposed to overcome this issue, such as MRI-to-CT synthesis [3,4]. The creation of an algorithm to create a synthetic CT (sCT) from an MRI scan may eliminate the need for a separate CT scan, thus reducing radiation dose and improving clinical workflows. This technique involves generating an sCT from the MRI, which can be combined with the MRI for applications such as radiation therapy planning and dose calculation. The integration of MRI and sCT allows for precise treatment planning while eliminating the necessity for further CT scans, and brings closer the vision of MRI-only radiotherapy. The success

of MRI-to-CT synthesis addresses both the problem of radiation exposure and enhances the efficiency and cost-effectiveness of medical imaging by eliminating the need for multiple scans. Advances in deep learning and computer vision have recently improved the quality of sCT images, and thus enabled the use of MRI-to-CT synthesis in clinical settings.

Conventional MRI-to-CT synthesis approaches have involved methods such as atlas-based methods [5−10] and segmentation-based methods [11−15]. These methods usually rely on hand-crafted features or well-registered images to link MRI and CT images [16]. For instance, Johansson et al. developed a voxel-based method for CT synthesis that used three different MRI sequences with a Gaussian Mixture Regression (GMR) model [17]. This model tries to convert voxel values of MRI into Hounsfield Units (HU) of CT using clustering of voxels and their mean and variance. This approach was able to model tissue structures like soft tissues and bones well, but remains limited by its inability to capture complex non-linear mappings and its high sensitivity to image misregistration. Moreover, methods such as patch-based regression and atlas-based MRI-CT mapping depend on reliable spatial normalization, which often assumes normal anatomical structures, restricting their applicability. In addition, Dowling et al. also used an atlas-based mapping approach, which involves the use of deformable registration methods for registering the MRI scans to a reference CT image to generate a corresponding sCT [18]. However, traditional techniques are limited in capturing intricate image details, especially in abnormal areas like tumors, and rely on accurate spatial mapping, which can be both time-consuming and prone to error [19]. This can affect the reliability and generalizability of these techniques, especially when dealing with datasets that may contain a wide variety of anatomical structures.

Moreover, while these conventional methods provided the foundation for MRI-to-CT synthesis, they often failed to handle the large volumes of training data and the diversity of clinical images. Their use of hand-designed features and hand intervention limited their scalability and effectiveness. Consequently, their accuracy and generalizability to other data sets were often limited.

Data-driven deep learning-based methods have largely overcome the shortcomings of conventional methods for MRI-to-CT synthesis by providing new solutions to model the intricate relationships between different imaging modalities [20]. Deep learning approaches, particularly those using convolutional neural networks (CNNs), can automatically learn complex features from the data, enabling them to establish the relationship between MRI and CT images [21]. Through extensive training, these models can automatically learn domain-specific features, such as the differences in soft tissues, bones, and other anatomical features, making them highly flexible and able to transform MRI images into pseudo-CT images with high accuracy.

There have also been significant improvements in the accuracy and speed of MRI-to-CT synthesis through the use of deep learning. Early deep learning networks, such as U-Net, have improved upon conventional approaches by generating high-quality sCT images from MR images in real-time or near real-time. A deep convolutional network (DCNN) using U-Net architecture for 2D axial sections of T1-weighted MRI volumes, for instance, achieved better performance in terms of speed and accuracy of the synthesis process, compared to atlas-based approaches [22]. To further improve the synthesis process, more sophisticated models have also been developed using 3D convolutions, residual connections and adversarial training [23,24]. These networks, which can take 3D T1-weighted MRI volumes or Dixon MRI sequences as inputs, can create pseudo-CT volumes with improved tissue contrast and in a shorter time, sometimes in just minutes. This is in contrast to conventional methods, which involved time-consuming manual editing and registration. Furthermore, deep learning approaches have shown improved stability, with the capacity to handle different body sites and MRI scan protocols, unlike traditional methods that make rigid assumptions.

In the subsequent sections, we will group the MRI-to-CT synthesis methods into four main methods—Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Transformers, and Diffusion Models—which each employ a different approach to further boost the development of the field.

2. Datasets

Research on MRI-to-CT synthesis depends on quality data and annotations. A number of these datasets have been used frequently, and some are publicly accessible, which offer researchers datasets to develop and test their image synthesis models. These datasets contain both paired and unpaired MRI-CT images, catering to various research purposes. In the following, we list the datasets used in the documents uploaded, which are fundamental in MRI to CT synthesis problems, especially for sCT generation from MRI. A summary of publicly available medical image synthesis datasets is shown in Table 1. In addition to MRI and CT, some datasets also include other imaging modalities, such as positron emission tomography (PET), single-photon emission computed tomography (SPECT), and cone-beam computed tomography (CBCT).

Table 1. Summary of public datasets for MRI-to-CT image synthesis.

Dataset	Pairing	Modality	Region	Patients	Format	Size
Paired CT and MRI Dataset [25]	Paired	CT, T1w MRI, T2w MRI	Brain, Abdomen, Neck	18	DICOM, PNG	1167 single-modality images
Harvard medical AANLIB dataset [26]	Paired	CT, MRI, PET, SPECT	Brain	184 pairs of CT-MRI, 269 pairs of PET-MRI, 357 pairs of SPECT-MRI	DICOM, PNG	810 paired images
Gold Atlas [27]	Paired	T1w MRI, T2w MRI, CT	Pelvis	19 patients	DICOM	76 volumes
SynthRAD2025 [28]	Paired	MRI, CBCT, CT	Head-and-neck, Thorax, Abdomen	890 MRI-CT pairs, 1472 CBCT-CT pairs	MetaImage	2362 paired cases
The Cancer Imaging Archive (TCIA) [29]	Mixed [1]	CT, MRI, PET	Brain, Breast, Lung, Colon	/	DICOM, JPEG, PNG	>50 million images
Unpaired MR-CT Brain Dataset [30]	Unpaired	T2w MRI, CT	Brain	20 patients	DICOM	179 images

[1] "Mixed" indicates that the dataset is a collection-level archive containing multiple subsets, and its paired or unpaired status depends on the specific subset used.

2.1. Paired CT and MRI Dataset

The data set includes CT and MRI (T1 and T2) scans of 18 patients: brain, abdomen and neck. The scans were taken with a Siemens Verio 1.5T MRI scanner (Siemens Healthineers, Erlangen, Germany) and Siemens Somatom CT scanner (Siemens Healthineers, Erlangen, Germany). This dataset contains 389 T1-weighted MRI images, 389 T2-weighted MRI images, and 389 CT 2D axial images. The images in the dataset have been resampled to 512 × 512 resolution, and co-registered to align the images across modalities. This dataset is tailored to enhance image-to-image translation methods like CycleGAN, contributing to the improvement of accuracy and effectiveness in multimodal medical imaging algorithms.

2.2. Harvard Medical AANLIB Dataset

This dataset includes medical images of various modalities such as CT, MRI, PET, and SPECT for brain regions. It consists of 184 pairs of MRI-CT images, 269 pairs of MRI-PET images, and 357 pairs of MRI-SPECT images. Each dataset has 24 pairs selected randomly as the test set, with the remaining pairs used for training. The images are in 256 × 256 resolution and are available in DICOM and PNG formats. The PET and SPECT images are color images, while CT and MRI are single-channel grayscale. The dataset has been pre-processed for fusion tasks by converting PET and SPECT images to the YCbCr color space. The fusion evaluation of this dataset helps assess various models in generating more accurate and realistic fusion results, which are beneficial for medical diagnoses such as detecting functional and structural abnormalities in the brain.

2.3. Gold Atlas (Male Pelvis)

The Gold Atlas dataset includes multi-modal MRI and CT images from 19 male patients, collected across three different Swedish radiotherapy departments. The images consist of T1- and T2-weighted MRI sequences as well as CT scans, all acquired with the same patient positioning devices. The dataset also includes multiobserver delineations for nine organs, which were independently defined by five experts. These delineations were then used to create consensus-based segmentations, making the dataset ideal for training and validating segmentation algorithms and methods for generating sCT from MRI images.

