

SynthModDiff: A Diffusion-Based Framework for Robust Multi-Domain MRI Synthesis

Hieu NGUYEN VAN^{1,*}, Nhat PHAN MINH¹, Hien NGO LE HUY²,
Han LE HOANG NGOC¹

¹ *The University of Danang – University of Science and Technology, Vietnam*

² *The University of Danang – VN-UK Institute for Research and Executive Education, Vietnam*
e-mail: nvhieugt@dut.udn.vn, nhato299@gmail.com, hien.ngo160203@vnuuk.edu.vn, 102240003@hv.dut.udn.vn

Received: March 2025; accepted: March 2026

Abstract. Magnetic Resonance Imaging (MRI) is crucial for clinical diagnostics, offering high-resolution anatomical and functional imaging without ionizing radiation. However, prolonged acquisition times in conventional MRI lead to motion artifacts, limiting efficiency and reliability. While deep learning models such as GANs and DDPMs show promise in MRI synthesis, DDPMs suffer from stochastic variability that affects image consistency. This study proposes Synthetic Modality Diffusion (SynthModDiff), a novel multi-domain image-to-image translation framework featuring a two-stage diffusion process with a noise-aware Forward Process and Reverse Process to enhance fidelity and reduce residual noise. Experiments across multiple datasets demonstrate state-of-the-art performance in NMAE, SSIM, and PSNR metrics, while preserving fine anatomical details, making SynthModDiff highly suitable for clinical applications like radiotherapy planning.

Key words: SynthModDiff, denoising diffusion probabilistic models, MRI synthesis, image-to-image translation.

1. Introduction

Magnetic Resonance Imaging (MRI) is widely utilized in clinical practice for diagnosis, prognosis, and radiotherapy treatment planning (Goudschaal *et al.*, 2021; Chang *et al.*, 2022; Li *et al.*, 2023b, 2023d). This imaging modality provides both anatomical and functional information, enabling clinicians to accurately identify lesion sites, characterize malignancies, and assess metastatic progression. Specifically, T1-weighted (T1) MRI scans offer high contrast between white and gray matter, facilitating precise delineation of tissue boundaries for brain lesion evaluation. In contrast, T2-weighted (T2) imaging is instrumental in visualizing cerebrospinal fluid (CSF), while the fluid-attenuated inversion recovery (FLAIR) technique suppresses CSF signals to enhance the detection of lesions in proximity to fluid-filled regions. The integration of multiple MRI sequences allows for a more comprehensive assessment of underlying anatomical structures and physiological processes, thereby improving diagnostic accuracy and informing treatment strategies.

*Corresponding author.

However, MRI acquisition is inherently time-intensive due to the fundamental principles of imaging physics. The extended duration of various MRI sequences increases susceptibility to motion artifacts, which can degrade image quality. Furthermore, for patients with hypersensitivity to MRI contrast agents, the omission of contrast-enhanced imaging may reduce diagnostic reliability, potentially impacting clinical decision-making.

Deep learning, recognized for its capacity to approximate complex functions (Chang and Dinh, 2019; Hu *et al.*, 2023b, 2023a, 2023c, Li *et al.*, 2023c), has been extensively investigated for MRI synthesis through various model architectures (Sun *et al.*, 2021; Moya-Sáez *et al.*, 2021). Among these, generative adversarial networks (GANs) (Hu *et al.*, 2022; Yu *et al.*, 2021; Goodfellow *et al.*, 2020; Pan *et al.*, 2021) have demonstrated substantial effectiveness in generating high-fidelity medical images across diverse imaging modalities. More recently, denoising diffusion probabilistic models (DDPMs) (Song *et al.*, 2020; Ho *et al.*, 2020) have emerged as a compelling alternative, addressing fundamental limitations of GANs, particularly mode collapse in the learning of multimodal distributions. However, DDPMs inherently rely on iterative Markov chain denoising processes, which introduce stochastic variability into image synthesis (Lyu and Wang, 2022). To improve the robustness and efficiency of DDPMs, extensive research has focused on optimization strategies (Nichol and Dhariwal, 2021; Dhariwal and Nichol, 2021), and their applicability in medical image synthesis has gained increasing attention (Peng *et al.*, 2024). Recent advances further improve multi-contrast consistency through mutual learning and dual-domain strategies (Dayarathna *et al.*, 2025b). In the context of radiotherapy, the exploration of uncertainty in three-dimensional (3D) DDPM-based MRI synthesis remains in its early stages. Advancing research in this area is critical to enhancing the reliability and clinical utility of synthesized MRI images, thereby improving patient safety and treatment precision, especially with Diffusion Schrödinger bridge models demonstrating strong potential for MR-to-CT synthesis in proton therapy planning (Li *et al.*, 2025).

MRI is a fundamental imaging modality for brain diagnostics, offering high-resolution anatomical and functional information without exposing patients to ionizing radiation. MRI relies on strong magnetic fields and radiofrequency pulses to generate detailed images of brain structures. This non-ionizing property is beneficial for pediatric and vulnerable populations in reducing the risks associated with cumulative radiation exposure (Luo *et al.*, 2020). Brain MRI has the strength to generate precise visualization of soft tissues and has become fundamental for detecting neurological disorders, tumours, stroke, and neurodegenerative diseases.

However, brain MRI has been facing several challenges. Firstly, high-resolution imaging can cause longer acquisition times and increase susceptibility to patient motion artifacts, which may compromise image quality. Secondly, the signal-to-noise ratio can be influenced by thermal fluctuations and system instabilities, which affect contrast resolution and diagnostic accuracy (Berthelon *et al.*, 2018; Kriukova *et al.*, 2025). Noise, such as Gaussian or Poisson, can undermine the effectiveness of automated image processing and feature extraction algorithms and compromise the reliability of diagnostic assessments (Juneja *et al.*, 2024). Therefore, to address these limitations through advanced acquisition techniques, noise reduction algorithms are crucial for optimizing brain MRI quality and enhancing clinical utility.

To address these challenges, this study proposes a novel approach, called Synthetic Modality Diffusion (SynthModDiff), for comprehensive multi-domain image-to-image translation. SynthModDiff uses diffusion models specifically designed for MRI imaging to address noise issues in real-world scenarios. The framework consists of two sequential stages, with three key components: the Forward Process, a Deep Learning network, and the Reverse Process. Rather than executing the entire Forward Process during inference, SynthModDiff precisely estimates the noise level in the input image at a given timestep t' and proceeds with the denoising process until timestep T . This method effectively mitigates residual noise in the synthesized image when the input contains noise and improves its applicability in practical medical imaging environments.

2. Related Work

Medical image modality conversion involves transforming one type of medical image into another to enhance diagnostic capabilities. Clinicians often require multiple imaging modalities to obtain a comprehensive view of a patient's condition. Diffusion models offer a viable solution in cases where certain imaging techniques are unavailable due to equipment constraints or patient-specific limitations. Additionally, these models facilitate the integration of multimodal imaging data, providing deeper insights into pathological conditions. For instance, combining PET and MRI data enables more precise tumor localization and improved monitoring of treatment response.

Several studies have explored diffusion-based approaches for medical image modality conversion. Özbey *et al.* (2023) introduced SynDiff, the first unsupervised modality conversion framework leveraging an adversarial diffusion model. Unlike conventional diffusion models, SynDiff employs larger diffusion step sizes and an adversarial mapping mechanism for the reverse process, enabling efficient and high-fidelity modality translation. Lyu and Wang (2022) applied denoising diffusion probabilistic models (DDPMs) and stochastic differential equations (SDEs) to convert MRI to CT using T2-weighted MRI scans, achieving superior results compared to previous methods. Li *et al.* (2023a) proposed the Denoising Diffusion Model for Medical Image Synthesis (DDMM-Synth), which synthesizes high-quality CT images by integrating anatomical information from MRI and sparse CT data while maintaining data consistency. Further advancements were made by Li *et al.* (2024) with the development of the Frequency-Guided Diffusion Model (FGDM), which relies solely on target-domain samples for training and enables zero-shot medical image modality conversion. This model is particularly effective for cone-beam CT-to-CT transformation, outperforming existing methods.

