A Comparative Study on Diffusion Sampling Methods Across Diverse Medical Imaging Modalities

Muhammad Ali Farooq[1], Ayman Abaid[2], Ihsan Ullah[2,3], and Peter Corcoran[1]

[1] School of Engineering, University of Galway, H91TK33, Ireland
[2] School of Computer Science, University of Galway, H91TK33, Ireland
[3] Insight SFI Research Centre for Data Analytics, University of Galway, Ireland
{muhammadali.farooq,a.abaid1,ihsan.ullah,peter.corcoran}@universityofgalway.ie

Abstract. The evaluation of diffusion-based image sampling methods is pivotal in improving the quality and reliability of synthetic data generation, particularly in medical imaging applications. Medical imaging requires high precision and fidelity, as even subtle artifacts or inconsistencies can significantly impact clinical decision-making. This study examines the effectiveness of four different image sampling techniques across various medical imaging modalities, focusing on dermoscopic skin lesion data, computed tomography angiography for Type B Aortic Dissection, and chest X-ray imaging. By systematically assessing these methods, we aim to enhance the fidelity of synthetic datasets, ensuring they more closely resemble real-world clinical data thereby supporting more accurate diagnostics, treatment planning, and prognostic predictions. In this work, we evaluate the performance of four different sampling techniques by incorporating Euler, Euler A, Denoising Diffusion Implicit Mode (DDIM), and Pseudolinear Multistep (PLMS) approaches for medical image synthesis. The study utilizes quantitative metrics including Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) to assess the realism and structural integrity of the generated images. Additionally, we employed t-SNE visualization to illustrate the latent feature representations of rendered synthetic medical images, providing an intuitive understanding of the underlying structure. We also analyzed and compared the computational complexity associated with each image sampling technique, offering insights into the efficiency of different approaches. The generated medical images are available at Diffusion-Sampling-for-Medical-Image-Synthesis.

Keywords: Synthesis · Stable Diffusion · CTA · Dermoscopic · X-ray · Text to Image

1 Introduction

Medical imaging plays a critical role in modern healthcare, providing clinicians with invaluable insights into the structure, function, and pathology of the human body. However, the medical data privacy concern and further acquisition of

high-quality medical imaging data often relies on expensive and time-consuming processes, limiting the availability of large datasets for training machine learning models. Additionally, the diversity and complexity of medical imaging data present unique challenges for traditional machine learning approaches. To address these challenges, recent advancements in artificial intelligence (AI) have shown promise in generating synthetic medical imaging data using generative models. In recent years, there has been a growing interest in the application of image diffusion techniques across different modalities, including X-ray, magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and positron emission tomography (PET), among others [1, 4, 7, 8].

This work focuses on the synthesis of three specific medical imaging modalities: computed tomography angiography (CTA), X-ray, and dermoscopic images. The first application concentrates on synthesizing dermoscopic images to enhance the diagnosis of malignant skin cancer. Dermoscopic imaging is particularly vital in dermatology, as it provides critical insights into the microscopic structures and pigmentation patterns of skin lesions. This research aims to generate synthetic dermoscopic images that closely resemble real-world examples, thereby facilitating the development and validation of machine learning algorithms specifically designed for skin cancer diagnosis. The second application involves the rendering of synthetic cardiac computed tomography angiography (CTA) images to aid in the diagnosis and prognosis of Type B Aortic Dissection (TBAD). TBAD is a life-threatening condition characterized by a tear in the inner layer of the aorta, necessitating prompt diagnosis and intervention. The generation of synthetic CTA images aims to enhance diagnostic capabilities for this critical condition, ultimately improving patient outcomes. The third application focuses on generating synthetic chest X-ray images for pediatric patients to assist in the diagnosis of pneumonia. Given the challenges associated with obtaining sufficient training data for machine learning models in pediatric populations, the synthesis of these images is crucial for developing effective diagnostic tools.

This study aims to provide a comprehensive exploration of adapting image diffusion modeling in medical imaging [16], encompassing its underlying principles, methodologies, applications, and future directions. While substantial work has been done in generating medical imaging data using diffusion models, this research seeks to evaluate how sampling methodologies impact each modality.

2 Medical Image Synthesis using Diffusion Modeling

This section explores the stable diffusion modeling techniques and integration of image sampling methods, for synthesizing high-quality medical images, enabling the generation of realistic and clinically relevant datasets.