2.4. SynthRAD2025 Dataset

The SynthRAD2025 dataset, as part of a challenge on sCT generation for radiotherapy, is a comprehensive collection of paired MRI and CT data from patients with head-and-neck, thoracic, and abdominal cancers. The dataset includes 2362 cases across three anatomical regions—head-and-neck, thorax, and abdomen. It contains 890 MRI-CT pairs for Task 1 (MRI-to-CT conversion) and 1472 CBCT-CT pairs for Task 2 (CBCT-to-CT conversion). The images were obtained from five university hospitals in Europe. Datasets were collected from a wide variety of scanners, manufacturers (Philips, Amsterdam, Netherlands; Siemens, Munich, Germany; Toshiba,

Kawasaki, Japan; GE, Boston, MA, USA) and protocols. The data is split into train (65%), validation (10%), and test (25%) subsets, enabling the comparison of algorithms for generating sCT images. The SynthRAD2025 dataset supports studies on sCT generation for applications in MRI-only and CBCT-based radiotherapy. It can be used to advance adaptive radiotherapy, enabling more tailored cancer treatments. Its broad applicability and size make it an important resource for validating sCT algorithms for photon and proton therapy.

2.5. The Cancer Imaging Archive (TCIA)

The Cancer Imaging Archive (TCIA) is a publicly available repository of cancer imaging information, which is curated by Washington University. It contains images from numerous collections across different institutions and research projects. These collections include CT scans, MRI scans, PET scans, and mammograms, covering various cancer types such as brain, breast, lung, and colon cancers. The archive also includes images from large-scale projects, such as The Cancer Genome Atlas (TCGA) and the Quantitative Imaging Network (QIN).

TCIA helps researchers by providing a portal for accessing high-quality images, used for cancer research, algorithm creation, and testing. The data contains de-identified images (to protect patient privacy) with metadata on the type of imaging, body part and protocol used. Through its de-identification process and adherence to strict standards, TCIA ensures that all patient information is removed and the data can be safely used in research.

The TCIA leverages open-source technologies, including the National Biomedical Imaging Archive (NBIA; National Cancer Institute, Bethesda, MD, USA), for secure data submission, distribution and sharing. The archive supports cancer research collaboration and offers critical data for the development, testing and validation of imaging algorithms. Furthermore, TCIA has an increasing user base, with more than 1000 users globally, and continues to increase the number of data collections.

2.6. Unpaired MRI-CT Brain Dataset

This dataset, prepared for unsupervised MRI-CT image translation, contains brain tumor images of 20 patients. There are 179 2D axial images: 90 MRI and 89 CT. The MRI images were collected with a Siemens Verio 3T scanner, T2-weighted, and CT images with a Siemens Somatom scanner. The images, acquired from April 2016 to December 2019 at Jordan University Hospital, have tumor masks outlined by radiologists. This dataset facilitates research in artificial intelligence (AI) for brain tumor detection, classification and segmentation. It can be used for deep learning networks, particularly for unsupervised translation techniques to synthesise MRI to sCT and sCT to MRI scans. The tumor volumes include annotations and clinical descriptions of the tumors, which can be used to train models for image synthesis and segmentation. The data is available on Mendeley Data and may help to overcome the lack of publicly available medical image data for AI research.

The main objective in MRI-to-CT synthesis is to train algorithms to convert MRI scans into sCT scans that are similar to CT scans. These algorithms are crucial for radiotherapy, where CT images are generally needed for treatment planning, but the use of MRI is desirable for its better soft tissue visibility and absence of radiation. The datasets listed above are essential in the training and validation of such algorithms through the provision of paired MRI-CT data (SynthRAD2025 and Gold Atlas) or unpaired data (Unpaired MRI-CT Brain), which are used to develop image translation models.

For example, SynthRAD2025, with its diverse anatomical regions and different imaging modalities, aids in the creation of models that can produce sCTs for radiotherapy, such as MRI-only workflows for photon and proton therapy. Likewise, the Gold Atlas dataset, with its organ segmentations, helps to train segmentation models and generate sCTs that can be directly used for radiotherapy treatment. These datasets are useful in addressing the difficulties in registering and synthesising images from multiple modalities, particularly in adaptive radiotherapy, where dynamic updates are required.

On the other hand, the Unpaired MRI-CT Brain Dataset focuses on creating unsupervised learning models for MRI-to-CT translation, where the lack of paired data poses an additional challenge but also allows for greater flexibility in cases where paired data is limited. The additional provision of tumor masks makes the dataset even more valuable, allowing for the development of dedicated models for tumor detection and segmentation in both MRI and CT images.

3. Deep Learning Methods for MRI-to-CT Synthesis: Categories and Applications

With the rapid advancement of deep learning models and the computational power to support them, image-to-image translation models have become increasingly popular for MRI-to-CT image synthesis. The models are intended to change domain-specific features while preserving domain-invariant features. In MRI-to-CT translation, the deep learning model must alter the image's texture while preserving the body's shape. But separating and

extracting image features (features) is a difficult task, and hence deep learning models with different networks and training approaches are used.

To facilitate the review, the current approaches can be grouped in two ways. One is model architecture, which refers to the computational structure used to learn and manipulate image features, such as CNN-based models, and Transformer-based models that incorporate both local (convolutional) and global (attention) processing. The second is generative paradigm, which refers to the approach to learn and generate the target CT image, such as adversarial learning (GAN-based methods) and diffusion-based generation. These two views are not uncorrelated. For instance, a large number of GAN-based methods adopt CNNs for generators, while some recent diffusion models use Transformer blocks for modeling long-range dependencies.

3.1. Model Architecture

3.1.1. CNN-Based Models

CNN-based methods are among the first and most common methods used for MRI-to-CT synthesis. Crucially, the synthesis CNNs are distinct from CNNs for image classification. Classification CNNs eventually map image features to a lower-dimensional semantic space, while synthesis CNNs make dense predictions, that is, they predict a continuous value for each pixel/voxel that corresponds to a CT intensity value. Thus, MRI-to-CT CNNs are usually based on fully convolutional encoder-decoder networks, instead of classification heads, to maintain spatial alignment between MRI and sCT.

These structures involve an encoder that captures features from MRI, and a decoder that synthesises CT-like images from features. U-Net and ResNet-inspired encoder-decoder networks are popular [31,32]. Skip connections can be incorporated to pass detailed anatomical information from the initial encoder layers to the decoder. Residual blocks help ease training and promote feature reuse, particularly at deeper layers. As shown in Figure 1, the network architecture is symmetric with multi-scale feature extraction and reconstruction branches, an encoder and a decoder. The encoder pathway successively downscales the MRI image to extract abstract anatomical and contextual information and the decoder pathway successively upscales the features to produce the sCT image. The paths between the encoder and decoder layers of each stage are connected to retain spatial details and restore lost details of anatomical structures that may occur during downsampling. Moreover, the bottleneck layers offer concise high-level representations and the fully convolutional architecture supports voxel-wise intensity prediction.

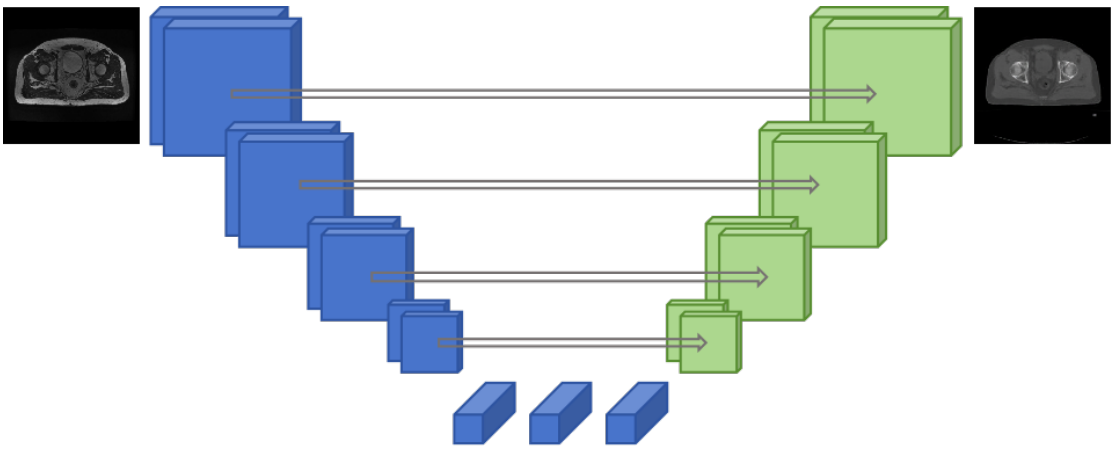

Figure 1. Architecture of a U-Net-like encoder-decoder network for MRI-to-CT synthesis.