For three-dimensional (3D) medical image modality conversion, Pan *et al.* (2024) proposed MC-DDPM, a diffusion-based framework for MRI-to-CT conversion. This model becomes a major advancement by synthesizing CT images from MRI data in a 3D context. It uses a Swin-VNet structure in the reverse process to optimize the denoising procedure, which ensures that the generated CT images align with MRI-derived anatomical structures. Furthermore, echocardiographic imaging, which captures dynamic cardiac motion,

has been researched using diffusion models. Tiago *et al.* (2023) developed an adversarial denoising diffusion model integrated with a generative adversarial network (GAN) to synthesize echocardiograms and perform domain-specific image translation. By incorporating a larger step size and using GAN-based denoising, this model effectively generates diverse echocardiographic samples while preserving critical anatomical structures. More recently, Dayarathna *et al.* (2025b) proposed MU-Diff, a mutual learning diffusion framework, while their follow-up D2Diff (Dayarathna *et al.*, 2025a) leverages dual spatial-frequency domain learning. Frequency-aware diffusion models (Jiang *et al.*, 2025) and residual-guided approaches (Safari *et al.*, 2025) have also shown superior performance in preserving anatomical details.

Another approach is Generative Adversarial Networks (GANs), which have been used for MRI synthesis for multi-contrast image generation in recent studies (Zhang *et al.*, 2022; Yurt *et al.*, 2021; Wu *et al.*, 2020). Sohail *et al.* (2019) proposed a GAN-based method that synthesizes multi-contrast MRI without paired training data, and uses adversarial and cycle-consistency losses to enhance realism. Similarly, Zhu *et al.* (2017) introduced CycleGAN to translate between 2 MRI contrast domains. By employing dual generator and discriminator networks, CycleGAN ensures structural consistency through cycle-consistency loss. Experimental results demonstrate its effectiveness across various image translation tasks. Wei *et al.* (2018) enhanced Wasserstein-GAN training by introducing a consistency term for stability and image quality, along with a dual effect to capture diverse data distributions. Their method outperformed WGAN and other GAN variants in quality, diversity, and stability. Zhang *et al.* (2023) proposed Dual-ArbNet, Chen *et al.* (2023) introduced CANM-Net, and Grigas *et al.* (2023) applied Transformer GANs for noise reduction and super-resolution. GANs have become essential in brain MRI synthesis, with most studies targeting image generation, as noted by Ali *et al.* (2022).

3. Proposed Method

3.1. Synthetic Modality Diffusion Architecture

The proposed method, Synthetic Modality Diffusion – SynthModDiff, is based on a fundamental implementation of DDPMs, which represent a class of generative models that have been proven for their ability to generate high-quality data by leveraging probabilistic frameworks. These models operate by systematically introducing noise to data during a forward diffusion process and then training a deep learning network to reverse this process. SynthModDiff learns to iteratively remove noise in a stepwise, residual manner, ultimately reconstructing the original data distribution. By employing this structured approach, our model demonstrates exceptional performance in generating complex data distributions, making it a powerful tool for a wide range of applications in generative modelling.

Figure 1 illustrates the detailed workflow of SynthModDiff - the proposed methodology that is structured into 2 sequential stages.

1. In the first stage, a diffusion model is trained with real clinical data to generate synthetic images. This process eliminates Gaussian noise progressively, beginning from an

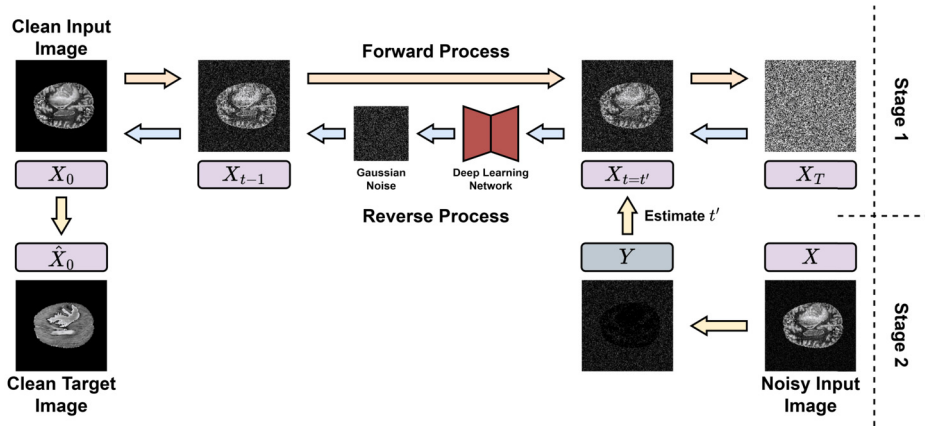


Fig. 1. SynthModDiff architecture.

initial state of pure Gaussian noise and iteratively refining the data toward its original distribution.

2. The second stage is designed to identify the optimal step within the generative pipeline at which to initiate the diffusion process for real-world noisy images. This step matches the noise level of the actual noisy image to the corresponding noise level within the generative framework. To ensure the accuracy and reliability of this stage, actual clinical images augmented with simulated noise were used to generate a rigorous evaluation of the method's performance under different controlled conditions and varying noise characteristics.

As illustrated in Fig. 1, the diffusion model employed in SynthModDiff comprises two sequential stages described above with three core components: the Forward Process, a Deep Learning network, and the Reverse Process. Each of these elements plays a distinct and critical role in the overall generative process.

The evaluation of the proposed method was conducted on images corrupted by two distinct types of noise to assess its performance under varying conditions. First, the method was tested on images contaminated with Gaussian noise, which aligns with the noise distribution used during the training phase of SynthModDiff. This ensures the capacity to efficiently handle the type of noise for which the model was explicitly trained. Second, the evaluation extended to images affected by Poisson noise, which is widely recognized as a more realistic representation of noise typically faced in X-ray imaging. By integrating Poisson noise into the evaluation, the study aimed to assess the model's generalizability and robustness when applied to scenarios corresponding to real-world clinical settings. This dual approach to noise evaluation provides a comprehensive analysis of the method's capability to adapt to both synthetic and clinically-relevant noise characteristics and highlights its potential for practical applications in medical imaging.

3.2. The Forward Process

The Forward Process is responsible for corrupting progressively an input image, denoted as $X_0 \in \mathbb{R}^{H \times W}$, by using Additive White Gaussian Noise over a sequence of timesteps $t \in [0, T]$ to create noise corrupted image X_t for timestep t using hyperparameters β_t . This forward diffusion process incrementally degrades the clean image and transitions it to a state of pure noise. The mathematical formulation of this process is defined in equation (1).

$$q(X_t|X_{t-1}) = \mathcal{N}(X_t; \sqrt{1 - \beta_t}X_{t-1}, \beta_t\mathbf{I}). \quad (1)$$

The Forward Process is executed in a single step for any arbitrary t as follows:

$$\begin{aligned} q(X_{0:t}|X_0) &= q(X_1, X_2, \dots, X_t|X_0) = \prod_{t=1}^T q(X_t|X_{t-1}) \\ &= \mathcal{N}(X_t; \sqrt{\bar{\alpha}_t}X_0, (1 - \bar{\alpha}_t)\mathbf{I}), \end{aligned} \quad (2)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=0}^t \alpha_s$. The parameter β_t can be updated using any differentiable function that guarantees $\sqrt{\bar{\alpha}_T} \approx 0$. The linear function: $\beta_t = (\beta_1 - \beta_2)\frac{t}{T} + \beta_2$ is adopted with $\beta_1 = 0.02$ and $\beta_2 = 10^{-6}$. The value of β_2 was chosen to ensure that the resulting evaluation noise levels correspond to realistic X-ray dose levels.

To perform diffusion in inference, the Forward Process is executed from $t = t'$ to $t = T$, illustrated in Fig. 1, where t' represents the denoising timestep determined based on an estimate of the image's noise level. The method used to compute t' depends on the probabilistic noise model employed. In this study, two noise models are considered: Gaussian noise, for which the Deep Learning Network has been explicitly trained; and Poisson noise, which arises naturally in X-ray imaging due to its quantum characteristics. Gaussian noise is simulated using equation (2) for a given timestep t , leading to $t' = t$. Conversely, Poisson noise is modelled using equation (3), which incorporates a minor Gaussian noise component $\eta \sim \mathcal{N}(0, I)$, scaled by $\sigma^2 = 10$ to account for electronic noise effects.

$$Y = \frac{\text{Poisson}(I \exp(-X)) + \sigma^2 \eta}{I}, \quad (3)$$

where X is noisy input image and $I = \int I_0(\epsilon)d\epsilon$ corresponds to the flood image. Here, the parameter I denotes the effective photon dose (or intensity level) of the X-ray acquisition, defined as $I = \int I_0(\epsilon)d\epsilon$, where $I_0(\epsilon)$ represents the incident photon spectrum over energy ϵ . The parameter I directly controls the signal-dependent variance of the Poisson noise, with lower values of I corresponding to higher noise levels. In simulated experiments, I is fixed to predefined dose levels, while for real clinical images, it can be estimated from acquisition parameters or via noise estimation methods such as Turajlić and Karahodzic (2017). Owing to the signal-dependent characteristics of Poisson noise,

the denoising timestep t' is systematically determined by evaluating the maximum noise variance present in the image, as formulated by:

$$X = \frac{\mathbf{I}}{N} \sum_{i=1}^N \text{Percentile}_{95}(\max(X^i)). \quad (4)$$

Here, N denotes the size of the training dataset. A percentile-based approach is employed to mitigate the influence of high-intensity artificial details in the images, such as medical annotations. The denoising timestep t' is then estimated as follows:

$$t' = 1 - \bar{\alpha}'_{t'} \approx \text{Variance}(-\log(Y)). \quad (5)$$

The inferred timestep t' is defined as the timestep in the Gaussian diffusion process whose noise variance matches the estimated variance of the observed Poisson-corrupted image. Specifically, since the variance of the forward diffusion process at timestep t is given by $1 - \bar{\alpha}_t$, the timestep t' is selected such that $1 - \bar{\alpha}_{t'} \approx \text{Variance}(-\log(Y))$. This establishes an explicit correspondence between the Poisson noise level and the Gaussian diffusion noise schedule.