2.1 Text-to-Image Stable Diffusion Model

Text-to-image synthesis in medical imaging refers to the process of generating realistic medical images from textual descriptions or medical reports, which

has significant potential for enhancing diagnostic accuracy and facilitating the training of AI-driven systems. Various generative models, including Variational Autoencoders (VAEs) [11], Generative Adversarial Networks (GANs) [15], and diffusion models [3], have been employed to achieve this task. Among these, diffusion-based models are particularly notable for their enhanced training stability and superior alignment with textual descriptions, while generating high-fidelity images with realistic textures and intricate anatomical details. These characteristics make diffusion models particularly well-suited for medical imaging, where the accurate portrayal of subtle anatomical structures is crucial for clinical utility. Diffusion's ability to maintain image quality with limited data availability further strengthens its applicability in scenarios where access to large annotated datasets is restricted, a common challenge in medical research.

Diffusion models, such as Stable Diffusion [12] and DALL-E [9, 10], have demonstrated remarkable capabilities in generating images based on textual prompts, having been trained on vast datasets containing billions of images. However, training a diffusion model from scratch poses significant challenges in the medical domain due to the limited availability of large, annotated datasets, primarily due to ethical and logistical constraints. This is where few-shot learning techniques become invaluable. Few-shot learning enables models to be trained with a minimal number of examples, making it particularly advantageous in scenarios where acquiring extensive annotated datasets is impractical. In this research, we propose an integrated approach that combines few-shot learning with Stable Diffusion to generate synthetic medical imaging data that accurately reflects real-world conditions. This augmentation of training data is essential for downstream tasks such as image segmentation, classification, and disease diagnosis.

By leveraging this synergistic approach, we aim to enhance the performance of AI-driven medical imaging models and improve diagnostic workflows, ultimately advancing the field of medical AI.

2.2 Sampling Techniques

Image diffusion modeling, as discussed in [2], is grounded in stochastic processes, where the goal is to propagate and diffuse information across image pixels or voxels in a structured and controlled manner. A key component of this methodology is the image samplers, which are responsible for generating new image samples that maintain a distribution closely aligned with the original seed data. The process begins by generating a random image within a latent space as demonstrated in Figure 1. A noise predictor, typically a U-Net architecture, is employed to estimate the noise present in the image. At each step of the process, the predicted noise is subtracted from the image, and this denoising procedure is iteratively repeated over multiple timesteps. Through this repeated refinement, the model gradually transforms the noisy image into a clean, coherent output. Formally, the process starts with the latent variable $\mathbf{x}_T$ sampled from a normal distribution $\mathcal{N}(0, \mathbf{I})$. At each timestep $T = t, t-1, \ldots, 1$, the next state $\mathbf{x}_{t-1}$ is computed as

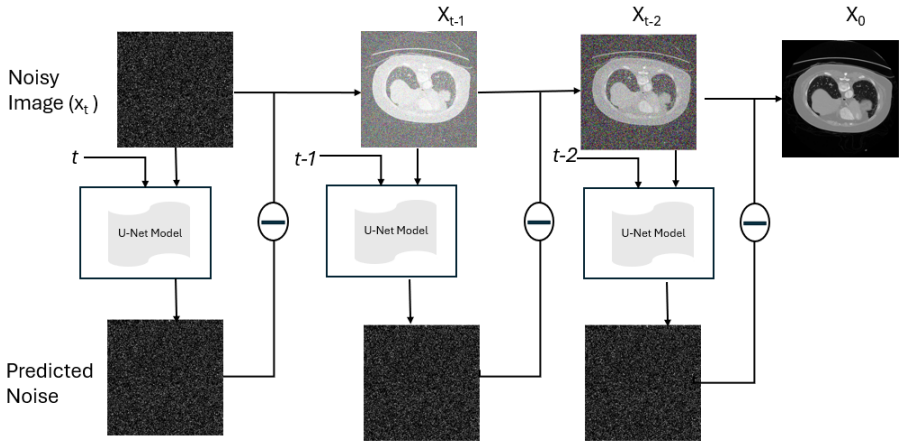

Fig. 1: Schematic representation of the reverse diffusion process using a U-Net in a Diffusion Model

follows:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \sqrt{\frac{1-\alpha_t}{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) + \sigma_t \mathbf{z}$$

where the noise term $\mathbf{z}$ is sampled from a standard normal distribution $\mathcal{N}(0, \mathbf{I})$ for $t > 1$, and is set to 0 when $t = 1$. Here, α_t and $\bar{\alpha}_t$ are predefined schedule parameters, typically derived from noise schedules such as linear or cosine schedules, while σ_t is the variance controlling the noise scale at timestep t. The function $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is a learned denoising model, usually implemented as a neural network. The denoising procedure iterates until $t = 1$, at which point the final denoised output $\mathbf{x}_0$ is obtained and returned. The below subsections provide details about different image sampling methods employed for rendering distinct medical modality data.