Nie et al. used a 3D fully convolutional neural network for patch-based translation, achieving better structural preservation and mean absolute error than atlas-based methods and random forests [33]. Han modified U-Net for 2D slice synthesis and integrated skip connections with pre-trained encoder weights to accelerate convergence and enhance the image quality [22]. Bahrami et al. presented a lightweight CNN with alternative activation functions and extra residual connections to enhance image clarity and reduce the number of parameters [19]. Xiang et al. proposed a deep embedding CNN, which integrates both shallow and deep subnetworks to address the modality differences between MRI and CT [34]. Furthermore, CNN models have been adapted to handle issues such as image misalignment, and data scarcity. Zhao et al. used a Huber-type loss to make the model less sensitive to registration errors [35], and Li et al. proposed a multisequence fusion network to combine complementary information from multiple MRI sequences [36].

CNN-based approaches are still the preferred choice for MRI-to-CT synthesis tasks due to their balance between efficiency, stability and effectiveness. Their local inductive bias is consistent with the continuity of anatomical structures at short spatial resolutions, and encoder-decoder architectures with skip connections have been proven to be highly effective for capturing organ contours and other structural details. Consequently, in the presence of paired MRI-CT datasets, CNNs often achieve strong performance in supervised tasks, and are frequently considered to be "safe" models to learn from.

However, as standard convolutions are primarily used for local feature extraction, CNN models may struggle to capture long-range contextual information or preserve anatomical consistency. Their performance may also be affected by the registration accuracy of paired data, as poor registration may adversely affect voxel-wise training. In addition, the most common pixel-level regression losses may tend to produce smooth synthesis results, which is common in regions with sharp anatomical boundaries (e.g., bone or air interfaces).

3.1.2. Transformer-Based Models

Transformer-based approaches are incorporated into MRI-to-CT synthesis to overcome a limitation of traditional CNNs: the limited receptive field. Self-attention allows Transformers to capture long-range interactions between remotely located segments of the image, enabling the model to capture global anatomical relationships, which are hard to learn using local convolutions. Self-attention works by projecting input feature maps to different embedding representations (query (Q), key (K), and value (V)) using three different 1 × 1 convolutions (Figure 2). These are used to compute the spatial correlations between pixels and the attention weights (normalized by SoftMax) constitute an attention map that identifies crucial long-range relationships. This attention map is then used to weight the value features and thereby integrate long-range spatial features from anatomically correlated but distant parts. In practice, researchers design hybrid solutions that use CNN encoders/decoders and Transformer bottlenecks or attention blocks. The combination makes sense given the respective features of the two sets of blocks: CNNs are efficient at learning local texture and shape features, but Transformers are crucial for learning global contextual information.

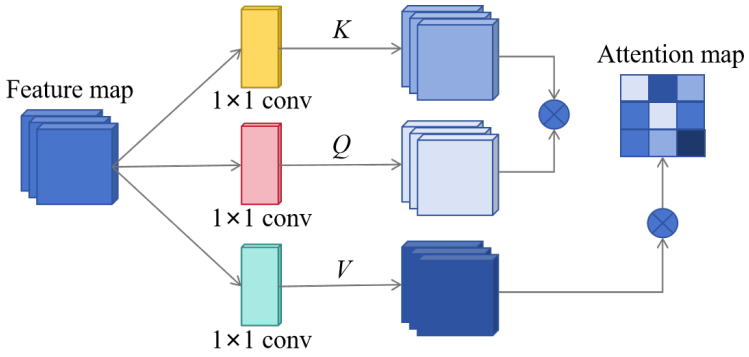

Figure 2. Illustration of the self-attention mechanism in transformer-based models for image processing. The feature maps are processed through separate convolutions to generate K, Q, and V maps, which are then used to compute an attention map, allowing the model to focus on the most relevant image regions.

ResViT features an encoder-decoder generator based on CNNs and aggregated residual Transformer blocks to simultaneously model local and global features for pelvic MRI-to-CT translation [37]. The HMSS-NET model splits global and local processing into two branches: a Transformer bottleneck for low-resolution contextual features and CNN for structural details [38]. MTT-Net takes this concept further with multiscale tokenization and multi-shape window self-attention that allows for more efficient contextual modelling of image patches [39]. Transformers have also been combined with multitask learning. For instance, GenESIS integrates CT synthesis, organ-at-risk segmentation and uncertainty estimation in a single Vision Transformer architecture to enhance CT synthesis and clinical applicability [40]. In the MRI-to-CT synthesis of the abdomen, SeRL leverages self-attention combined with style-embedding learning to decrease the distribution shift between MRI and CT and enhance the synthesis result [41].

The Transformers have an inherent advantage in MRI-to-CT synthesis due to their ability to model global spatial relationships better than conventional CNNs. This makes them well suited in cases where large anatomical structures need to be preserved and in situations where global structural information is crucial for the synthesis.

On the other hand, the enhanced global modeling capabilities of these models may also lead to higher data, memory and computational requirements. They can be difficult to train in small medical datasets, in which training

instability and overfitting can be problematic. Furthermore, the benefits of attention are not always commensurate with its cost: for simple synthesis problems, pure self-attention models may be overcomplicated and their improvement marginal unless the design is carefully crafted. As such, pure Transformer-based and, in particular, CNN-Transformer hybrid approaches are best suited to cases with large or complex anatomy, applications in which global coherence is essential, and cases where sufficient data and resources are available for training.

3.2. Generative Paradigm

3.2.1. GAN-Based Methods

GANs are one of the key generative frameworks for MRI-to-CT synthesis. GAN-based methods, unlike direct regression-based methods which only aim to reconstruct images that are similar in a pixel-wise manner, have a discriminator that assists the generator to produce visually realistic CT images. This design is appealing as pixel-wise losses tend to result in blurry edges and excessively smooth bones.

The MRI-to-CT synthesis using GANs, as shown in Figure 3, involves two players: a generator and a discriminator. The generator, which is usually a U-shaped network with skip connections, receives the MRI image and generates a sCT image. The discriminator, on the other hand, takes the real CT and the synthetic CT (sCT) and tries to classify them as real CT or sCT. The adversarial training process helps the generator not only to reduce the reconstruction error, but also to generate CT images with more realistic anatomical structures, clearer tissue contours and detailed bones.

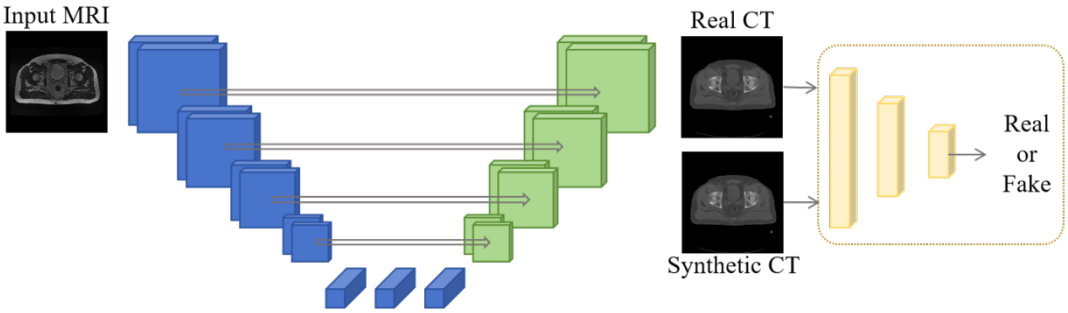

Figure 3. The generator learns to synthesize realistic CT images from MRI, while the discriminator learns to distinguish real CT from sCT.

