Sri Lankan Journal of Applied Statistics, Vol 26 (Special Issue)
Selected Papers from International Statistics Conference (ISC) 2024

Advancements in Cross-modality Imaging: A Survey of CNN and GAN Applications in Medical Imaging

K. Thahir[1,*] and N. Meyen[2]

[1]Department of Physics, Faculty of Applied Sciences,
University of Sri Jayewardenepura, Sri Lanka

[2]Cogniata (Pvt) Ltd, Colombo, Sri Lanka

*khadeeja20184@gmail.com

Received: 31st August, 2025/ Revised: 03rd November, 2025/ Published: 30th November, 2025

ABSTRACT

This paper presents a comprehensive survey of recent advancements in cross-modality imaging techniques. It focuses on the application of Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) in medical imaging. Cross-modality imaging involves translating images from one modality to another (E.g.: MRI to CT scans) and can assist medical experts in accurate diagnostics and patient care. CNNs and GANs have been quite popular in various fields and their abilities to work with images and generate high quality outputs are quite an advantage for a situation like this. This survey highlights several studies utilizing CNN methodologies for various applications, including synthetic MRI generation, MRI to CT translations for different anatomical regions and bone structure identification. GANs, on the other hand, excel in generative modeling by training two neural networks, the generator and the discriminator, in a competitive setting. We discuss their application in generating missing MRI modalities, creating 3D medical images, and producing synthetic CT scans for radiotherapy. The integration of CNN and GAN models in cross-modality imaging shows significant promise in improving image quality, reducing noise, and enhancing the accuracy of diagnostic tools. This survey underscores the importance of advanced deep learning techniques in medical imaging and sets the stage for future research in this rapidly evolving field.

Keywords: Medical imaging, Synthetic CT/MRI, Machine learning, Image generation, Image synthesis, Deep learning

1 Introduction

In recent years, the intersection of deep learning techniques and medical imaging has led to groundbreaking advancements in image-to-image translation within the realm of medical physics. The ability to use Machine Learning techniques to convert between Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI) scans & MRI to CT scans has emerged as a transformative capability with profound implications for medical diagnosis and treatment planning. Through the use of Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and other sophisticated models, researchers have made significant strides in bridging the modalities of CT and MRI imaging.

The present study surveys a selection of research papers that utilize CNN and GAN model frameworks for medical image translation. The referenced papers have been obtained from the publication "Machine Learning for Medical Image Translation: A Systematic Review" (McNaughton et al., 2023). The study has focused on contemporary research employing these models and insights have been provided for each reference paper.

By highlighting the nuances of image-to-image translation model frameworks, this research aims to contribute to the ongoing discussion on the advancement and optimization of deep learning techniques in medical imaging. The goal is to compile and analyze key research studies in this field to deepen our understanding of these cutting-edge technologies and to lay the groundwork for future research. Through this effort, further exploration and development in the promising area of medical image translation are expected to be encouraged, advancing both research and clinical applications.

2 Model Frameworks

The analysis focused on two main model frameworks, and the research papers have been categorized based on the model type used. The two types of model frameworks identified in the selected research papers are Conventional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). Each study has been reviewed in detail and the key findings related to cross-modality imaging and the broader field of medical imaging have been presented.

2.1 Conventional Neural Networks (CNN)

CNNs represent a widely used model framework in deep learning, designed for processing and analyzing visual data. These networks consist of multiple layers, including convolutional, pooling and fully connected layers, which operate together to automatically learn and extract features from input images. Through successive transformations, patterns, edges, textures, and other complex structures are effectively recognized, producing highly accurate outputs for various tasks such as image classification, image segmentation and translation.

In the studies reviewed, two cross-modality imaging techniques have been utilized: Cross-MRI imaging and MRI-to-CT conversion.

2.1.1 Cross-modality MRI image translation

Magnetic Resonance Imaging (MRI) comprises various modalities, such as T1-weighted, T2 weighted, T1-contrast imaging, each highlighting different anatomical features. Cross-modality MRI image translation refers to the process of converting one type of MRI scan into another. In this study conducted by Pal, Dutta and Maitra (2023), this method was implemented using a CNN model. Deep Image Prior (DIP) with a U-net denoising architecture was utilized to generate synthetic MRI scans. The generated scans were assessed using three evaluation metrics:

- Root Mean Square Percentage Error (RMSPE) – Measures the difference between the original and synthetic MRI scans, placing greater emphasis on larger errors. This metric highlights the precision of the generated scans, especially in regions displaying substantial discrepancies.