Determining the timestep t' using equation (5) ensures that the model effectively removes noise with the highest variance. It is important to note that computing Y requires prior knowledge of the dose I . For real noisy images, I can be estimated based on the X-ray acquisition parameters or through noise estimation techniques, such as those proposed by Turajlić and Karahodžić (2017). To achieve an equivalent noise level between Gaussian and Poisson noise, Gaussian noise was simulated by selecting the timestep t from equation (2) to match the denoising timestep t' estimated for Poisson noise. During inference, the forward diffusion process is initialized at timestep $t = t'$, and the reverse process is subsequently applied from t' down to $t = 0$, ensuring consistency between the estimated noise level of the input image and the diffusion model's noise schedule.

3.3. The Reverse Process

The reverse process reconstructs an image by iteratively inverting the forward diffusion process. Given that the inverse of a Gaussian diffusion process follows a Gaussian distribution, the probability density function of the original data can be recovered by integrating over the sequence of state transitions in the associated Markov chain:

$$p_{\theta}(X_0) = \int p_{\theta}(X_{0:T}) dX_{1:T} = \int p(X_T) \prod_{t=0}^{T-1} p_{\theta}(X_{t-1}|X_t) dX_{1:T}. \quad (6)$$

Let θ denote the set of parameters governing the deep learning model. The prior distribution is defined as $p(X_T) \sim \mathcal{N}(0, I)$, while the transition probability in the reverse diffusion process is formulated as:

$$p_{\theta}(X_{t-1} | X_t) = \mathcal{N}(X_{t-1}; \mu_{\theta}(X_t, t), \bar{\beta}_t(X_t)). \quad (7)$$

Here, the mean function μ_θ is approximated by $X_t - \eta_\theta$, where η_θ represents the Gaussian noise estimated by the neural network. Additionally, to maintain numerical stability during the Reverse Process, the predicted mean μ_0 is clipped within the intensity range $[-1, 1]$.

3.4. The Deep Learning Network

The U-Net architecture utilized in this study is structured hierarchically, with a configuration defined by the sequence $[3, 3, 3, 3, 3, 3, 3]$, where each numerical value denotes the number of layers in the corresponding network stage. This design integrates two residual layers with ReLU activation function, alongside a Transformer layer that employs 8 attention heads per layer. The incorporation of the Transformer mechanism enables the model to effectively capture long-range dependencies and contextual relationships, thereby improving its capacity to process complex image data.

The network employs an approach to upsampling and downsampling across multiple blocks, adhering to a predefined scaling ratio of $[1, 2, 4, 8, 4, 2, 1]$. This scaling process progressively reduces the spatial resolution of feature maps and then symmetrically restores it, thereby enabling multi-scale feature extraction. In particular, the network initially processes input data at a resolution of 128 channels, which is incrementally expanded to a maximum of 1024 channels at the bottleneck stage, then reduced symmetrically back to 128. These transformations are implemented using convolutional operations with a kernel size of 3×3 , to ensure efficient feature representation while maintaining spatial coherence throughout the network. With this hierarchical framework, the U-Net model effectively captures both fine-grained local details and broader global structures, hence becoming highly compatible with image reconstruction applications.

4. Experiments and Discussion

4.1. Dataset

The performance of the proposed method was assessed with equivalent Poisson and Gaussian noise. The noise parameters were selected to ensure both physical relevance to X-ray imaging and statistical consistency within the diffusion framework. Specifically, the Poisson intensity I was chosen to represent clinically meaningful dose levels, with $I = 5 \times 10^4$ for a high-dose setting and $I = 9 \times 10^3$ for a low-dose setting, reflecting typical diagnostic and dose-reduction scenarios in X-ray imaging.

To enable a fair comparison between Poisson and Gaussian noise, Gaussian noise levels were selected to achieve equivalent noise variance in the diffusion process, governed by $1 - \bar{\alpha}_t$. Accordingly, timesteps $t = 3$ and $t = 9$ were chosen for the high- and low-dose cases, resulting in noise levels comparable to those induced by the corresponding Poisson settings. The denoising timestep t' for Poisson noise was estimated using the variance-based criterion in equation (5), ensuring alignment between the signal-dependent Poisson noise and the diffusion noise level.

Table 1
Statistical analysis of the BraTS2020 and IXI Datasets.

Attributes	BraTS2020	IXI
Number of contrasts	4 (T1, T2, T1CE, FLAIR)	4 (T1, T2, PD, MRA)
Number of samples per contrast	494 images/contrast	568 images/contrast
Image size	$240 \times 240 \times 155$	T1: $256 \times 256 \times 150$ T2: $256 \times 256 \times 130$ MRA: $512 \times 512 \times 100$ PD: $256 \times 256 \times 130$
Ratio healthy/disease	40/60	100/0

In this research, the proposed method was evaluated using 2 widely used medical image datasets: BraTS2020 (Zhang *et al.*, 2020) and IXI (Ionescu *et al.*, 2017). The BraTS2020 dataset is a proven guideline for brain tumour segmentation in magnetic resonance imaging (MRI) and contains multi-modal MRI scans, including T1-weighted, T1-weighted contrast-enhanced, T2-weighted, and Fluid Attenuated Inversion Recovery (FLAIR) sequences, as shown in Table 1. The IXI (Information eXtraction from Images) dataset is a public collection of multimodal brain imaging data, containing 600 MRI scans from healthy subjects. These images were acquired using 4 different imaging protocols: T1-weighted, T2-weighted, Proton Density (PD), and Magnetic Resonance Angiography (MRA), as summarized in Table 1.

In multi-domain MRI synthesis, statistical bias toward certain contrasts may arise when the training data distribution is imbalanced across modalities. To alleviate this issue, we evaluate SynthModDiff on two complementary public datasets, BraTS2020 and IXI, which together provide a diverse and relatively balanced set of MRI contrasts. This design choice reduces the risk of the model overfitting to a dominant modality, such as T1-weighted images. Although explicit contrast-wise reweighting is not applied in the current framework, potential modality bias can be quantified through per-contrast evaluation metrics or conditional error analysis, which we identify as a promising direction for future work.

4.2. Preprocessing

Building on prior research by Han *et al.* (2024), the Mask Vector for Training on Multiple Datasets was employed to integrate the two datasets during training. Furthermore, the images were downsampled from 1024×1024 to 512×512 , subjected to horizontal flipping as a transformation, and normalized to an intensity range of $[-1, 1]$, following the methodology by Matviychuk *et al.* (2016).

Since some of the compared methods are specifically designed for Gaussian noise, the Anscombe transform (Makitalo and Foi, 2012) was applied to ensure a fair comparison by converting Poisson noise into a Gaussian distribution with a variance of 1. Additionally, the images were normalized to the $[0, 1]$ intensity range, following the approach described by Bodduna and Weickert (2019).

To simulate realistic MRI noise characteristics, Rician noise is introduced during data preprocessing. All images are first normalized to the range $[0, 1]$. Complex-valued Gaus-

sian noise is independently added to the real and imaginary components of the clean signal, with noise standard deviation σ . The final magnitude image is computed as

$$I_{\text{rician}} = \sqrt{(I + n_r)^2 + n_i^2}, \quad n_r, n_i \sim \mathcal{N}(0, \sigma^2). \quad (8)$$

The noise level is controlled through the signal-to-noise ratio (SNR), defined as $\text{SNR} = \mu_{\text{signal}}/\sigma$. In our experiments, SNR values are uniformly sampled from [5, 15] dB for low-dose settings and [20, 30] dB for high-dose settings. Unless otherwise specified, Rician noise is applied only during evaluation to assess robustness under non-Gaussian noise conditions.