GANs for MRI-to-CT synthesis can be generally classified into supervised and unsupervised. In supervised GANs, paired MRI-CT images are used to learn a conditional mapping (Pix2Pix-like). In unsupervised GANs, models like CycleGAN eliminate the need for paired data by pairing adversarial learning with cycle-consistency loss.

Supervised approaches use GANs to better preserve realistic and anatomical features than CNN-based regression alone [42]. Emami et al. demonstrated that incorporating adversarial loss to a ResNet-based generator led to better mean absolute error (*MAE*) and structural similarity index measure (*SSIM*), particularly in bone areas [43]. Arabi et al. proposed SynGAN, which uses segmentation guidance to speed up synthesis and generate better attenuation maps, with reduced failure rates [44]. Longuefosse et al. introduced an anatomical feature-focused loss that incorporates features from a pretrained segmentation model to prioritise features such as airways, bone and soft tissue [45]. Other research further improved the supervised GAN by using data augmentation, deformable registration, and customised MRI sequences such as zero echo time imaging [46,47].

These approaches are valuable when high-quality paired data are available and a goal is to render images as realistic and accurate as possible.

For unsupervised tasks, CycleGAN and its variants have become very popular as paired MRI-CT data are costly and may be misaligned. Yang et al. added modality independent neighborhood descriptors, self-attention and spectral normalization to CycleGAN for better anatomical consistency and stability [48]. Ge et al. introduced mutual-information loss and shape-consistency loss to improve anatomical accuracy for different body parts [49]. Kang et al. introduced 2.5D CycleGAN for reduced computational load while maintaining volume information [50]. CUT and CycleCUT also simplified models while preserving structure [51,52]. Other variants like PA-CycleGAN, CycleSGAN and DMFI-Net further incorporated perceptual, structural or multi-task constraints to enhance visual quality and translation performance [53−55]. Quasi-supervised approaches have also been used to generate pseudo-paired data from unpaired MRI and CT images, using measures of image similarity such as

normalized mutual information [56]. This work shows that adversarial learning is still very adaptable, particularly in the context of paired data scarcity.

GAN-based models are a major category of approaches used for MRI-to-CT synthesis since they are able to produce more realistic looking sCT images than pure regression-based models. Another important practical attribute of GANs is their adaptability to data, especially in the case of unpaired MRI-to-CT translation tasks, where they enable simultaneous learning from MRI and CT data without exact correspondence.

But the benefits of GANs also come with their notorious challenges as generative models. Training is generally less stable than that of typical discriminative models, and the results can be highly dependent on the network architecture, loss weighing, and other hyperparameters. Issues such as mode collapse may also reduce the quality and variability of the outputs. In MRI-to-CT synthesis, this can translate into the appearance of artifacts, or anatomically unlikely structures, and may not consistently retain fine anatomy. This is especially critical in unpaired scenarios, where the absence of exact voxel-wise supervision may further exacerbate anatomical misalignment between sCT and MRI. Therefore, even though GAN-based synthesis is desirable when we need sharper sCT results or when paired data are limited, we need to be more careful when using them in some medical applications that require extremely accurate anatomy.

3.2.2. Diffusion-Based Generation

Another recently emerging approach to MRI-to-CT synthesis is diffusion models. Rather than mapping the input MRI to the target CT or learning an adversarial game, diffusion models learn to generate images via denoising. It begins with noisy latent variables and progressively denoises to produce the final target image, given the MRI as a condition. This sampling-based generation approach has gained increasing attention because it can generate high-quality images while being free from some of the adversarial training issues of GANs.

For MRI-to-CT synthesis, diffusion models are usually conditioned on MRI anatomical information so that denoising gradually recovers CT information while maintaining anatomical consistency. Lyu and Wang evaluated various diffusion-based sampling strategies (DDPM and stochastic differential equation variants) and found they produced high-quality images with some level of oversmoothing and artifacts, but still outperformed the state-of-the-art GAN models, albeit with a trade-off between sampling speed and function [57]. Özbey et al. proposed SynDif, which integrates diffusion mechanisms with CycleGAN-like components to handle unaligned data more effectively [58]. Graf et al. evaluated GAN and diffusion models for paired and unpaired spinal MRI-CT translations, with DDIM and SynDif performing on par with the best GAN-based methods [59]. Recently, diffusion models have incorporated Transformer-based components, including Swin Transformer, to enhance reverse diffusion with global information, even with a reduced number of timesteps [60].

Diffusion models have gained popularity for MRI to CT synthesis due to their promising training stability and synthesis quality. Compared to GAN-based methods, they are generally more stable to training and less prone to optimization challenges like mode collapse, and they are well adapted to capturing uncertainty in the synthesis process through their iterative probabilistic nature. By effectively conditioning on MRI features, diffusion models can produce high-quality CT synthesis that is often both plausible and anatomically accurate. However, the major drawback of diffusion models is that they are computationally expensive. They often require substantial resources to train, and inference can be slow due to the need to denoise the image several times. Therefore, their real-world applications can rely on a trade-off between generation speed and quality, achieved through the model and sampling strategy.

While diffusion models are one of the most exciting recent developments, they follow on from a long line of model innovations in MRI-to-CT synthesis. To better understand these developments, Table 2 highlights some exemplar studies with respect to model family, training strategy and anatomical site. As reflected in the table, the research has evolved from initial CNN-based supervised learning to recent diffusion models, transformers, and annotation-free methods. Together, these advances have enhanced image generation and segmentation capabilities, especially in the context of radiotherapy treatment planning, where collecting paired data can be challenging.

Table 2. Overview of MRI-to-CT synthesis models and techniques.

Base Model	Training Data	Feature	Body Region	Year	Paper
CNN	Paired MRI/CT	3D Fully Convolutional Networks (FCN)	Pelvis	2016	[33]
		DCNN	Brain	2017	[22]
		Deep Embedding CNN	Brain/Prostate	2018	[34]
		Multi-2D sub-networks	Abdomen/Pelvis	2018	[35]
		Efficient CNN	Pelvis	2020	[19]
		Adaptive Multi-sequence Fusion	Head/Neck	2023	[36]
Transformer	Paired MRI/CT	Transformer and residual/dense connections	Head/Neck	2023	[38]
		Multi-scale tokens-aware Transformer	Head and neck, abdomen	2024	[39]
		Vision Transformer with multi-task learning	Brain	2025	[40]
	Unpaired MRI/CT	Style-embedding and self-attention	Abdomen	2024	[41]
GAN	Paired MRI/CT	ResNet generator	Brain	2018	[43]
		Anatomical Feature-Prioritized (AFP) loss	Lung/Pelvis	2025	[45]
		Preprocessing pipeline with online data augmentation	Pelvis	2025	[46]
		DL-based CT synthesis from Zero Echo Time (ZTE) MRI sequences	Pelvis	2024	[47]
		Unified image generation and registration	Head/Neck	2024	[61]
		Structure-guided attention and contrastive learning	Whole-body	2025	[62]
	Paired CT/Segmentation maps	GANs for lung image synthesis from CT	Lung	2023	[42]
	Paired PET/CT/MRI	Adversarial Semantic Structure	Brain	2019	[44]
	Weakly paired MRI/CT	CycleGAN with perceptual loss	Abdomen/Thorax	2021	[50]
		Structure-consistency loss with CycleGAN	Brain/Abdomen	2020	[48]
		Cycle-GAN with mutual information loss and shape consistency	Whole-body	2019	[49]
		CycleCUT with contrastive learning	Brain	2022	[52]
		Perceptual attention	Brain/Pelvis	2024	[53]
		Semantic structure preservation	Hip	2024	[54]
	Unpaired MRI/CT	Dynamic Multi-scale Feature Integration	General anatomical regions	2025	[55]
		Staged Multi-Sequence Fusion	General radiotherapy planning	2025	[63]
		Quasi-supervised learning	General radiotherapy planning	2025	[56]
Diffusion	Paired MRI/CT	Score-matching	Pelvis	2022	[57]
		Landmark registration	/	2023	[59]
		Lightweight pipeline	Pelvis	2024	[64]
	Unpaired MRI/CT	Adversarial diffusion modeling	/	2023	[58]

4. Evaluation Strategies for Synthetic CT

The assessment of MRI-to-CT methods can be roughly grouped into three categories: reconstruction quality, performance in downstream tasks and visual quality. The first relates to the reconstruction quality of the sCT, which concerns the accuracy of the generated sCT compared to the reference CT in terms of voxel intensity, anatomical structure, and visual quality. The second focuses on whether the sCT is good enough for certain clinical tasks such as segmentation tasks, registration or radiotherapy planning.