- Mean Absolute Percentage Error (MAPE) - Represents the average percentage difference between the original and synthetic scans, depicting the resemblance of the two scans.

- Structural Similarity Index (SSIM) – Compares the overall structural quality between the original and synthetic scans, focusing on visual similarity by accounting for brightness, contrast and texture.

The DIP method has been introduced to address the limitations of traditional methods such as:

- Least Squares (LS) - A statistical method that minimizes the sum of the squared differences between predicted and actual values. In the context of image translation, LS often struggles with noisy data, as it assigns equal weight to all the errors, which reduces accuracy in complex or high-noise environments.

- Maximum Likelihood Estimators (MLE) – A method that determines the parameters most likely to produce the observed data by maximizing the likelihood function. Although effective in multiple contexts, MLE can be highly sensitive to noise, leading to reduced reliability in situations involving complex image data or significant noise levels.

By applying the DIP method, more accurate and visually consistent MRI scans have been generated, compared to traditional approaches, which often fail to maintain performance under high-noise conditions.

2.1.2 MRI to CT cross-modality imaging

Several papers have focused on MRI to CT translations, addressing different anatomical regions such as hip, knee, head, neck, abdomen, brain, and whole body.

2.1.2.1 Bone MRI Software

In the study conducted by (Florkow et al., 2022), hip MRI and CT scans from 30 male patients were used to develop a model for generating synthetic CT (sCT) scans. Synthetic CT (sCT) refers to computed tomography images generated from MRI scans using machine learning techniques rather than a CT machine. This method provides CT-equivalent visualizations of bone structures without subjecting patients to additional radiation exposure.

A commercially available software, BoneMRI (MR Guidance, n.d.), was utilized to generate sCT scans from a single MRI scan. The model was developed using a 3D patch-based U-Net like architecture. The synthetic CT images produced possess the same resolution, orientation, and matrix size as the corresponding MRI scans, ensuring high compatibility and facilitating accurate comparisons between MRI and sCT outputs. By using this approach, clinicians can obtain both MRI and sCT data from a single scan, thereby achieving a more efficient and patient-friendly imaging process.

The above figure 1 demonstrates the striking similarity between the CT and synthetic CT (sCT) scans, with no major differences observed. The arrows highlight minute details that the framework successfully captures during the translation process. According to the study, the uncertainty associated with each parameter was calculated at a 95% confidence level, indicating high reliability of the model's performance.

However, certain limitations have been identified. The study has been conducted on asymptomatic patients who do not exhibit any clinical signs of the disease. Despite this, the model successfully detected subtle variations in bone shape, leading the authors to suggest that patients with bone bumps or abnormal growths could also be identified using sCT, with results comparable to conventional CT scans. Additionally, demographic factors such as age or sex

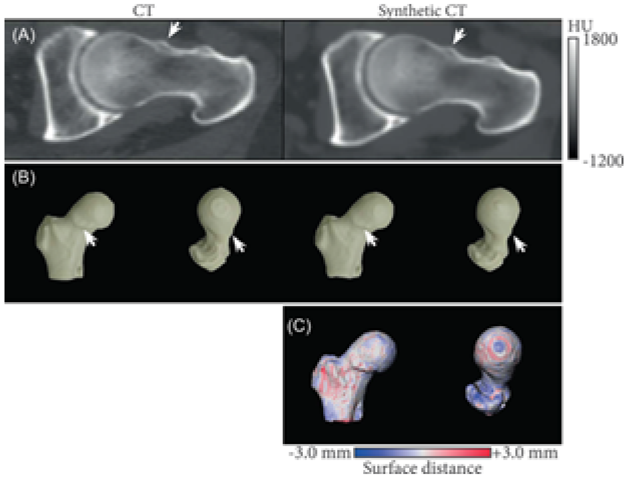

Fig. 1: CT and synthetic CT scans obtained in the BoneMRI study.

have not been explicitly considered. While the authors suggest that that sex may not significantly influence the model, age could play a role, as bone density varies with age and may impact the accuracy of synthetic CT scans.

2.1.2.2 sCT generator for Proton Therapy

The study conducted by Zimmermann et al. (2022) has aimed to develop and evaluate a synthetic CT (sCT) generator that is independent of specific MRI sequences, for use in MRI-guided proton therapy targeting the head and neck regions. The purpose has been to create a versatile model capable of generating accurate sCT scans from various MRI sequences, including T1-weighted, T2-weighted, and contrast-enhanced T1 (T1CM), without dependence on a particular imaging protocol. Such flexibility allows for more efficient and accurate treatment planning in proton therapy.