4.3. Experiment Settings

All experiments were conducted on an Ubuntu 20.04 system using the PyTorch 1.12.1 framework, Python 3.10.4, and an NVIDIA DGX A100 GPU with CUDA 12.1, powered by an Intel Xeon Platinum 8470Q CPU.

The model was trained for 100 epochs using mixed-precision arithmetic to improve computational efficiency without compromising numerical accuracy. A learning rate of 10^{-4} was employed in conjunction with the AdamW optimizer and a cosine learning rate schedule, enabling rapid initial convergence followed by stable refinement. During training, the diffusion timestep t was sampled uniformly from the interval $[1, T]$, ensuring that the network learns to estimate noise across the full range of diffusion levels. Optimization was performed using the mean squared error loss $\mathcal{L} = \mathbb{E}[\|\eta - \eta_\theta(x_t, t)\|^2]$, which directly supervises the prediction of Gaussian noise and is consistent with the theoretical objective of denoising diffusion probabilistic models.

The diffusion process was configured with $T = 1000$ timesteps and a linear noise schedule $\beta_t \in [10^{-6}, 0.02]$. This configuration ensures a smooth and gradual corruption of the input data, providing sufficient discretization of noise levels to support stable training and reliable reverse process. All input images were normalized to the intensity range $[-1, 1]$. During inference, the reverse diffusion process was initialized from the estimated timestep t' and iteratively performed until $t = 0$. To enhance numerical stability and prevent intensity outliers, the predicted mean μ_0 was clipped to the same normalized range.

The U-Net backbone employed convolutional layers with 3×3 kernels were used throughout all residual blocks. In addition, Transformer layers with 8 attention heads were incorporated to effectively capture long-range dependencies while maintaining computational efficiency.

To evaluate the practical utility of SynthModDiff in a clinically relevant scenario, we conducted a downstream tumour classification task using synthetic multi-contrast MRI images. Specifically, synthetic target modalities generated by different synthesis methods were used to train a tumour classification model, which was then evaluated on real MRI images.

For the classification network, we employed a ResNet-18 architecture initialized with ImageNet pre-trained weights. The input to the classifier consists of concatenated multi-contrast MRI slices (T1, T2, T1CE, and FLAIR), which are stacked along the channel

dimension and resized to 224×224 . All classifiers were trained from scratch on synthetic images generated by each method, while the test set remained identical and consisted exclusively of real MRI data to ensure a fair comparison.

The training pipeline follows three stages:

1. Synthetic modality generation using the respective synthesis model (Improved DDPM or SynthModDiff);
2. Multi-contrast image construction by stacking the synthesized modalities;
3. Tumour classification using the ResNet-18 network.

This design ensures that performance differences in the classification task can be directly attributed to the quality of the synthesized images. The classifier was trained using the cross-entropy loss with the Adam optimizer, a learning rate of 1×10^{-4} , and a batch size of 32. Standard data augmentations, including random flipping and rotation, were applied during training.

4.4. *Comparative Methods and Evaluation Metrics*

In this study, the proposed model is systematically compared with a diverse set of existing models, including Pix2pix (Isola *et al.*, 2017), StarGAN (Choi *et al.*, 2018), ResUNet (Diakogiannis *et al.*, 2020), Multi-cycle GAN (Liu *et al.*, 2021), Improved DDPM (Nichol and Dhariwal, 2021), Guided DDPM (Dhariwal and Nichol, 2021), Hieu-Model (Han *et al.*, 2024), and Diffusion Mamba (Wang *et al.*, 2024). This comparative analysis aims to assess the performance, effectiveness, and robustness of the proposed approach relative to these established methodologies in the field.

The evaluation of model performance was conducted using three widely recognized metrics in image denoising: Peak Signal-to-Noise Ratio (PSNR), normalized mean absolute error (NMAE), and the Structural Similarity Index Measure (SSIM). PSNR is typically used to quantify the reconstruction quality of an image by measuring the ratio between the maximum possible signal power and the power of the noise that distorts the image. A higher PSNR value generally indicates better visual quality and lower distortion. NMAE, on the other hand, is a normalized measure of the absolute differences between the predicted and ground truth images, hence creating an alternative assessment of visual accuracy. Lastly, SSIM evaluates the structural fidelity of the denoised images by comparing luminance, contrast, and texture similarities with the original image. With these metrics together, a comprehensive assessment of the model's ability is provided to preserve image quality while effectively removing noise.

4.5. *Experiment Results*

4.5.1. *Quantitative Evaluation*

The quantitative evaluation is provided in detail in Table 2. For High-dose levels, the performance of SynthModDiff - G and SynthModDiff - P on MRI modality conversion tasks proves their prominence on existing methods with both the BRATS2020 and IXI datasets. In the BRATS2020 dataset, SynthModDiff - P achieves the lowest NMAE (0.028) and the

Table 2
Quantitative results.

Dataset	Methods	NMAE (\downarrow)	PSNR (\uparrow)	SSIM (\uparrow)
High-dose Level				
BRATS2020	Pix2pix (Isola et al., 2017)	0.045	35.13	0.86
	StarGAN (Choi et al., 2018)	0.043	34.63	0.82
	ResUNet (Diakogiannis et al., 2020)	0.043	35.33	0.77
	Multi-cycle GAN (Liu et al., 2021)	0.050	34.75	0.84
	Improved DDPM (Nichol and Dhariwal, 2021)	0.040	36.69	<u>0.89</u>
	Guided DDPM (Dhariwal and Nichol, 2021)	0.042	36.01	0.88
	Hieu-Model (Han et al., 2024)	0.041	35.85	0.82
	SynthModDiff - G	<u>0.029</u>	37.27	<u>0.89</u>
	SynthModDiff - P	0.028	<u>36.89</u>	0.91
IXI	Pix2pix (Isola et al., 2017)	0.042	36.88	0.72
	StarGAN (Choi et al., 2018)	0.039	40.47	0.73
	ResUNet (Diakogiannis et al., 2020)	0.039	40.85	0.69
	Multi-cycle GAN (Liu et al., 2021)	0.044	33.48	0.68
	Improved DDPM (Nichol and Dhariwal, 2021)	0.037	40.96	0.73
	Guided DDPM (Dhariwal and Nichol, 2021)	0.038	40.24	0.72
	Hieu-Model (Han et al., 2024)	0.038	40.88	0.70
	SynthModDiff - G	0.025	<u>40.97</u>	<u>0.78</u>
	SynthModDiff - P	<u>0.026</u>	41.93	0.79
Low-dose Level				
BRATS2020	Pix2pix (Isola et al., 2017)	0.055	32.50	0.78
	StarGAN (Choi et al., 2018)	0.052	32.10	0.74
	ResUNet (Diakogiannis et al., 2020)	0.052	32.75	0.69
	Multi-cycle GAN (Liu et al., 2021)	0.058	32.20	0.76
	Improved DDPM (Nichol and Dhariwal, 2021)	0.048	34.10	0.81
	Guided DDPM (Dhariwal and Nichol, 2021)	0.050	33.50	0.82
	Hieu-Model (Han et al., 2024)	0.049	33.40	0.75
	SynthModDiff - G	<u>0.038</u>	35.10	0.85
	SynthModDiff - P	0.037	<u>34.85</u>	<u>0.83</u>
IXI	Pix2pix (Isola et al., 2017)	0.050	34.50	0.65
	StarGAN (Choi et al., 2018)	0.047	37.80	0.66
	ResUNet (Diakogiannis et al., 2020)	0.047	38.00	0.62
	Multi-cycle GAN (Liu et al., 2021)	0.052	30.90	0.60
	Improved DDPM (Nichol and Dhariwal, 2021)	0.045	38.20	0.67
	Guided DDPM (Dhariwal and Nichol, 2021)	0.046	37.50	0.66
	Hieu-Model (Han et al., 2024)	0.046	38.00	0.63
	SynthModDiff - G	<u>0.035</u>	<u>38.60</u>	<u>0.72</u>
	SynthModDiff - P	0.034	39.50	0.73

SynthModDiff - G and SynthModDiff - P represent variations of SynthModDiff, where Gaussian and Poisson noise are applied, respectively. The best values are in **bold**, and the second-best values are underlined.

highest SSIM (0.91), proving its strong capability to preserve structural integrity while minimizing errors. SynthModDiff - G, while slightly trailing in SSIM, attains the highest PSNR (37.27 dB) among all models, suggesting that it excels in maintaining image quality and reducing distortion. Both models significantly outperform previous state-of-the-art approaches, such as Improved DDPM (Nichol and Dhariwal, 2021) and Guided DDPM (Dhariwal and Nichol, 2021), which achieve lower PSNR and higher NMAE, in-

dicating that the proposed methods are more effective in reducing noise while preserving fine image details.