4.1. Reconstruction-Level Evaluation

At the reconstruction level, the similarity between the sCT and the reference CT is commonly quantified using *MAE*, peak signal-to-noise ratio (*PSNR*), *SSIM*, and normalized cross-correlation (*NCC*). Let x_i and y_i denote the voxel intensities of the reconstructed image and the reference image at voxel i, respectively, and let N be the total number of voxels.

The *MAE* measures the average absolute voxel-wise intensity difference between the two images:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - y_i\right| \tag{1}$$

A lower *MAE* indicates better intensity agreement.

The *PSNR* evaluates reconstruction fidelity relative to the mean squared error (*MSE*), and is defined as:

$$PSNR = 10\log_{10}(\frac{MAX^2}{MSE}) \tag{2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{3}$$

where *MAX* denotes the maximum possible intensity value of the image, or the upper bound of the intensity range after preprocessing or normalization. A higher *PSNR* indicates better reconstruction quality.

The *SSIM* evaluates the similarity between two images from the perspectives of luminance, contrast, and structural information:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4}$$

where μ_x and μ_y are the mean intensities of images x and y, σ_x^2 and σ_y^2 denotes are their variances, σ_{xy} is the covariance between them. The constants C_1 and C_2 are small positive values introduced to stabilize the computation when the denominator is close to zero. In practice, *SSIM* is usually computed within local windows and then averaged over the whole image. A higher *SSIM* indicates greater structural similarity.

The *NCC* measures the linear correlation between the reconstructed and reference images:

$$NCC(x,y) = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{N}(y_i - \mu_y)^2}} \tag{5}$$

NCC reflects the degree of intensity pattern correspondence between the two images. Its value ranges from −1 to 1, with values closer to 1 indicating stronger positive similarity.

MAE quantifies voxel-wise absolute intensity discrepancy, *PSNR* reflects signal fidelity relative to reconstruction error, *SSIM* captures structural similarity, and *NCC* measures global linear correspondence between intensity patterns. These metrics provide complementary assessment of reconstruction quality.

4.2. Downstream-Task-Level Evaluation

While reconstruction-level metrics provide important information on intensity consistency and structural similarity, they are inherently limited in evaluating task-oriented usefulness. In medical image synthesis, visual or voxel-wise similarity to the reference image does not always translate into reliable performance in clinically relevant downstream applications. For this reason, many studies further evaluate synthetic images using task-specific metrics, such as segmentation overlap, registration accuracy, or dosimetric agreement, to better characterize their practical value.

4.2.1. Segmentation Overlap

For segmentation-related downstream evaluation, the most commonly used metric is the Dice coefficient (Dice), which quantifies the overlap between a structure derived from sCT and the corresponding reference structure. Given two sets of voxels, A and B, representing the predicted and reference segmentations, respectively, Dice is defined as:

$$\mathrm{Dice}(A,B) = 2\frac{|A \cap B|}{|A| + |B|} \tag{6}$$

where $|A|$ and $|B|$ denote the numbers of voxels in the two segmentations, and $|A \cap B|$ is the number of overlapping voxels. The Dice ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap. In MRI-to-CT synthesis, Dice is used to assess whether tissue classes or anatomically relevant structures derived from sCT remain sufficiently consistent with those obtained from the reference image [65].

4.2.2. Registration Accuracy

In the sCT literature, sCT has been used as an intermediate or bridge modality to improve MRI-CT registration, including head-and-neck and abdominopelvic applications [66,67]. For registration-related evaluation, the target registration error (TRE) measures the residual spatial mismatch between corresponding landmarks after registration. If p_j denotes the j-th landmark in the moving image, q_j the corresponding landmark in the fixed image, $T()$ the estimated transformation, and M the number of landmark pairs, then TRE can be written as:

$$\mathrm{TRE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left\|T(p_j) - q_j\right\|^2} \tag{7}$$

where || || denotes the Euclidean norm. A lower TRE indicates better registration accuracy.

4.2.3. Dosimetric Agreement

For radiotherapy-oriented downstream evaluation, the key question is whether the synthesized image supports dose calculation comparable to that obtained from the reference CT [68−70]. A basic quantity is the dose difference, which can be defined voxel-wise as:

$$\Delta D(r) = D_{\mathrm{sCT}}(r) - D_{\mathrm{CT}}(r) \tag{8}$$

or, in relative form, as:

$$\delta D(r) = \frac{\left|D_{\mathrm{sCT}}(r) - D_{\mathrm{CT}}(r)\right|}{D_{\mathrm{ref}}} \times 100\% \tag{9}$$

where $D_{\mathrm{sCT}}(r)$ and $D_{\mathrm{CT}}(r)$ are the doses calculated on sCT and reference CT at spatial location r, respectively, and D_{ref} is the normalization dose, such as the prescription dose or a chosen reference dose. Smaller values indicate better dosimetric consistency.

A more comprehensive criterion is the gamma index, originally proposed to jointly evaluate dose difference and distance-to-agreement. For a reference point r_m and an evaluated point r_c, the gamma index is defined as:

$$\gamma(r_m) = \min_{r_c}\sqrt{\left(\frac{\left\|r_c - r_m\right\|}{\Delta d_M}\right)^2 + \left(\frac{D_c(r_c) - D_m(r_m)}{\Delta D_M}\right)^2} \tag{10}$$

where Δd_M is the distance-to-agreement criterion, ΔD_M is the dose-difference criterion, $D_m(r_m)$ is the reference dose, and $D_c(r_c)$ is the evaluated dose. A point is commonly considered to pass if $\gamma \leq 1$, and the gamma pass rate is the percentage of evaluated points satisfying this condition.

In the evaluation of sCT, different metrics characterize different aspects of image usefulness. Reconstruction-level metrics mainly assess voxel-wise intensity fidelity and structural resemblance to the reference CT, whereas downstream-task-level metrics evaluate whether the synthesized image is sufficiently reliable for anatomical analysis, spatial alignment, and treatment planning. Therefore, these metrics should be interpreted as complementary rather than interchangeable.

4.3. Qualitative Visual Evaluation

Numerical indicators like *MAE*, *PSNR*, *SSIM*, *NCC*, Dice coefficient, and dose indicators are important quantitative evidence for sCT assessment. But these measures do not always capture variations in visual realism, local anatomical accuracy, or artifacts. Thus, qualitative visual inspections are an integral part of quantitative evaluation, particularly when it comes to discovering the capabilities of different model families in MRI-to-CT synthesis.

For a more visual comparison, a set of sCT images generated under a consistent experimental setting are presented in Figure 4. The first row shows the input real MR image, the reference real CT image, and the sCT images generated by the U-Net-based methods (U-Net, Attention U-Net, R2U-Net and R2Attention U-Net) with the same enlarged regions shown in red boxes for visual comparison. The second row presents the error maps of the sCTs relative to the reference CT. The third row shows the sCT images produced by other methods, i.e., the GAN-based methods (ResVit, Pix2Pix, Pix2PixHD, and RTC-GAN) and the diffusion-based methods (DDPM

and DDIM) with red boxes indicating enlarged local regions for a local view, and the fourth row shows the corresponding error maps for these methods.