- T1-weighted MRI: Primarily used to highlight fat tissue and provide clear anatomical details, making it suitable for visualizing the structure of the brain, spine, and other organs.

- T2-weighted MRI: Highlights fluid in the body, such as cerebrospinal fluid, and is effective in identifying areas of inflammation, edema, or tumors.

- Contrast-enhanced T1 (T1CM): Uses a contrast agent to enhance the visibility of blood vessels, tumors, or other structures, providing more detailed imaging compared to standard T1 scans.

The final model has been trained using several key hyperparameters, including group normalization and multiple down-convolution steps. Group normalization normalizes data within group of features, improving the model's learning efficiency, while down-convolution reduces the data size and increases the number of features, enabling the extraction of critical patterns and details from the input images.

The model has demonstrated strong performance on test datasets, achieving low Mean Absolute Error (MAE) values and high Structural Similarity Index (SSIM) scores across different MRI sequences, indicating accurate image generation. Additionally, minimal differences have been observed in the original proton therapy treatment plans and those recalculated on sCTs demonstrating high consistency. These results confirm the reliability of the sCT generator for clinical implementation. Importantly, the real-world data from proton therapy patients have been utilized, enhancing the practical relevance and potential for real-world application of the study's findings.

2.1.2.3 MRI-Only Radiotherapy for Brain Tumors

The study conducted by Lerner et al., (2022) has explored the feasibility of using MRI-only radiotherapy (RT) for the treatment of brain tumors, aiming to reduce potential errors and streamline the planning process by eliminating the need for CT scans. While synthetic CT (sCT) methods had been validated for head regions in previous studies, they had not yet been clinically implemented for brain tumor treatment. This study has included 21 glioma patients, with MRI Dixon images utilized to generate sCTs through a commercially available deep learning software.

The software has been developed using a 3D deep convolutional neural network (CNN), designed specifically to generate sCT images from Dixon MRI images. Dixon imaging is an MRI technique that separates fat and water components in the body, producing four image types: fat, water, in-phase, and out-of-phase. The generated sCTs have been applied for radiotherapy planning and radiation dose calculation. To verify the accuracy of these sCTs, CT scans have been performed for each patient as part of the quality assurance (QA) process.

The QA results have demonstrated a strong correspondence between the sCT and actual CT images, showing excellent agreement in dose distribution. Dose deviations remain within $\pm 1\%$, and patient positioning accuracy is maintained within ± 1 mm. These findings demonstrate that MRI-only radiotherapy represents a viable approach for brain tumor treatment, enabling reliable treatment planning and precise patient alignment without the need for CT scans.

2.1.2.4 Bone MRI to sCT translation

The study conducted by Bambach and Ho (2022), has focused on generating synthetic CT (sCT) images from bone MRI scans in the head and neck region through deep learning techniques. MRI and CT scans have been obtained from 39 patients, and these images have been aligned and trained using two deep learning models: Light_U-Net and VGG-16 U-Net.

- Light_U-Net represents a simplified version of the U-Net architecture, commonly applied in medical imaging. Its effectiveness lies in capturing both local and global image features while maintaining computational efficiency, making it suitable for smaller datasets.
- VGG-16 U-Net integrates the U-Net structure with the VGG-16 model, a recognized convolutional neural network (CNN) known for deep feature extraction. This combination enables the model to leverage the pre-trained VGG-16 layers for enhanced feature learning in medical images.

Both models, Light_U-Net (with 2 million parameters) and VGG-16 U-Net (with 29 million parameters) have been evaluated with and without transfer learning. Transfer learning is a technique in which a model pre-trained on a large dataset is fine-tuned for a different but related task. This approach is particularly useful when the available data is limited, as it improves model performance by transferring previously learned features from one task (like general image classification) to another (such as medical imaging).

The models were assessed using several loss functions and performance metrics:

- Loss functions quantify the difference between the model's predicted outputs and the actual target values, guiding the learning process. The study has used Mean Absolute Error (MAE), Mean Squared Error (MSE) and weighted averages as loss functions.
- Performance metrics evaluate the accuracy and quality of model predictions. Metrics such as Pearson's correlation efficient (R), Mean Absolute Error (MAE), Mean Squared Error (MSE), bone precision and bone recall have been used to measure the degree of similarity between generated sCT images and the reference CT scans.

The Light_U-Net model demonstrated the best overall performance, particularly in the accurate detection of bone structures. Findings have indicated that models trained with data collected from diverse sources perform more effectively, suggesting that this technique can be widely adopted in clinical practice. This approach enhances medical imaging workflows and reduces the reliance on ionizing radiation in diagnostic procedures.