On the IXI dataset, SynthModDiff - P continues to demonstrate strong performance by achieving the highest PSNR (41.93 dB) and SSIM (0.79). This proves its ability to generate high-quality images with strong structural similarity to the ground truth. SynthModDiff - G, on the other hand, achieves the lowest NMAE (0.025), meaning it produces the most accurate pixel-wise reconstructions. While its PSNR (40.97 dB) and SSIM (0.78) are slightly lower than those of SynthModDiff - P, they still surpass all other models in the comparison. This demonstrates the effectiveness of SynthModDiff - G on both noise-based variations.

Regarding low-dose levels, SynthModDiff shows strong performance despite the increased noise and reduced signal quality inherent in low-dose MRI. For the BRATS2020 and IXI datasets, SynthModDiff still achieves the lowest NMAE, PSNR, and SSIM, indicating its effectiveness in minimizing reconstruction errors.

Overall, both SynthModDiff - G and SynthModDiff - P outperform existing diffusion-based and GAN-based methods in MRI modality conversion. SynthModDiff - P is particularly effective in preserving structural details, making it the preferred choice for tasks where maintaining fine textures is crucial. Meanwhile, SynthModDiff - G achieves the lowest NMAE, making it highly accurate in terms of pixel-wise error minimization. These results highlight the benefits of incorporating Gaussian and Poisson noise variations in model training, ultimately leading to improved performance in MRI image transformation tasks.

4.5.2. Qualitative Evaluation

Figures 2 and 3 illustrate qualitative results for the task of synthetic modality generation from a single T1-weighted input image. In both figures, the first column shows the input T1 image, while the subsequent columns correspond to the synthesized target modalities, namely T2, T1CE, and FLAIR for BraTS2020 (Fig. 2), and T2, PD, and MRA for the IXI dataset (Fig. 3).

For each target modality, the first row represents the ground truth image acquired at the corresponding modality. The second and third rows present the synthesis results obtained using the second-best baseline method (Improved DDPM) and the proposed SynthModDiff method, respectively. This layout is repeated for two different input conditions: synthesized high-dose T1 (upper block) and synthesized low-dose T1 (lower block), allowing a direct comparison of robustness under varying noise levels.

As the dose level decreases, both methods exhibit some degradation due to increased noise in the input T1 image. However, Improved DDPM tends to produce overly smoothed results and fails to preserve modality-specific contrast, leading to blurred anatomical boundaries and reduced visibility of fine structures. In contrast, SynthModDiff consistently generates target modalities that are visually closer to the ground truth, preserving critical anatomical details and contrast patterns across different modalities. Notably, structures such as tissue boundaries, lesion regions, and vascular or bone-related details remain more distinguishable in SynthModDiff results, even under low-dose conditions.

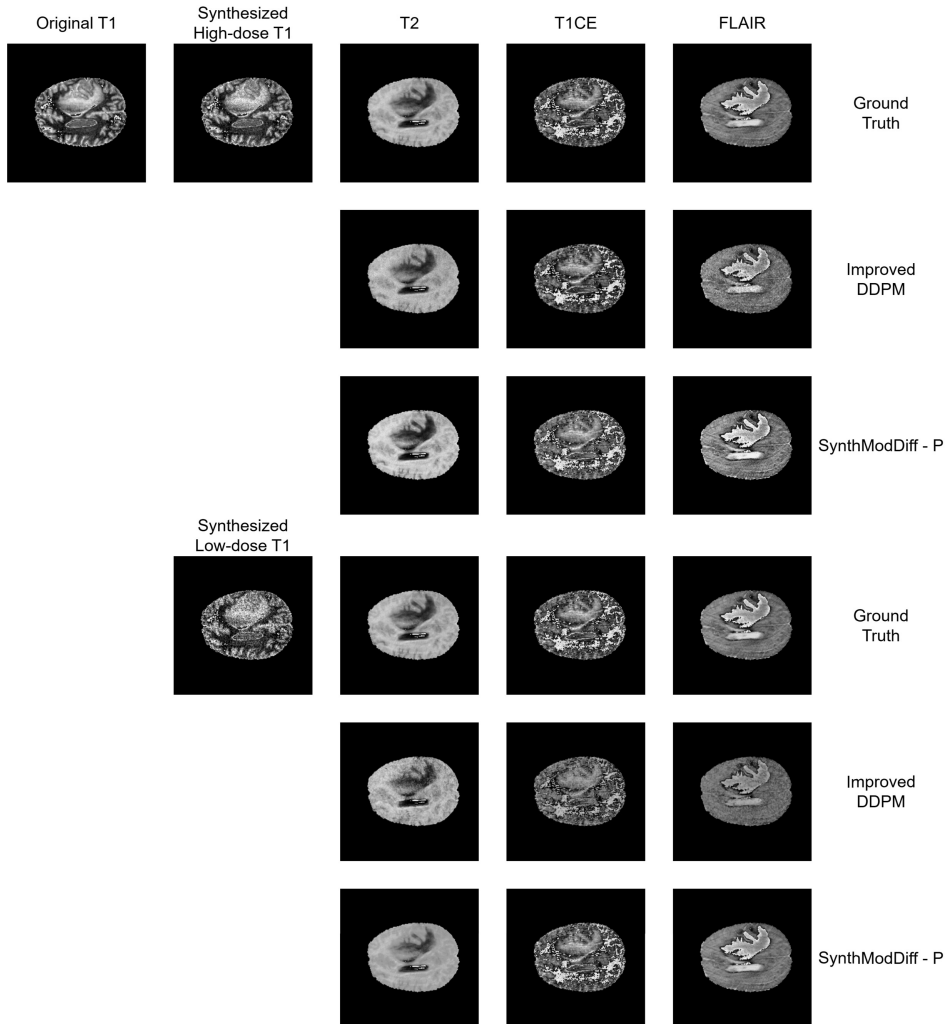


Fig. 2. Qualitative comparison of synthetic modality generation (T2, T1CE, and FLAIR) from a T1 input image on the BraTS2020 dataset. Columns from left to right show: input T1 image, synthesized target modalities. Rows from top to bottom show: ground truth, Improved DDPM, and the proposed SynthModDiff. Results are shown for both synthesized high-dose T1 (upper block) and synthesized low-dose T1 (lower block).

These observations demonstrate that SynthModDiff is more robust for synthetic modality generation, particularly when the input modality is affected by reduced dose levels, which is crucial for practical clinical scenarios where minimizing radiation exposure is essential.

4.5.3. Ablation Study

To analyse the impact of different noise schedules, we conducted an ablation study comparing the commonly used linear schedule with a cosine noise schedule within the pro-