2.2.1 Euler A Euler A, where 'A' denotes Ancestral, represents an ancestral variant of the Euler sampling method, commonly employed in diffusion models. The term "Ancestral" indicates that this approach incorporates randomness from earlier steps within the sampling process, reintroducing noise that had previously been reduced or removed. This step-wise reintroduction of noise adds an element of unpredictability, as the process revisits earlier states while progressing. Notably, Euler A maintains a constant presence of noise throughout the process, which prevents the sampler from fully converging to a single, deterministic solution. Instead, this method introduces a degree of flexibility, enabling small variations in the results with each step, even in later stages of the sampling procedure. This stochastic nature makes Euler A a useful tool for exploring the solution space in a more comprehensive manner, ultimately fostering innovation through the generation of multiple, slightly varied outputs.

2.2.2 Euler Euler is a numerical sampling method used in diffusion-based generative models, particularly in Stable Diffusion and other generative models like Denoising Diffusion Probabilistic Models (DDPMs). These methods are originally derived from Euler's method, which is based on numerical method for differential equations. For critical applications like medical imaging, where anatomical accuracy is crucial, the choice between Euler and Euler A depends on the balance between speed and image fidelity. Euler is preferable for generating more controlled outputs, while Euler A might be useful when slight variations in images are beneficial for tasks like data augmentation.

2.2.3 Denoising Diffusion Implicit Mode The Denoising Diffusion Implicit Mode (DDIM) [14] method introduces a more efficient sampling process by using deterministic, implicit denoising steps. It leverages a non-Markovian forward process, meaning the noise added at each time-step depends not just on the previous step, but also on other steps in the process. This results in the same final data distribution with far fewer steps, making the process faster and more computationally efficient.

2.2.4 Pseudolinear Multistep Pseudolinear Multistep (PLMS) derives from numerical methods known as Linear Multistep Methods (LMS) [6]. It is a multistep method, such that it uses information from multiple previous time steps to predict the current image state. Instead of relying only on the current time-step's noise and gradient, PLMS takes advantage of the history of the reverse diffusion process, making it more efficient and accurate. By using a pseudolinear strategy, PLMS allows for faster convergence compared to single-step methods like Euler. PLMS reduces the number of timesteps needed to transform noise into a clear image, making it more computationally efficient without compromising quality.

3 Dataset

In this study, we acquired three different public medical imaging datasets as input data for fine-tuning diffusion models and defined short text prompts for each data class used during the training phase. The dermoscopic images were sourced from the International Skin Imaging Collaboration (ISIC) Archive, consisting of 2,800 images evenly distributed between benign and malignant lesions (1,400 images in each category). For the ImageTBAD dataset, which contains 100 annotated 3D computed tomography angiography (CTA) images labeled for true lumen (TL), false lumen (FL), and false lumen thrombosis (FLT), the 3D volumes were converted into 2D by treating each axial slice as an individual image. The dataset was divided into training and testing sets, with three primary classes: Class 1 for TL, Class 2 for FLT, and Class 3 containing both TL and FL. Additionally, a chest X-ray dataset consisting of anterior-posterior images from pediatric patients aged one to five years was employed, with images collected from Guangzhou Women and Children's Medical Center. After initial

quality control, the images were diagnostically validated by two board-certified physicians, and a third expert independently reviewed the evaluation set to minimize labeling errors. The detailed train-test split for each of these datasets is presented in Table 1. The train data samples were used for fine-tuning diffusion model whereas the test data samples were used for the quantitative evaluation and data visualization phase.

Type	Classes	Train	Test
Dermoscopic	**Benign**	1440	360
	Malignant	1197	300
CTA	**True Lumen**	501	235
	True Lumen + False Lumen	16616	3395
	False Lumen Thrombus	121	52
Chest X-Ray	**Normal**	1342	242
	Pneumonia	3876	398

Table 1: Overview of the dataset with the number of training and test samples for each class of three different medical imaging modalities.