2.2 Generative Adversarial Networks (GANs)

Generative Adversarial Networks, also known as GANs, have been recognized as a prominent class of deep learning models used for generative tasks. Each GAN is composed of two sub-models: the generator and the discriminator. The generator is designed to create new data samples based on the model's specifications, while the discriminator is tasked with evaluating the authenticity of these generated outputs.

The objective of the generator is to produce data that closely resembles real data, whereas the discriminator aims to distinguish between real and synthetic samples. This interaction functions as an adversarial process in which both networks are trained simultaneously. Over successive iterations, the generator progressively improves its output until the discriminator can no longer reliably differentiate between real and generated data, resulting in highly realistic outputs.

GANs have become widely used for their capability to generate highly realistic images, perform image-to-image translation, and synthesize photorealistic representations of objects or human faces. Their use in various fields, including medical imaging, has demonstrated strong potential for improving data generation and domain adaptation tasks.

2.2.1 Generating Missing MRI Modalities

The study conducted by Raju et al. (2022) has focused on generating missing MRI modalities from existing ones using a deep learning Generative Adversarial Network (GAN) framework. The evaluation has been conducted on the BraTS'18 dataset (Menze et al., 2015), which contains records from 75 patients with Low-Grade Glioma (LGG) and 210 patients with High-Grade Glioma (HGG), encompassing T1, T1c (T1-contrast), T2, and FLAIR modalities. The model has been trained to perform target modality generation by synthesizing one missing MRI modality from any three of the four available inputs.

Model performance has been assessed through qualitative and quantitative analyses using metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The PSNR metric is applied to identify image quality by comparing reconstructed images against reference ones. Results have indicated that the synthesized T1c images closely correspond to real

T1c images for both HGG and LGG patients, demonstrating the model's effectiveness in generating contrast-enhanced images without the need for contrast agents.

Furthermore, the proposed model has been compared against baseline architecture such as MM-GAN, capable of producing multiple output types from a single input, and StarGAN, designed for image-to-image translation across diverse domains. A comprehensive style analysis has been performed, in which the model has successfully extracted and reflected distinct styles within and across MRI modalities, further demonstrating its efficiency.

Visual evaluations have been conducted by two expert radiologists, who blindly evaluated 400 images (both real and synthesized) based on image quality and structural detail. The generated images have received ratings nearly equivalent to real images, with synthesized FLAIR images achieving even higher scores. The model has been commended for effectively capturing multiple styles within each domain, a feature particularly advantageous for longitudinal studies.

Overall, this study demonstrates that generative models for style transfer can effectively produce missing MRI modalities, as validated by both quantitative metrics and expert evaluations. The approach proves highly practical for clinical applications, especially in generating contrast-enhanced images without the use of contrast agents, thereby reducing cost, scan time, and patient exposure while maintaining diagnostic accuracy and image quality.

2.2.2 Generating Cross-MRI images for newborns

T1- and T2-weighted images play a crucial role in distinguishing tissues and anatomical structures in MRI scans. However, obtaining these images from awake newborns presents significant challenges due to frequent motion and the lower tissue contrast typical in early development stages. Such factors often result in poor-quality or incomplete images, thereby complicating image processing and leading to the exclusion of some subjects from neuroimaging studies. While several recent methods for generating synthetic MRI images have been evaluated in adults, their application in infants has remained limited. The study conducted by Kaplan et al. (2022) has introduced two novel methods for generating pseudo-T2-weighted (pseudo-T2w) images specifically designed for newborns:

Method 1: 3DGAN-T2w

The 3D GAN-T2w approach uses a three-dimensional Generative Adversarial Network (GAN) to translate T1-weighted (T1w) into T2-weighted (T2w) images. The network comprises two generators and two discriminators:

one generator produces T2w images from T1w inputs, while the other performs the reverse mapping. Two loss functions are implemented: adversarial loss to enhance image realism and cycle-consistency loss to ensure accurate reconstruction of the original images. The model processes full volumetric data, thereby avoiding banding artifacts often encountered in slice-by-slice generation. Trained on paired T1w and T2w scans from 107 neonates, the model has been shown to generate anatomically accurate pseudo-T2w images.