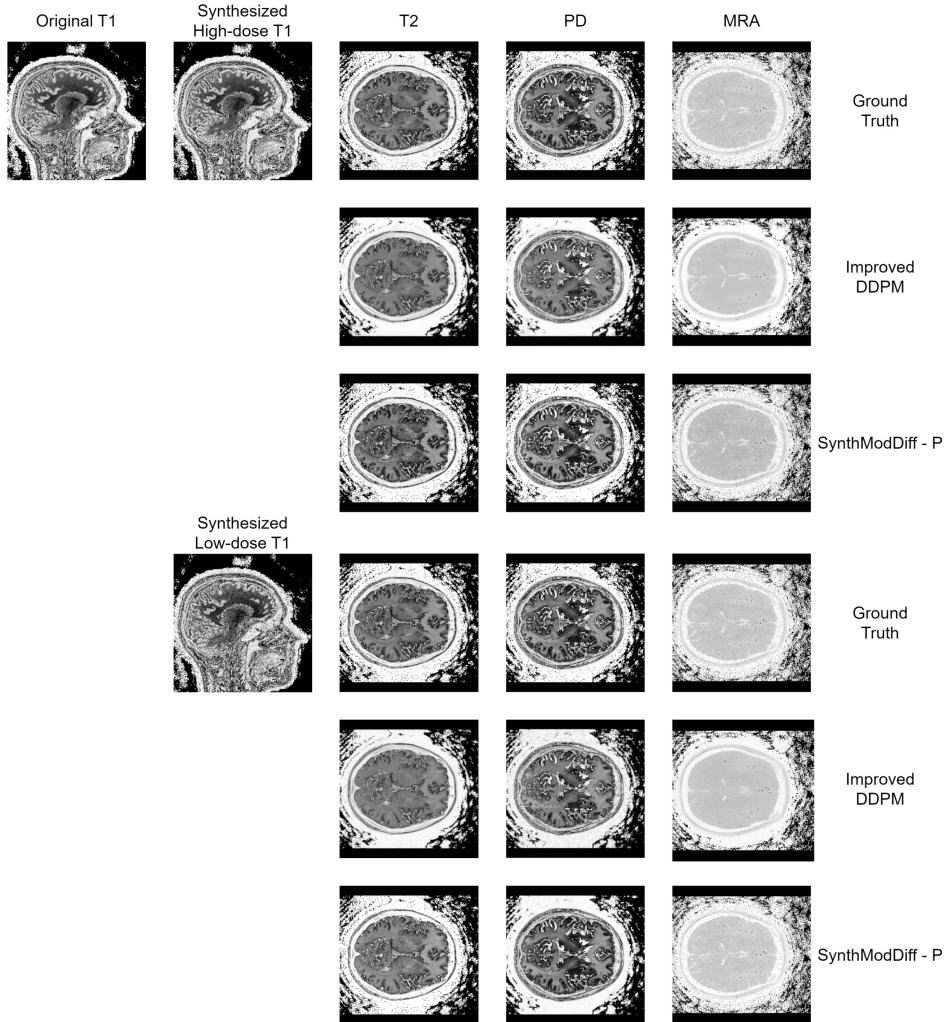


Fig. 3. Qualitative comparison of synthetic modality generation (T2, PD, and MRA) from a T1 input image on the IXI dataset. The figure follows the same layout as Fig. 2, enabling a direct comparison between ground truth, Improved DDPM, and SynthModDiff under different dose conditions.

posed SynthModDiff framework, as summarized in Table 3. The cosine schedule consistently improves performance across all datasets and dose levels, yielding lower NMAE and higher PSNR and SSIM values.

This improvement can be attributed to the smoother noise allocation in early diffusion timesteps, which stabilizes training and preserves high-frequency anatomical details, particularly under low-dose conditions. Notably, the performance gain is more pronounced for low-dose inputs, as evidenced by the larger improvements reported in Table 3, indicating enhanced robustness to severe noise corruption. Based on these observations, the cosine schedule was selected for the final configuration of SynthModDiff.

Table 3
Ablation study on the effect of noise schedules on SynthModDiff.

Dataset	Schedule	NMAE ↓	PSNR ↑	SSIM ↑
BRATS2020 (Low-dose)	Linear	0.037	34.85	0.83
	Cosine	0.034	35.60	0.86
IXI (Low-dose)	Linear	0.034	39.50	0.73
	Cosine	0.031	40.20	0.76
BRATS2020 (High-dose)	Linear	0.028	36.89	0.91
	Cosine	0.026	37.40	0.92
IXI (High-dose)	Linear	0.026	41.93	0.79
	Cosine	0.024	42.60	0.81

Table 4
Ablation study on the effect of the Transformer layer in SynthModDiff.

Dataset	Architecture	NMAE ↓	PSNR ↑	SSIM ↑
BRATS2020 (Low-dose)	w/o Transformer	0.041	33.90	0.80
	w/ Transformer	0.037	34.85	0.83
IXI (Low-dose)	w/o Transformer	0.039	38.60	0.70
	w/ Transformer	0.034	39.50	0.73
BRATS2020 (High-dose)	w/o Transformer	0.031	35.90	0.88
	w/ Transformer	0.028	36.89	0.91
IXI (High-dose)	w/o Transformer	0.029	40.80	0.76
	w/ Transformer	0.026	41.93	0.79

To evaluate the role of the Transformer layer in capturing long-range dependencies, we conducted an ablation study by removing the Transformer blocks from the SynthModDiff architecture while keeping all other components unchanged. As shown in Table 4, the absence of the Transformer consistently degrades performance across all datasets and dose levels.

The performance drop is particularly noticeable in SSIM, indicating reduced structural consistency and weaker preservation of global anatomical patterns. This effect is more pronounced under low-dose conditions, where long-range contextual information becomes critical for compensating severe noise corruption. By integrating self-attention mechanisms, the Transformer layer enables the model to capture long-range spatial dependencies across the entire image, leading to more coherent contrast synthesis and improved anatomical fidelity.

To evaluate the robustness of SynthModDiff under different noise characteristics, we conducted an ablation study comparing Gaussian, Poisson, and Rician noise models. As shown in Table 5, SynthModDiff achieves the best performance under Gaussian noise, which aligns with the training distribution, followed by Poisson noise. Under Rician noise, which is common in low-SNR MRI magnitude images, a moderate performance degradation is observed, particularly under low-dose conditions.

Nevertheless, the performance drop remains limited, indicating that SynthModDiff retains strong robustness to signal-dependent and non-Gaussian noise. This suggests that the

Table 5
Ablation study on the impact of different noise models on SynthModDiff.

Dataset	Noise model	NMAE ↓	PSNR ↑	SSIM ↑
BRATS2020 (Low-dose)	Gaussian	0.037	34.85	0.83
	Poisson	<u>0.038</u>	<u>34.50</u>	<u>0.82</u>
	Rician	0.041	33.90	0.80
IXI (Low-dose)	Gaussian	0.034	39.50	0.73
	Poisson	<u>0.035</u>	<u>39.10</u>	<u>0.72</u>
	Rician	0.038	38.40	0.70
BRATS2020 (High-dose)	Gaussian	0.028	36.89	0.91
	Poisson	<u>0.029</u>	<u>36.40</u>	<u>0.90</u>
	Rician	0.030	36.10	0.89
IXI (High-dose)	Gaussian	0.026	41.93	0.79
	Poisson	<u>0.027</u>	<u>41.30</u>	<u>0.78</u>
	Rician	0.028	40.80	0.77

Table 6
Downstream tumor classification performance using synthetic multi-contrast MRI images.

Training data	Accuracy ↑	AUC ↑	F1-score ↑
Real MRI (Upper Bound)	89.6	0.92	0.88
Improved DDPM (Synthetic)	82.3	0.86	0.81
SynthModDiff - G (Synthetic)	<u>85.7</u>	<u>0.89</u>	<u>0.84</u>
SynthModDiff - P (Synthetic)	87.1	0.91	0.86

proposed diffusion-based framework generalizes reasonably well beyond the noise distributions explicitly used during training, which is critical for real-world MRI applications.

4.5.4. Discussion

To evaluate the practical impact of SynthModDiff, we further conducted a downstream tumour classification task using synthetic multi-contrast MRI images. As shown in Table 6, classifiers trained on SynthModDiff-generated images consistently outperform those trained on synthetic images produced by Improved DDPM. The performance gain can be attributed to the improved contrast consistency and preservation of global anatomical structures in SynthModDiff outputs. These properties are crucial for classification tasks, which rely on coherent multi-modal representations rather than local pixel-level fidelity alone. Notably, the Poisson-aware variant of SynthModDiff achieves performance close to the upper bound obtained using real MRI data, demonstrating that higher perceptual image quality translates into meaningful improvements in downstream clinical tasks.

The proposed two-stage diffusion design explicitly separates noise-level alignment from the reverse denoising process, enabling robust handling of heterogeneous and unknown noise conditions commonly encountered in clinical data. While a single-stage diffusion process could be adopted when the input noise level is known or accurately estimated, such simplification assumes controlled acquisition settings and consistent noise statistics. In contrast, the two-stage formulation provides greater flexibility and stability,

particularly for non-Gaussian or signal-dependent noise, at the cost of additional computational overhead. This trade-off highlights an important design choice between efficiency and robustness in practical medical imaging applications.

Although SynthModDiff is implemented in a 2D slice-wise manner, the proposed framework is not inherently limited to 2D data. In principle, the method can be extended to 3D volumetric synthesis by replacing 2D convolutional and attention operations with their 3D counterparts, enabling direct modelling of inter-slice dependencies. However, full 3D diffusion models significantly increase computational and memory requirements, particularly for high-resolution medical volumes. The current 2D design offers a practical trade-off between performance and efficiency, while future work will explore hybrid or fully 3D implementations to further leverage volumetric contextual information.