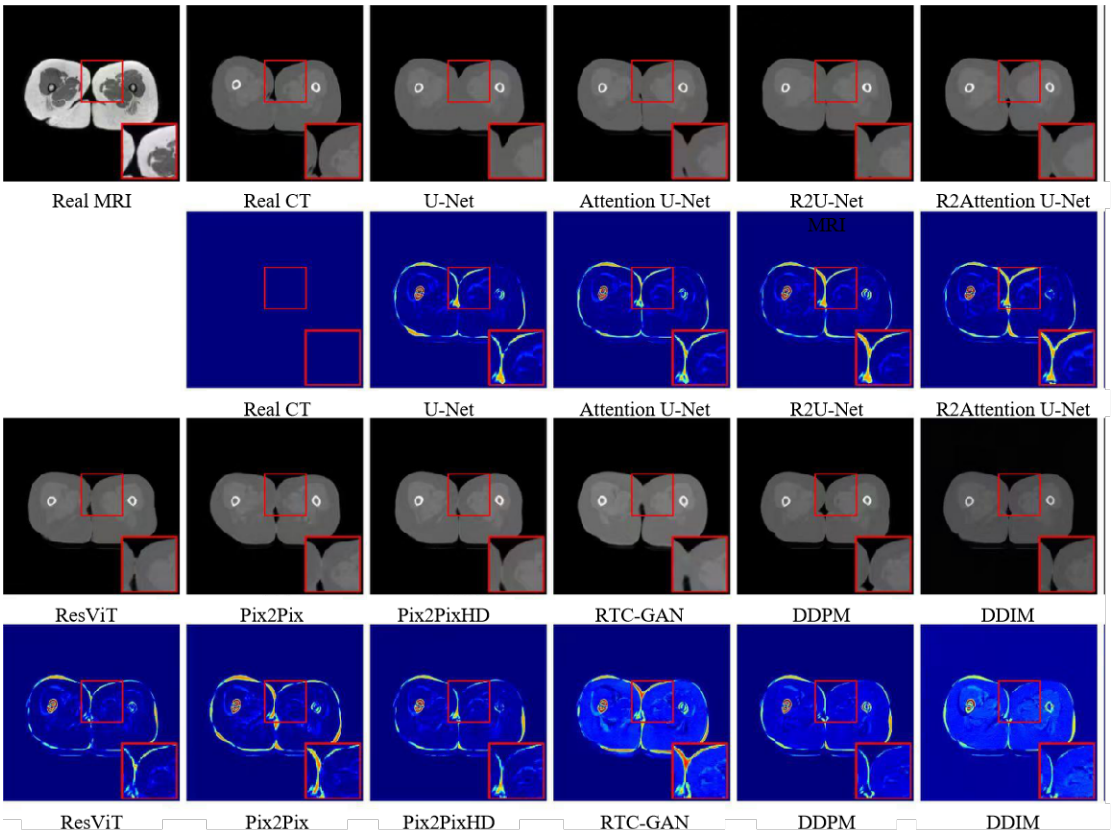

Figure 4. Representative sCT results from different model families under a unified experimental setting. The same MR input and reference CT are used for all methods. Red boxes indicate enlarged local regions for detailed visual comparison. The error maps show the absolute difference between the generated CT and the reference CT.

Qualitatively, the results show that different models have different features. The U-Net-based methods typically exhibit relatively consistent global performance in reconstruction and retain the major anatomical structures, but their local textures are relatively over-smoothed. The GAN-based methods generally produce sharper edges and more visually significant local textures; however local artifacts can also be observed. On the other hand, the diffusion-based methods demonstrate better global consistency, and generate more anatomically plausible images, but these methods typically involve much higher training and inference costs.

It's worth noting that purely visual comparison cannot draw a clear conclusion on the superiority of one model family over another, as the visual performance may still be affected by other aspects, such as the data distribution, data preparation strategies, training schemes, and the network architecture details. However, when conducted under the same experimental framework, representative visual comparison provides a convenient and intuitive extension to quantitative assessment.

5. Discussion

Synthetic medical data sets in medical image analysis, such as MRI-to-CT synthesis, have several benefits, including improved privacy and the ability to create large data sets to train deep learning networks. But there are some challenges to be overcome to guarantee their safe, effective and ethical application. These revolve around data privacy, interpretability, explainability, bias in the training data and clinical barriers of practical usage which can result in the generation of biased and inaccurate synthetic images.

5.1. Privacy and Ethics

Although synthetic data sets support patient privacy by deidentifying data, there are concerns about copying data and reidentification because of the presence of patient-identifying features in medical images [71]. Although explicit identifiers may be excluded, patterns in the data (e.g., facial landmarks in brain MRI scans, or distinctive anatomical features) can still potentially identify patients. To address this concern, it is important to apply

sophisticated privacy-preserving methods such as differential privacy, which adds noise to data to safeguard individual data points while preserving overall data utility. In addition, transparency about the origin and purpose of datasets is essential to prevent unintentional leakage or misuse of data. Academics need to consider the implications of the privacy risks of synthetic datasets and implement techniques that ensure anonymization, enabling safe and ethical use in medical imaging.

5.2. Interpretability and Explainability

As generative models, including deep learning models, become more complex, their interpretability and explainability become more complicated. Knowing how these models learn and generate data is crucial in establishing confidence in their predictions and to safely apply them in high-stake applications such as medical imaging. Although there are some techniques for explainability, like feature map visualisations, or adding uncertainty to the model, these have not yet been adapted and tested for medical imaging. This can lead to a lack of trust in these models for clinical diagnostic and treatment planning. In the future, we need more research on explainability techniques for generative models in medical imaging to ensure their output can be interpreted, understood, trusted and used safely for clinical decision-making [72,73].

5.3. Bias and Representation

A significant challenge with synthetic datasets is the potential for biases in the training data, which can result in generative models that lack diversity and produce discriminatory outcomes. If the training data underrepresents certain demographics, pathologies, or imaging protocols, the synthetic images generated by the model may perpetuate these biases. In MRI-to-CT synthesis, this concern also applies to atypical anatomical structures or lesion-related changes, which are often less represented in the training distribution and may therefore be translated less accurately than common anatomical patterns. This is particularly concerning in healthcare, where biased models can lead to skewed research findings or discriminatory applications of medical technology. To mitigate this risk, strategies such as diversity-aware sampling, adversarial debiasing, and fairness constraints should be incorporated into the training process. These techniques ensure that the generated datasets better represent the diversity of the patient population, improving the generalization of the models and reducing the risk of bias in the generated synthetic images. Regular audits of generated data are also necessary to assess and address any emerging biases [74,75].

5.4. Clinical Translation Barriers

The clinical translation of MRI-to-CT synthesis for MRI-only radiotherapy remains constrained by regulatory, workflow, and validation challenges. First, these systems may fall within the scope of AI-enabled medical device software and therefore require evidence of safety, effectiveness, and lifecycle risk management before deployment. Second, clinical usability depends not only on fast model inference, but also on robust integration with MRI acquisition, preprocessing, treatment planning systems, image guidance, and quality assurance procedures within a time-sensitive workflow. Third, rigorous validation is still needed, because clinically acceptable mean performance may obscure occasional but important failure cases. Therefore, future work should emphasize commissioning standards, prospective validation, and clinically practical fallback mechanisms for safe hospital implementation [76,77].

6. Conclusions

In this paper, we give an overview of recent deep learning methods for the synthesis of CT images from MRI, focusing on four major categories: CNN models, GAN models, Transformer models and Diffusion models. CNN models are basic, providing robust local feature extraction through convolutional layers but are not suited for global feature extraction. GAN models, incorporating adversarial learning mechanisms, are effective at generating realistic images and have been widely used in both paired and unpaired image synthesis tasks. Transformer models, with their attention-based mechanisms, can enhance global feature extraction for better image generation, particularly for large-scale structures such as bones and tissues.