4 Methodology for Synthesizing Dermoscopic, CTA TBAD, and Chest X-ray Imaging Data

This section will discuss the adapted methodology to synthesize high quality medical imaging data as a reliable source to enlarge the existing datasets and indulge more additional variety and diversity. In this context the recent success in text to image translation using pretrained stable diffusion models by employing language encoders such as CLIP model has gained enough popularity. By leveraging diffusion modelling in text-to-image translation can assist in tasks such as generating synthetic medical images for training machine learning algorithms, augmenting medical image datasets, or aiding in medical education and communication. Additionally, these models have the potential to facilitate interdisciplinary collaboration by enabling seamless communication between clinicians and imaging experts through textual descriptions of medical findings or diagnostic impressions translated into visual representations. Figure 2 shows the adapted methodology for tuning large-scale diffusion models using transfer learning, tailored for customized medical image rendering tasks. As it can be observed from Figure 2 that we have employed Dreambooth tool [13] for the purpose of transfer learning and fine-tuning the pretrained stable diffusion model. Further Low Rank Adaptation (LoRA) [5] is used as a lightweight training methodology to fine-tune Large Language and Stable Diffusion Models without necessitating complete model retraining. The conventional approach of fully fine-tuning larger models, characterized by billions of parameters, entails inherent resource intensiveness and temporal demands. LoRA operates by introducing a reduced

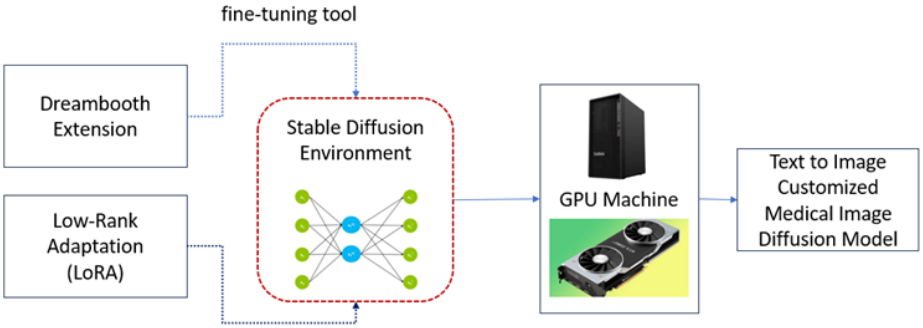

Fig. 2: Block diagram of the proposed methodology. DreamBooth [13] with LoRA (Low-Rank Adaptation) [5] utilized for fine-tuning pre-trained diffusion models in medical imaging, enabling the generation of highly specialized and domain-specific synthetic data, including dermoscopic, CTA and X-ray images, while maintaining high fidelity and anatomical accuracy.

set of new weights to the model during training, thereby bypassing the requirement to retrain the entirety of the model's parameter space. Consequently, this strategy significantly reduces the number of trainable parameters, leading to expedited training durations and more manageable file sizes, typically spanning a few hundred megabytes. The next stage includes passing on the trained models in the inference pipelines. The image inference pipeline in stable diffusion models comprises a series of sequential steps aimed at generating high-quality images from textual prompts/ descriptions. Figure 3 shows four sequential steps for image rendering using optimized Text to Image Diffusion models. The same pipeline has been adapted in our experimental work for generating synthetic dermoscopic, CTA, and X-ray medical imaging data. The further details on data sourcing and training the models via few shot learning methodology are available in our accepted papers [4], [1].

5 Experimental Results

The complete experimental work was carried on workstation machine equipped with A6000 graphic card with 48GB of graphical video memory. Further we have used Pytorch framework for code implementation.