Method 2: Kaplan-T2w

The Kaplan-T2w method generates pseudo-T2w images by leveraging high-quality, age-specific reference data. Aligned T1w and T2w images are grouped into "banks" of 10 subjects according to development stage, and a brain mask is applied to restrict computation to brain voxels. Advanced Normalization Tools (ANTs) are used to estimate deformation fields between the reference bank's T1w images and the target individual's T1w image, producing multiple T1w and T2w estimations. Multi-atlas registration, a technique combining information from multiple reference images, is used to merge these estimations into an optimal pseudo-T2w image. The resulting image undergoes additional refinement through denoising and histogram-based contrast enhancement to improve texture and structural clarity.

Both methods have demonstrated the ability to generate anatomically precise pseudo-T2w images that integrate seamlessly into existing MRI analysis workflows. Notably, the use of pseudo-T2w images in resting-state functional MRI has yielded results nearly identical to those derived from real T2w scans, underscoring their readability. These advances contribute to expanding the inclusion of newborns in MRI- based research while improving image consistency and diagnostic potential.

2.2.3 "Make-A-Volume" Method

The study conducted by Zhu et al. (2023) has introduced a novel Cross-MRI technique known as the "Make-A-Volume" method, which combines 2D and 3D approaches to construct volumetric medical images from multiple imaging modalities. A latent diffusion model has been utilized to efficiently map 2D slices, which are then refined using 3D data. This approach extends 2D image synthesis into 3D image generation with minimal additional computational requirements.

The method has been evaluated through both quantitative and qualitative results. For the quantitative assessment, "Make-A-Volume" has been compared to existing medical image generation techniques, including 2D-based models such as Pix2pix and Palette, and 3D-based models such as 3D Pix2pix and CycleGAN. Metrics including Mean Absolute Error (MAE), Structural

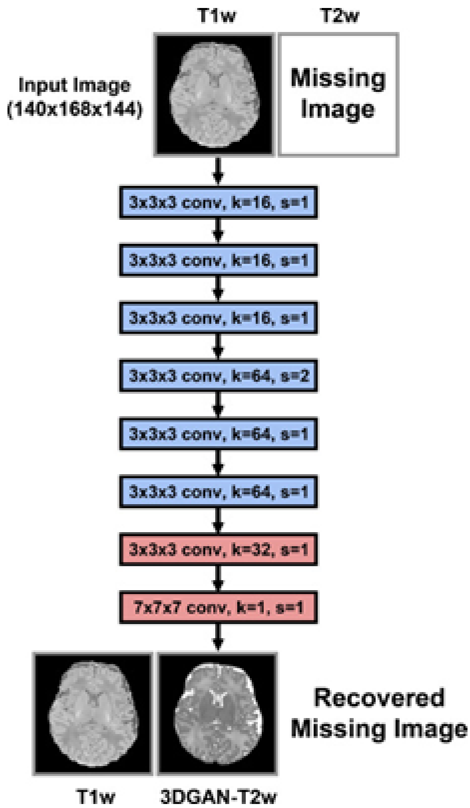

Fig. 2: Method to generate cross-MRI images.

Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR) have been used to measure comparative performance. The results have shown that "Make-A-Volume" outperforms the other methods across all metrics. Among the compared models, 3D versions generally demonstrate improved image quality over the 2D counterparts that require greater memory resources. The 2D Palette method has exhibited stability issues, producing inconsistent image quality across slices.

The qualitative analysis, based on visual inspection, has revealed that the proposed "Make-A-Volume" method generates more detailed and stable volumetric outputs, especially in representing structures such as blood vessels and ventricles. In contrast, 2D-based methods, particularly Palette, have been observed to produce volumes of inconsistent quality across slides.

An Ablation analysis has been conducted to confirm the contribution of the volumetric fine-tuning process, demonstrating its role in enhancing overall model performance and structural fidelity.

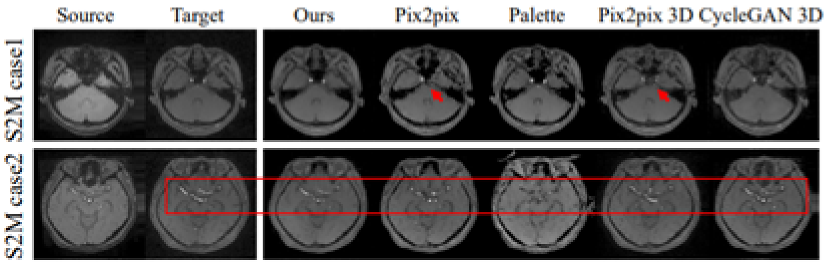

Fig. 3: Qualitative comparison of volumetric outputs.