Author Contributions

R.Z.: conceptualization, methodology, software, data curation, writing—original draft preparation; X.L.: investigation, validation; W.Q.: supervision; Y.T.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Science Foundation of Liaoning Province (Grant No. 2022-MS-114) and the International High Impact Paper of Northeast Agriculture University (Grant No. 54960812).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were generated or analyzed in this study. The publicly available datasets discussed in this review can be accessed through the original publications or official data repositories cited in the article.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work the authors used Chat-GPT in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Armanious, K.; Jiang, C.; Fischer, M.; et al. MedGAN: Medical image translation using GANs. *Comput. Med. Imaging Graph.* **2020**, *79*, 101684.
2. Zhan, B.; Li, D.; Wang, Y.; et al. LR-cGAN: Latent representation based conditional generative adversarial network for multi-modality MRI synthesis. *Biomed. Signal Process. Control.* **2021**, *66*, 102457.
3. Khosravi, B.; Purkayastha, S.; Erickson, B.J.; et al. Exploring the potential of generative artificial intelligence in medical image synthesis: Opportunities, challenges, and future directions. *Lancet Digit. Health.* **2025**, *7*, 100890.
4. Wang, Y.R.; Baratto, L.; Hawk, K.E.; et al. Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, *48*, 2771–2781.
5. Merida, I.; Costes, N.; Heckemann, R.; et al. Pseudo-CT generation in brain MR-PET attenuation correction: Comparison of several multi-atlas methods. *EJNMMI Phys.* **2015**, *2*(Suppl. 1), A29.
6. Burgos, N.; Cardoso, M.J.; Thielemans, K.; et al. Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies. *IEEE Trans. Med. Imaging* **2014**, *33*, 2332–2341.
7. Uh, J.; Merchant, T.E.; Li, Y.; et al. MRI-based treatment planning with pseudo CT generated through atlas registration. *Med. Phys.* **2014**, *41*, 051711.
8. Torrado-Carvajal, A.; Herraiz, J.L.; Alcain, E.; et al. Fast patch-based pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in brain studies. *J. Nucl. Med.* **2016**, *57*, 136–143.
9. Sjölund, J.; Forsberg, D.; Andersson, M.; et al. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys. Med. Biol.* **2015**, *60*, 825–839.
10. Torrado-Carvajal, A.; Herraiz, J.L.; Hernandez-Tamames, J.A.; et al. Multi-atlas and label fusion approach for patient-specific MRI based skull estimation. *Magn. Reson. Med.* **2016**, *75*, 1797–1807.
11. Berker, Y.; Franke, J.; Salomon, A.; et al. MRI-based attenuation correction for hybrid PET/MRI systems: A 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence. *J. Nucl. Med.* **2012**, *53*, 796–804.
12. Hsu, S.H.; Cao, Y.; Huang, K.; et al. Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy. *Phys. Med. Biol.* **2013**, *58*, 8419–8435.
13. Zheng, W.; Kim, J.P.; Kadbi, M.; et al. Magnetic resonance–based automatic air segmentation for generation of synthetic computed tomography scans in the head region. *Int. J. Radiat. Oncol. Biol. Phys.* **2015**, *93*, 497–506.
14. Ladefoged, C.N.; Benoit, D.; Law, I.; et al. Region specific optimization of continuous linear attenuation coefficients based on UTE (RESOLUTE): Application to PET/MR brain imaging. *Phys. Med. Biol.* **2015**, *60*, 8047–8065.

15. Izquierdo-Garcia, D.; Hansen, A.E.; Förster, S.; et al. An SPM8-based approach for attenuation correction combining segmentation and nonrigid template formation: Application to simultaneous PET/MR brain imaging. *J. Nucl. Med.* **2014**, *55*, 1825–1830.
16. Yu, B.; Wang, Y.; Wang, L.; et al. Medical image synthesis via deep learning. In *Deep Learning in Medical Image Analysis*; Advances in Experimental Medicine and Biology; Springer: Cham, Switzerland, 2020; Vol. 1213, pp. 23–44.
17. Johansson, A.; Garpebring, A.; Karlsson, M.; et al. Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information–potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. *Acta Oncol.* **2013**, *52*, 1369–1373.
18. Dowling, J.A.; Lambert, J.; Parker, J.; et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2012**, *83*, 5–11.
19. Bahrami, A.; Karimian, A.; Fatemizadeh, E.; et al. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. *Med. Phys.* **2020**, *47*, 5158–5171.
20. Wang, C.; Uh, J.; He, X.; et al. Transfer learning-based synthetic CT generation for MR-only proton therapy planning in children with pelvic sarcomas. In *Medical Imaging 2021: Physics of Medical Imaging*; SPIE Digital Library: Bellingham, WA, USA, 2021; Vol. 11595, pp. 1112–1118.
21. Spadea, M.F.; 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.
22. Han, X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* **2017**, *44*, 1408–1419.
23. Florkow, M.C.; Zijlstra, F.; Willemsen, K.; et al. Deep learning-based MR-to-CT synthesis: The influence of varying gradient echo-based MR images as input channels. *Magn. Reson. Med.* **2020**, *83*, 1429–1441.
24. Zhao, B.; Cheng, T.; Zhang, X.; et al. CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN. *Comput. Med. Imaging Graph.* **2023**, *103*, 102150.
25. Siam, Z.S.; Akon, M.Y.; Munmun, I.J.; et al. A paired CT and MRI dataset for advanced medical imaging applications. *Data Brief.* **2025**, *61*, 111768.
26. Gu, X.; Wang, L.; Deng, Z.; et al. Adaptive spatial and frequency experts fusion network for medical image fusion. *Biomed. Signal Process. Control.* **2024**, *96*, 106478.
27. Nyholm, T.; Svensson, S.; Andersson, S.; et al. MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project. *Med. Phys.* **2018**, *45*, 1295–1300.
28. Thummerer, A.; van der Bijl, E.; Galapon, A.J.; et al. SynthRAD2025 Grand Challenge dataset: Generating synthetic CTs for radiotherapy from head to abdomen. *Med. Phys.* **2025**, *52*, e17981.
29. Clark, K.; Vendt, B.; Smith, K.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057.
30. Al-Kadi, O.S.; Almallahi, I.; Abu-Srhan, A.; et al. Unpaired MR-CT brain dataset for unsupervised image translation. *Data Brief.* **2022**, *42*, 108109.
31. He, K.; Zhang, X.; Ren, S.; et al. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
33. Nie, D.; Cao, X.; Gao, Y.; et al. Estimating CT image from MRI data using 3D fully convolutional networks. In Proceedings of the International Workshop on Deep Learning in Medical Image Analysis, Athens, Greece, 21 October 2016; pp. 170–178.
34. Xiang, L.; Wang, Q.; Nie, D.; et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med. Image Anal.* **2018**, *47*, 31–44.
35. Zhao, Y.; Liao, S.; Guo, Y.; et al. Towards MR-only radiotherapy treatment planning: Synthetic CT generation using multi-view deep convolutional neural networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 286–294.
36. Li, Y.; Xu, S.; Chen, H.; et al. CT synthesis from multi-sequence MRI using adaptive fusion network. *Comput. Biol. Med.* **2023**, *157*, 106738.
37. Dalmaz, O.; Yurt, M.; Çukur, T. ResViT: Residual vision transformers for multimodal medical image synthesis. *IEEE Trans. Med. Imaging* **2022**, *41*, 2598–2614.
38. Li, Y.; Xu, S.; Lu, Y.; Qi, Z. CT synthesis from MRI with an improved multi-scale learning network. *Front. Phys.* **2023**, *11*, 1088899.