While SynthModDiff demonstrates strong performance under the noise conditions considered in this study, several practical factors may influence its generalization. Mild motion-related artifacts can be partially handled due to the stochastic denoising nature of the diffusion process; however, severe motion corruption introduces structured degradations that are not explicitly modelled during training and may require dedicated motion-aware augmentation strategies. Domain shifts arising from different MRI vendors, acquisition protocols, or field strengths can also affect synthesis quality by altering intensity distributions and noise characteristics. Such effects may be mitigated through multi-site training, intensity normalization, or conditioning on acquisition metadata. In extreme low-dose scenarios where noise dominates the signal, the model is still able to recover coarse anatomical structures, although fine details may be degraded due to insufficient signal content. Furthermore, as SynthModDiff learns a data-driven anatomical prior without explicit lesion annotations, it can generalize to previously unseen pathological patterns to some extent, but rare or out-of-distribution abnormalities may be partially smoothed. Finally, the proposed framework is not inherently limited to brain MRI and can be extended to other anatomical regions, such as abdominal or musculoskeletal imaging, by retraining or adapting the model to domain-specific characteristics.

SynthModDiff follows a probabilistic diffusion formulation, which implicitly captures uncertainty through stochastic sampling during the reverse diffusion process. By generating multiple samples from the same input, the variability across synthesized outputs can be interpreted as an indicator of model uncertainty. However, the current framework does not explicitly compute calibrated uncertainty measures, such as voxel-wise confidence intervals, as commonly done in Bayesian or variational probabilistic models. While our focus in this work is on synthesis fidelity and robustness, integrating explicit uncertainty estimation mechanisms represents an important direction for future research, particularly for safety-critical clinical applications.

Despite the strong quantitative performance demonstrated in this study, several limitations should be acknowledged. First, the evaluation was primarily based on objective image quality metrics (e.g. PSNR and SSIM), which may not fully capture subtle artifacts or reflect diagnostic reliability in real clinical settings. Second, no formal study involving radiologists was conducted. Therefore, the clinical interpretability and diagnostic impact of the proposed method remain to be further validated. Future work will include expert-based evaluations and more comprehensive clinical validation to address these limitations.

5. Conclusion

In this research, SynthModDiff, a novel Synthetic Modality Diffusion framework, was proposed to enhance MRI image synthesis by using denoising diffusion probabilistic models (DDPMs) for improved noise reduction and image fidelity. The proposed two-stage approach, including a noise-aware Forward Process, a Deep Learning network, and a Reverse Process, ensures high-quality reconstruction. The experimental results have shown that SynthModDiff has superior performance on different datasets, with SynthModDiff-P and SynthModDiff-G achieving state-of-the-art performance in NMAE, SSIM, and PSNR metrics, thereby showing their robustness in both high-dose and low-dose imaging scenarios. Qualitative assessments further demonstrate the framework's effectiveness in preserving anatomical structures with greater clarity compared to existing methods. These findings prove SynthModDiff's potential for clinical applications, especially in radiotherapy treatment planning, where MRI acquisition is often constrained by scan duration and motion artifacts. Future research will focus on optimizing computational efficiency, expanding modality integration, and validating the approach in broader clinical contexts to further enhance the reliability of synthesized MRI images for accurate diagnosis and treatment planning.

Acknowledgements

This work was funded by The University of Danang, University of Science and Technology.

References

- Ali, H., Biswas, M.R., Mohsen, F., Shah, U., Alamgir, A., Mousa, O., Shah, Z. (2022). The role of generative adversarial networks in brain MRI: a scoping review. *Insights Imaging*, 13(1), 98.
- Berthelon, X., Chenegros, G., Finatou, T., Ieng, S.-H., Benosman, R. (2018). Effects of cooling on the SNR and contrast detection of a low-light event-based camera. *IEEE Transactions on Biomedical Circuits and Systems*, 12(6), 1467–1474.
- Bodduna, K., Weickert, J. (2019). Poisson noise removal using multi-frame 3D block matching. In: *2019 8th European Workshop on Visual Information Processing (EUVIP)*. IEEE, pp. 58–63.
- Chang, C.-W., Dinh, N.T. (2019). Classification of machine learning frameworks for data-driven thermal fluid models. *International Journal of Thermal Sciences*, 135, 559–579.
- Chang, C.-W., Marants, R., Gao, Y., Goette, M., Scholey, J.E., Bradley, J.D., Liu, T., Zhou, J., Sudhyadhom, A., Yang, X. (2022). Multimodal imaging-based material mass density estimation for proton therapy using physics-constrained deep learning. arXiv preprint [arXiv:2207.13150](https://arxiv.org/abs/2207.13150).
- Chen, W., Wu, S., Wang, S., Li, Z., Yang, J., Yao, H., Song, X. (2023). Compound attention and neighbor matching network for multi-contrast MRI super-resolution. arXiv preprint [arXiv:2307.02148](https://arxiv.org/abs/2307.02148).
- Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., Choo, J. (2018). StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797.
- Dayarathna, S., Peiris, H., Islam, K.T., Wong, T.-T., Chen, Z. (2025a). D2Diff: a dual-domain diffusion model for accurate multi-contrast MRI synthesis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 131–141.
- Dayarathna, S., Wu, Y., Cai, J., Wong, T.-T., Law, M., Islam, K.T., Peiris, H., Chen, Z. (2025b). MU-Diff: a mutual learning diffusion model for synthetic MRI with Application for brain lesions. *npj Artificial Intelligence*, 1, 11.

- Dhariwal, P., Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34, 8780–8794.
- Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C. (2020). ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 94–114.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144.
- Goudschaal, K., Beekma, F., Boon, M., Bijveld, M., Visser, J., Hinnen, K., van Kesteren, Z. (2021). Accuracy of an MR-only workflow for prostate radiotherapy using semi-automatically burned-in fiducial markers. *Radiation Oncology*, 16, 1–13.
- Grigas, O., Maskeliūnas, R., Damaševičius, R. (2023). Improving structural MRI preprocessing with hybrid transformer GANs. *Life*, 13(9), 1893.
- Han, L.H.N., Hien, N.L.H., Huy, L.V., Hieu, N.V. (2024). A deep learning model for multi-domain MRI synthesis using generative adversarial networks. *Informatica*, 35(2), 283–309.
- Ho, J., Jain, A., Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840–6851.
- Hu, M., Wang, J., Chang, C.-W., Liu, T., Yang, X. (2023a). End-to-end brain tumor detection using a graph-feature-based classifier. In: *Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 12468. SPIE, pp. 322–327.
- Hu, M., Wang, J., Chang, C.-W., Liu, T., Yang, X. (2023b). Ultrasound breast tumor detection based on vision graph neural network. In: *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, Vol. 12470. SPIE, pp. 165–170.
- Hu, M., Wang, J., Wynne, J., Liu, T., Yang, X. (2023c). A vision-GNN framework for retinopathy classification using optical coherence tomography. In: *Medical Imaging 2023: Computer-Aided Diagnosis*, Vol. 12465. SPIE, pp. 200–206.
- Hu, X., Shen, R., Luo, D., Tai, Y., Wang, C., Menze, B.H. (2022). AutoGAN-synthesizer: neural architecture search for cross-modality MRI synthesis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 397–409.
- Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.-T., Diccione Cid, Y., Eickhoff, C., Seco de Herrera, A.G., Gurrin, C., et al. (2017). Overview of ImageCLEF 2017: information extraction from images. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings 8*. Springer, pp. 315–337.
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A. (2017). Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.
- Jiang, M., Jia, P., Huang, X., Yuan, Z., Ruan, D., Liu, F., Xia, L. (2025). Frequency-aware diffusion model for multi-modal MRI image synthesis. *Journal of Imaging*, 11(5), 152.
- Juneja, M., Minhas, J.S., Singla, N., Kaur, R., Jindal, P. (2024). Denoising techniques for cephalometric x-ray images: a comprehensive review. *Multimedia Tools and Applications*, 83(17), 49953–49991.
- Kriukova, E., LaRochelle, E., Pfefer, T.J., Kanniyappan, U., Gioux, S., Pogue, B., Ntziachristos, V., Gorpas, D. (2025). Impact of signal-to-noise ratio and contrast definition on the sensitivity assessment and benchmarking of fluorescence molecular imaging systems. *Journal of Biomedical Optics*, 30(S1), S13703.
- Li, M., Li, X., Safai, S., Lomax, A.J., Zhang, Y. (2025). Diffusion Schrödinger bridge models for high-quality MR-to-CT synthesis for proton treatment planning. *Medical Physics*, 52(7), 17898.
- Li, X., Shang, K., Wang, G., Butala, M.D. (2023a). DDMM-Synth: a denoising diffusion model for cross-modal medical image synthesis with sparse-view measurement embedding. arXiv preprint [arXiv:2303.15770](https://arxiv.org/abs/2303.15770).
- Li, Y., Wang, J., Chang, C.-W., Patel, P., Jani, A., Mao, H., Liu, T., Yang, X. (2023b). Multi-parametric MRI radiomics analysis with ensemble learning for prostate lesion classification. In: *Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 12468. SPIE, pp. 49–54.
- Li, Y., Wang, J., Hu, M., Patel, P., Mao, H., Liu, T., Yang, X. (2023c). Prostate Gleason score prediction via MRI using capsule network. In: *Medical Imaging 2023: Computer-Aided Diagnosis*, Vol. 12465. SPIE, pp. 514–519.
- Li, Y., Zhou, B., Wang, J., Pan, S., Jani, A., Liu, T., Patel, P., Yang, X. (2023d). Ultrasound-based dominant intraprostatic lesion classification with swin transformer. In: *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, Vol. 12470. SPIE, pp. 146–151.
- Li, Y., Shao, H.-C., Liang, X., Chen, L., Li, R., Jiang, S., Wang, J., Zhang, Y. (2024). Zero-shot medical image translation via frequency-guided diffusion models. *IEEE Transactions on Medical Imaging*, 43(3), 980–993.