2.2.4 COSMOS Framework

In cross-modality imaging, pixel-level annotation plays a critical role in accurately labeling each individual pixel of an image for tasks such as segmentation. In MRI or CT scans, every pixel is typically classified into specific categories, such as tissue type, organ or abnormality. This level of annotation is essential for training precise deep learning models that can reliably identify structures and boundaries within medical images. However, the annotation process is highly labor-intensive and requires specialized expertise. This study by Shin et al., (2023) has addressed this challenge through an unsupervised domain adaptation framework designed to minimize dependency on detailed labeled data.

The proposed method, referred to as COSMOS (Contrastive Self-training-based Unsupervised Domain Adaptation for Medical Segmentation), is a self-training-based framework for 3D medical image segmentation. The approach has been validated using automatic segmentation of Vestibular Schwannoma (VS) and cochlea on high-resolution T2-weighted MRI scans. VS is a benign tumor of the vestibular nerve that connects the inner ear to the brain and affects both balance and hearing. The cochlea, another critical part of the inner ear, converts sound vibrations into neural signals transmitted to the brain.

The COSMOS model incorporates a contrast conversion network that transforms annotated T1-weighted MRI scans into pseudo-T2-weighted MRI images while preserving anatomical fidelity. This is followed by an iterative self-training process that uses unlabeled data to refine pseudo-labels and incrementally improve segmentation accuracy. The framework has achieved first place in the Cross-modality Domain Adaptation (crossMoDA) challenge at MICCAI 2021, with high Dice scores and low Average Symmetric Surface Distance for both VS and cochlea segmentation tasks.

The network structure used in the COSMOS method is based on a 2D CycleGAN comprising two 3-level U-net based generators and two PatchGAN discriminators. The generators perform bidirectional image translation between domains, while the discriminators distinguish between real and generated images. A segmentation network attached to the generators shares their encoder and predicts segmentation masks for Vestibular Schwannoma (VS) and cochlea using paired annotated images. The model inputs are normalized using min-max scaling and trained for 50 epochs with the Adam optimizer.

The 3D nnU-Net is then trained to estimate updated pseudo-labels. This network uses a deep supervision scheme with six down sampling operations. Preprocessing steps include image cropping, resampling, and normalization. Training is performed using Dice and Cross Entropy loss functions, optimized with stochastic gradient descent, and enhanced with data augmentation techniques. The study's main findings are summarized as follows:

1. Target-aware Translation: Domain translation utilizing segmentors has preserved the shapes of VS and cochlea more effectively, leading to improvements in segmentation accuracy.

2. Dataset Combination: The integration of labeled pseudo-T2 and pseudo-labeled real T2 images has enhanced model generalizability and segmentation performance.

3. Iterative Pseudo-label Updating: The iterative self-training approach has progressively refined pseudo-labels, resulting in steady performance improvement across training iterations.

4. Comparison with Existing Methods: The COSMOS framework outperformed competing methods in the Cross-Modality Domain Adaptation challenge, achieving state-of-the-art results for both VS and cochlea segmentation.

2.2.5 Synthetic CT for Carbon Ion Radiotherapy

Generating synthetic CT (sCT) images for carbon ion radiotherapy (CIRT) presents significant challenges, as precise treatment planning and accurate dose delivery are critical, particularly in anatomically complex regions such as the abdomen. The study conducted by Parrella et al. (2023) has aimed to address this challenge by developing a Generative Adversarial Network (GAN) model optimized for sCT generation.

In this work, 39 paired MRI and CT scans of the abdomen have been utilized to train a specialized three-channel conditional Generative Adversarial Network (cGAN) designed to generate sCTs representing air, bones, and soft tissue components. The model has been evaluated using five additional MRI scans through two complementary approaches: (i) using CT-based segmentations to evaluate the quality of the sCTs and (ii) using manual MRI segmentations to mimic an MRI-only treatment setup.

Quantitative evaluations have been conducted using similar metrics such as MAE (Mean Absolute Error) and geometrical criteria such as Dice Coefficients (DC). In addition, recalculated CIRT treatment plans have been analyzed through dose volume histograms, gamma analysis, and range shift analysis. CT-based evaluations have demonstrated high accuracy with low MAE values across bones, soft tissues, and air regions. In the MRI-only setup, MAE values for bone regions have been slightly higher, though overall performance has remained clinically acceptable.

The model has achieved a high gamma pass rate and minimal range shift, indicating strong consistency between the synthetic and reference datasets. These results demonstrate that the 3-channel cGAN framework is capable of generating reliable abdominal sCTs suitable for precise CIRT dose calculations, thereby supporting the feasibility of MRI-only workflows in carbon ion radiotherapy planning.