39. Zhong, L.; Chen, Z.; Shu, H.; et al. Multi-scale tokens-aware transformer network for multi-region and multi-sequence MR-to-CT synthesis in a single model. *IEEE Trans. Med. Imaging* **2023**, *43*, 794–806.
40. Mekki, L.; Ladra, M.; Acharya, S.; et al. Generative evidential synthesis with integrated segmentation framework for MR-only radiation therapy treatment planning. *Med. Phys.* **2025**, *52*, e17828.
41. You, L.; Wang, H.; Matta, E.J.; et al. SeRL: Style-embedding representation learning for unsupervised CT images synthesis from unpaired MR images. *Biomed. Signal Process. Control.* **2024**, *94*, 106280.
42. Mendes, J.; Pereira, T.; Silva, F.; et al. Lung CT image synthesis using GANs. *Expert Syst. Appl.* **2023**, *215*, 119350.
43. Emami, H.; Dong, M.; Nejad-Davarani, S.P.; et al. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med. Phys.* **2018**, *45*, 3627–3636.
44. Arabi, H.; Zeng, G.; Zheng, G.; et al. 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.
45. Longuefosse, A.; Denis de Senneville, B.; Dournes, G.; et al. Anatomical feature-prioritized loss for enhanced MR to CT translation. *Phys. Med. Biol.* **2025**, *70*, 145012.
46. Han, S.; Hémon, C.; Texier, B.; et al. Balancing data consistency and diversity: Preprocessing and online data augmentation for multi-center deep learning-based MR-to-CT synthesis. *Pattern Recognit. Lett.* **2025**, *189*, 56–63.
47. Getzmann, J.M.; Deininger-Czermak, E.; Melissanidis, S.; et al. Deep learning-based pseudo-CT synthesis from zero echo time MR sequences of the pelvis. *Insights Imaging* **2024**, *15*, 202.
48. Yang, H.; Sun, J.; Carass, A.; et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Trans. Med. Imaging* **2020**, *39*, 4249–4261.
49. Ge, Y.; Wei, D.; Xue, Z.; et al. Unpaired MR to CT synthesis with explicit structural constrained adversarial learning. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1096–1099.
50. Kang, S.K.; An, H.J.; Jin, H.; et al. Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. *Biomed. Eng. Lett.* **2021**, *11*, 263–271.
51. Wang, J.; Wu, X.; Jiang, X.; et al. MRI to CT synthesis using contrastive learning. In Proceedings of the 2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE), Hefei, China, 13–14 November 2021; pp. 1–5.
52. Wang, J.; Yan, B.; Wu, X.; et al. Development of an unsupervised cycle contrastive unpaired translation network for MRI-to-CT synthesis. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13775.
53. Zhu, R.; Liu, X.; Li, M.; et al. A Novel Perceptual Constrained cycleGAN with Attention Mechanisms for Unsupervised MR-to-CT Synthesis. *Int. J. Imaging Syst. Technol.* **2024**, *34*, e23169.
54. Wang, R.; Heimann, A.F.; Tannast, M.; et al. CycleSGAN: A cycle-consistent and semantics-preserving generative adversarial network for unpaired MR-to-CT image synthesis. *Comput. Med. Imaging Graph.* **2024**, *117*, 102431.
55. Tang, M.; Jiang, J.; Zhang, X.; et al. Dynamic Multi-scale Feature Integration Network for unsupervised MR-CT synthesis. *Neural Netw.* **2025**, *189*, 107584.
56. Zhu, R.; Ruan, Y.; Li, M.; et al. Quasi-supervised MR-CT image conversion based on unpaired data. *Phys. Med. Biol.* **2025**, *70*, 125010.
57. Lyu, Q.; Wang, G. Conversion between CT and MRI images using diffusion and score-matching models. *arXiv* **2022**, *preprint*, arXiv:2209.12104.
58. Özbey, M.; Dalmaz, O.; Dar, S.U.; et al. Unsupervised medical image translation with adversarial diffusion models. *IEEE Trans. Med. Imaging* **2023**, *42*, 3524–3539.
59. Graf, R.; Schmitt, J.; Schlaeger, S.; et al. Denoising diffusion-based MRI to CT image translation enables automated spinal segmentation. *Eur. Radiol. Exp.* **2023**, *7*, 70.
60. Pan, S.; Wang, T.; Qiu, R.L.; et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Phys. Med. Biol.* **2023**, *68*, 105004.
61. Li, X.; Bellotti, R.; Bachtiary, B.; et al. A unified generation-registration framework for improved MR-based CT synthesis in proton therapy. *Med. Phys.* **2024**, *51*, 8302–8316.
62. Zheng, J.; Shen, Z.; Zhang, L.; et al. Structure-guided MR-to-CT synthesis with spatial and semantic alignments for attenuation correction of whole-body PET/MR imaging. *Med. Image Anal.* **2025**, *103*, 103622.
63. Deng, L.; Sun, H.; Huang, S.; et al. Research on multi-sequence MR image synthesis CT algorithm based on unsupervised method. *Biomed. Signal Process. Control.* **2025**, *99*, 106819.
64. Zhuang, Y.; Mathai, T.S.; Mukherjee, P.; et al. Segmentation of pelvic structures in T2 MRI via MR-to-CT synthesis. *Comput. Med. Imaging Graph.* **2024**, *112*, 102335.
65. Hsu, S.H.; DuPre, P.; Peng, Q.; et al. A technique to generate synthetic CT from MRI for abdominal radiotherapy. *J. Appl. Clin. Med. Phys.* **2020**, *21*, 136–143. https://doi.org/10.1002/acm2.12816.

66. McKenzie, E.M.; Santhanam, A.; Ruan, D.; et al. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. *Med. Phys.* **2020**, *47*, 1094–1104. https://doi.org/10.1002/mp.13976.
67. Heo, J.U.; Zhou, F.; Jones, R.; et al. Abdominopelvic MR to CT registration using a synthetic CT intermediate. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13731. https://doi.org/10.1002/acm2.13731.
68. Low, D.A.; Harms, W.B.; Mutic, S.; et al. A technique for the quantitative evaluation of dose distributions. *Med. Phys.* **1998**, *25*, 656–661. https://doi.org/10.1118/1.598248.
69. Low, D.A.; Dempsey, J.F. Evaluation of the gamma dose distribution comparison method. *Med. Phys.* **2003**, *30*, 2455–2464. https://doi.org/10.1118/1.1598711.
70. Wang, H.; Chandarana, H.; Block, K.T.; et al. Dosimetric evaluation of synthetic CT for magnetic resonance-only based radiotherapy planning of lung cancer. *Radiat. Oncol.* **2017**, *12*, 108.
71. Akbar, M.U.; Wang, W.; Eklund, A. Beware of diffusion models for synthesizing medical images—A comparison with GANs in terms of memorizing brain MRI and chest X-ray images. *Mach. Learn. Sci. Technol.* **2025**, *6*, 015022.
72. Tang, R.; Liu, L.; Pandey, A.; et al. What the daam: Interpreting stable diffusion using cross attention. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Canada, 9–14 July 2023; pp. 5644–5659.
73. Horwitz, E.; Hoshen, Y. Conffusion: Confidence intervals for diffusion models. *arXiv* **2022**, *preprint*, arXiv:2211.09795.
74. Luccioni, A.S.; Akiki, C.; Mitchell, M.; et al. Stable bias: Analyzing societal representations in diffusion models. *arXiv* **2023**, *preprint*, arXiv:2303.11408.
75. Seyyed-Kalantari, L.; Zhang, H.; McDermott, M.B.; et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat. Med.* **2021**, *27*, 2176–2182.
76. Earwong, P.; Puttanawarut, C.; Suphaphong, S.; et al. Clinical implementation of deep learning-based synthetic CT for MRI-only volumetric modulated arc therapy in head and neck and pelvic cancer patients. *Radiat. Oncol.* **2025**, *20*, 166.
77. Aljaafari, L.; Speight, R.; Buckley, D.L.; et al. Clinical validation of using a commercial synthetic-computed tomography solution for brain MRI-only radiotherapy treatment planning. *Tech. Innov. Patient Support Radiat. Oncol.* **2025**, *35*, 100328.