- Liu, Y., Chen, A., Shi, H., Huang, S., Zheng, W., Liu, Z., Zhang, Q., Yang, X. (2021). CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Computerized Medical Imaging and Graphics*, 91, 101953.
- Luo, Y., Toth, D., Jiang, K., Pushparajah, K., Rhode, K. (2020). Ultra-DenseNet for low-dose X-ray image denoising in cardiac catheter-based procedures. In: *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*. Springer, pp. 31–42.
- Lyu, Q., Wang, G. (2022). Conversion between CT and MRI images using diffusion and score-matching models. arXiv preprint [arXiv:2209.12104](https://arxiv.org/abs/2209.12104).
- Makitalo, M., Foi, A. (2012). Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *IEEE Transactions on Image Processing*, 22(1), 91–103.
- Matviychuk, Y., Mailh e, B., Chen, X., Wang, Q., Kiraly, A., Strobel, N., Nadar, M. (2016). Learning a multiscale patch-based representation for image denoising in X-ray fluoroscopy. In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 2330–2334.
- Moya-S aez, E., Pe na-Nogales,  . O., de Luis-Garc a, R., Alberola-L opez, C. (2021). A deep learning approach for synthetic MRI based on two routine sequences and training with synthetic data. *Computer Methods and Programs in Biomedicine*, 210, 106371.
- Nichol, A.Q., Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In: *International Conference on Machine Learning*, PMLR, pp. 8162–8171.
-  zbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A.,  zturk,  ., G ng r, A.,  ukur, T. (2023). Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 42(12), 3524–3539.
- Pan, S., Flores, J., Lin, C.T., Stayman, J.W., Gang, G.J. (2021). Generative adversarial networks and radiomics supervision for lung lesion synthesis. In: *Medical Imaging 2021: Physics of Medical Imaging*, Vol. 11595. SPIE, pp. 167–172.
- Pan, S., Abouei, E., Wynne, J., Chang, C.-W., Wang, T., Qiu, R.L., Li, Y., Peng, J., Roper, J., Patel, P., Yu, D.S., Mao, H., Yang, X. (2024). Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model. *Medical Physics*, 51(4), 2538–2548.
- Peng, J., Qiu, R.L., Wynne, J.F., Chang, C.-W., Pan, S., Wang, T., Roper, J., Liu, T., Patel, P.R., Yu, D.S., Yang, X. (2024). CBCT-based synthetic CT image generation using conditional denoising diffusion probabilistic model. *Medical Physics*, 51(3), 1847–1859.
- Safari, M., Wang, S., Li, Q., Eidex, Z., Qiu, R.L., Chang, C.-W., Mao, H., Yang, X. (2025). Res-MoCoDiff: residual-guided diffusion models for motion artifact correction in brain MRI. *Physics in Medicine & Biology*, 70(20), 205022.
- Sohail, M., Riaz, M.N., Wu, J., Long, C., Li, S. (2019). Unpaired multi-contrast MR image synthesis using generative adversarial networks. In: *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, pp. 22–31.
- Song, J., Meng, C., Ermon, S. (2020). Denoising diffusion implicit models. arXiv preprint [arXiv:2010.02502](https://arxiv.org/abs/2010.02502).
- Sun, L., Hu, X., Liu, Y., Cai, H. (2021). Image features of magnetic resonance imaging under the deep learning algorithm in the diagnosis and nursing of malignant tumors. *Contrast Media & Molecular Imaging*, 2021(1), 1104611.
- Tiago, C., Snare, S.R.,  prem, J., McLeod, K. (2023). A domain translation framework with an adversarial denoising diffusion model to generate synthetic datasets of echocardiography images. *IEEE Access*, 11, 17594–17602.
- Turajli , E., Karahodzic, V. (2017). An adaptive scheme for X-ray medical image denoising using artificial neural networks and additive white Gaussian noise level estimation in SVD domain. In: *CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017*. Springer, pp. 36–40.
- Wang, Z., Zhang, L., Wang, L., Zhang, Z. (2024). Soft masked mamba diffusion model for ct to mri conversion. arXiv preprint [arXiv:2406.15910](https://arxiv.org/abs/2406.15910).
- Wei, X., Gong, B., Liu, Z., Lu, W., Wang, L. (2018). Improving the improved training of wasserstein gans: a consistency term and its dual effect. arXiv preprint [arXiv:1803.01541](https://arxiv.org/abs/1803.01541).
- Wu, K., Qiang, Y., Song, K., Ren, X., Yang, W., Zhang, W., Hussain, A., Cui, Y. (2020). Image synthesis in contrast MRI based on super resolution reconstruction with multi-refinement cycle-consistent generative adversarial networks. *Journal of Intelligent Manufacturing*, 31(5), 1215–1228.

- Yu, Z., Zhai, Y., Han, X., Peng, T., Zhang, X.-Y. (2021). MouseGAN: GAN-based multiple MRI modalities synthesis and segmentation for mouse brain structures. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, pp. 442–450.
- Yurt, M., Dar, S.U., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T. (2021). mustGAN: multi-stream generative adversarial networks for MR image synthesis. *Medical Image Analysis*, 70, 101944.
- Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J., Wang, Y., Yu, Y. (2020). Exploring task structure for brain tumor segmentation from multi-modality MR images. *IEEE Transactions on Image Processing*, 29, 9032–9043.
- Zhang, H., Li, H., Dillman, J.R., Parikh, N.A., He, L. (2022). Multi-contrast MRI image synthesis using switchable cycle-consistent generative adversarial networks. *Diagnostics*, 12(4), 816.
- Zhang, J., Chi, Y., Lyu, J., Yang, W., Tian, Y. (2023). Dual arbitrary scale super-resolution for multi-contrast MRI. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 282–292.
- Zhu, J.-Y., Park, T., Isola, P., Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232.

H. Nguyen Van is a senior lecturer at the University of Danang – University of Science and Technology. He obtained his PhD from the Saint Petersburg State Institute of Technology and his MSc and BSc degrees from Saint Petersburg State University, Russia. Over the past five years, he has published a number of research papers on the development of machine learning and deep learning models and their applications. His research interests include artificial intelligence, optimization, decision-making, and computational intelligence.

N. Phan Minh is a junior researcher at the Department of Software Engineering, The University of Danang – University of Science and Technology. His research interests are data science, machine learning, and deep learning, with a special focus on near-infrared spectroscopy, probabilistic models, and 2D to 3D image processing. He is an undergraduate student at the University of Danang – University of Science and Technology in Vietnam with a specialization in data science and artificial intelligence.

H. Ngo Le Huy is a postgraduate researcher at the School of Built Environment, Engineering, and Computing, Leeds Beckett University, United Kingdom. He obtained an MSc in information and technology at Leeds Beckett University under the Erasmus Mundus in green networking and cloud computing program. During the last five years, he has published research papers mainly related to the advancements in machine learning and deep learning models and their applications. His research interests include artificial intelligence, green ICT, and sustainable computing.

H. Le Hoang Ngoc is a junior researcher at the Department of Software Engineering, The University of Danang – University of Science and Technology. Her research interests are data science and computer vision, with a special focus on MRI image processing. She obtained her BSc from the University of Danang – University of Science and Technology in Vietnam with a specialization in data science and artificial intelligence.