3 Discussion

The study used the Deep Image Prior (DIP) method combined with a U-net denoising architecture has demonstrated promising results in cross-MRI image translation. The synthetic MRI scans have exhibited high structural similarity to real images, as supported by evaluation metrics such as RMSPE, MAPE and SSIM. These findings indicate that the DIP method effectively addresses noise-related limitations that traditional methods are unable to overcome.

Across reviewed studies on MRI-to-CT image translation, diverse anatomical regions have been explored. In the study on hip morphology, the generated sCTs have been validated against real hip morphology, and independent readers have confirmed their accuracy. These results suggest that sCTs can be reliably applied in clinical assessments of hip anatomy. The BoneMRI software, a commercially available solution utilizing a 3D patch-based U-Net like architecture, has produced sCT scans of bone morphology with a high level of confidence, indicating its potential as a diagnostic tool for conditions like osteoarthritis. In studies addressing MRI-based radiotherapy applications, several promising outcomes have been reported. The MRI sequence-independent sCT generator developed for proton therapy has shown minimal differences in recalculated treatment accuracy. Similarly, the study on MRI-only radiotherapy for brain tumors has demonstrated the feasibility of using MRI-generated sCTs for treatment planning, offering a viable alternative to conventional CT-based approaches. In bone MRI-to-sCT translation, the Ligh_U-Net model has outperformed other architectures, effectively identifying bone structures that are critical for accurate medical diagnosis and treatment planning.

Generative Adversarial Network (GAN) based-frameworks have also shown considerable potential in multimodal imaging. The model used for generating missing MRI modalities from existing ones using the BraTS'18 dataset has exhibited high performance, validated through PSNR and SSIM metrics. This demonstrates GANs' capability to reconstruct or substitute missing modalities, thereby improving diagnostic accuracy in cases with incomplete imaging data. The "Make-A-Volume" method has effectively integrated 2D and 3D approaches to produce high-quality volumetric medical images, outperforming comparable models in quantitative and qualitative evaluations. This hybrid approach enhances 3D visualization, offering clinicians more comprehensive and accurate representation of anatomical structures.

The COSMOS framework achieved state-of-the-art performance in the Cross-Modality Domain Adaption challenge by integrating a 2D CycleGAN with a 3D nnU-Net architecture. This highlights the effectiveness of integrating multiple network paradigms for domain adaptation in medical imaging. Similarly, the GAN model designed for generating sCT images for carbon ion radiotherapy has achieved high image accuracy and minimal range deviations in abdominal sCTs, confirming its suitability for precise radiotherapy dose planning.

To provide a clearer synthesis of these findings, a comparative table (Table 1) has been developed summarizing the imaging modality, anatomical regions, datasets, evaluation metrics, and key findings of the reviewed studies. This structured overview facilitates systematic comparison of methodologies and outcomes, highlighting both shared trends and unique contributions within the current research landscape.

4 Conclusion

The reviewed studies have highlighted significant advancements in cross-modality imaging through the application of CNN and GAN models. CNN-based models have been shown to be effective in tasks such as MRI to CT translation and the generation of synthetic images that preserve high structural fidelity. These developments have been supported by advances in multidetector technology and the introduction of refined architecture, including U-Net and its variants.

GAN-based approaches have demonstrated remarkable potential in generating missing modalities and producing high-quality synthetic data. Methods such as "Make-A-Volume" and frameworks such as COSMOS have further illustrated the adaptability and robustness of GANs in addressing complex challenges within medical imaging.

Table 1: Comparative Summary of this study

Study	**Model Framework**	**Anatomical Regions**	**Evaluation Metric**	**Key Findings**
Deep Image Prior (DIP) with U-Net (Cross-MRI Translation)	CNN	Brain MRI (structural regions, noise suppression)	RMSPE, MAPE, SSIM	Produced synthetic MRIs with high similarity to real images; effective at noise reduction.
Hip Morphology Study (MRI → CT (sCT))	CNN (BoneMRI, 3D U-Net-like)	Hip joint & surrounding bone morphology	Reader validation	sCT scans accurate for hip evaluation; reliable for diagnosing osteoarthritis.
Proton Therapy Study (MRI → CT (sCT))	CNN (sCT generator)	Abdominal & pelvic organs (proton dose planning)	Dosimetric differences	Minimal differences in treatment plans; feasible for clinical integration.
Brain Tumor Radiotherapy (MRI → CT (sCT))	3D CNN	Brain (tumor sites & surrounding tissues)	SSIM, clinical validation	Generated accurate sCTs for tumor planning; feasible for MRI-only radiotherapy.
Generating Missing Modalities (MRI → Missing modalities)	GAN	Brain (tumor regions, multimodal MRI synthesis)	PSNR, SSIM	High performance in generating missing modalities; improved multimodal imaging.
Make-A-Volume (MRI → 3D synthesis)	GAN (Hybrid 2D + 3D)	Whole brain & structural volumetric MRI synthesis	Quantitative + qualitative	Produced high-quality 3D images; outperformed other models.
COSMOS Framework (Cross-Modality Adaptation)	GAN (CycleGAN + nnU-Net)	Brain & general neuroimaging cross-modality tasks	Benchmark metrics	Achieved state-of-the-art results in cross-modality adaptation.

Despite these advancements, key challenges remain. The requirement for large, high-quality datasets continues to limit model training and generalizability, while the computational intensity of these techniques poses barriers to widespread implementation. Future research should therefore prioritize improving computational efficiency, developing methods that perform effectively with limited data, and validating existing models through extensive clinical trials to ensure their robustness and applicability in real-world healthcare environments.

The integration of artificial intelligence (AI) models into clinical workflows is anticipated to enhance diagnostic accuracy, streamline medical imaging processes, and improve overall patient outcomes. Continued interdisciplinary collaboration among computer scientists, radiologists, and clinicians remains crucial to achieving the safe and effective translation of these technologies into clinical practice.

Acknowledgement

The author would like to thank Nuzhi Meyen from Cogniata (Pvt) Ltd for his valuable assistance and insights during the research.

References

Bambach, S. and Ho, M.-L. (2022) 'Deep Learning for Synthetic CT from Bone MRI in the Head and Neck', *American Journal of Neuroradiology*, 43(8), p. 1172. doi: 10.3174/ajnr.A7588.

Florkow, M. C., Willemsen, K., Zijlstra, F., et al. (2022) 'MRI-based synthetic CT shows equivalence to conventional CT for the morphological assessment of the hip joint', *Journal of Orthopaedic Research*, 40(4), pp. 954–964. doi: 10.1002/jor.25127.

Lerner, M., Medin, J., Jamtheim Gustafsson, C., Alkner, S., Olsson, L. E. (2022) 'Prospective Clinical Feasibility Study for MRI-Only Brain Radiotherapy', *Frontiers in Oncology*, 11. doi: 10.3389/fonc.2021.812643.

McNaughton, J., Fernandez, J., Holdsworth, S., Chong, B., Shim, V., & Wang, A. (2023) 'Machine Learning for Medical Image Translation: A Systematic Review', *Bioengineering*, 10(9), p. 1078. doi: 10.3390/bioengineering10091078.

Menze, B.H. et al. (2015) 'The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)', *IEEE Transactions on Medical Imaging*, 34(10), p. 1993. doi: 10.1109/TMI.2014.2377694.

MR Guidance (n.d.) *MR Guidance: Quantitative MRI Solutions*. Available at: https://mriguidance.com/(Accessed: 29 May 2023)

Pal, S., Dutta, S. and Maitra, R. (2023) 'Personalized synthetic MR imaging with deep learning enhancements', *Magnetic Resonance in Medicine*, 89(4), pp. 1634–1643. doi: 10.1002/mrm.29527.

Parrella, G., Vai, A., Nakas, A., Garau, N., Meschini, G., Camagni, F., Molinelli, S., Barcellini, A., Pella, A., Ciocca, M., et al. (2023) 'Synthetic CT in Carbon Ion Radiotherapy of the Abdominal Site', *Bioengineering*, 10(2), p. 250. doi: 10.3390/bioengineering10020250.

Shin, H., Kim, H., Kim, S., Jun, Y., Eo, T., & Hwang, D.S. (2023) 'COSMOS: Cross-Modality Unsupervised Domain Adaptation for 3D Medical Image Segmentation based on Target-aware Domain Translation and Iterative Self-Training', *arXiv preprint*. Available at: http://arxiv.org/abs/2203.16557

Zimmermann, L., Knäusl, B., Stock, M., Lütgendorf-Caucig, C., Georg, D., Kuess, P. (2022) 'An MRI sequence independent convolutional neural network for synthetic head CT generation in proton therapy', *Zeitschrift für Medizinische Physik*, 32(2), pp. 218–227. doi: 10.1016/j.zemedi.2021.10.003.

Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., & Yu, L. (2023) 'Make-A-Volume: Leveraging Latent Diffusion Models for Cross-Modality 3D Brain MRI Synthesis', *arXiv preprint*. doi: 10.48550/arXiv.2307.10094.