5.1 Inference Results using text guided prompts with various Diffusion Sampling Methods

The first phase of experimental results focuses on image rendering results using four different sampling methods as discussed in subsection 2.2.1, 2.2.2, 2.2.3, and 2.2.4. All the images were generated in 512x512 image resolution and stored in lossless PNG format. Figure 4 shows the rendered benign and malignant

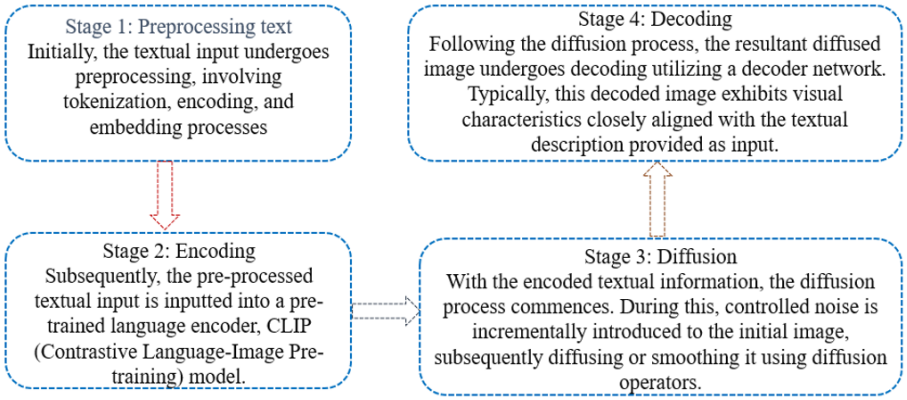

Fig. 3: Four stage pipeline for fine-tuning large scale pretrained stable diffusion model via transfer learning for customized medical imaging application.

dermoscopic data samples. Whereas Figure 5 and Figure 6 shows the synthetic CTA TBAD and chest X-ray imaging data generated via tuned diffusion pipeline elaborated in Figure 2. The complete inference results for each of the medical imaging modality is available at out GitHub repository Diffusion-Sampling-for-Medical-Image-Synthesis.

5.2 Performance Comparison of Sampling Methods for Medical Image Generation

In this section, we conduct a comprehensive evaluation of the impact of various sampling methods on different medical imaging modalities. Our assessment is based on two critical factors: (1) the quality of the generated data, and (2) computational complexity.

5.2.1 Image Quality Evaluation The visual samples from each imaging modality, generated using four distinct sampling techniques, are presented in Figures 4, 5, and 6. Notably, as seen in Figure 5, most sampling methods yielded robust and realistic results across modalities. However, PLMS sampler exhibited considerable variability, particularly in the case of CTA data, where random sampling noise led to inconsistent outputs. This inconsistency suggests that PLMS struggled to generate realistic CTA images, especially from a non-expert observer's perspective. The sampler's limitation can likely be attributed to the intricate anatomical details required in CTA data, which PLMS failed to accurately capture. This observation is further reinforced by the t-SNE visualizations in Figure 7, where synthetic CTA images generated by PLMS cluster significantly farther from their real counterparts, compared to the closer clustering of images generated by other samplers, which more closely align with the real data. A similar trend is noted with dermoscopic images, where malignant samples generated

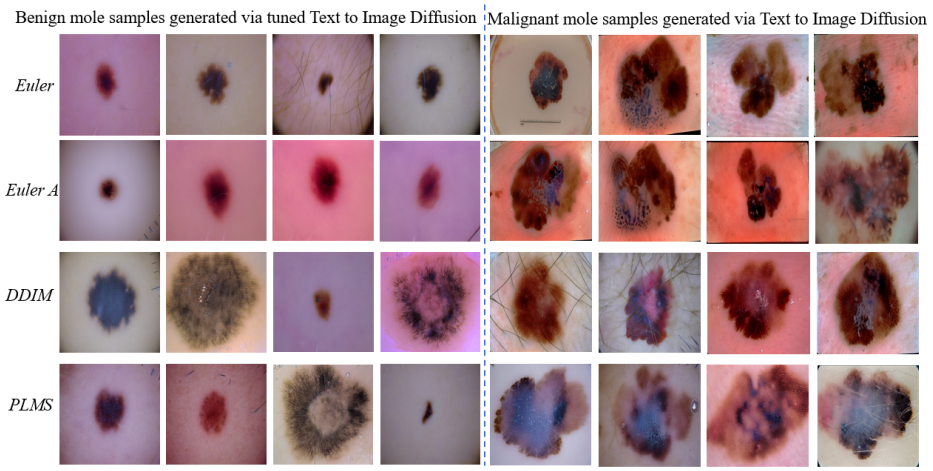

Fig. 4: Synthetic bening and malignant mole samples generated via Derm-T2IM model using four different sampling methods which includes Euler, Euler A, DDIM and PLMS.

by PLMS lacked diversity, producing similar structures repeatedly. Interestingly, despite these limitations, PLMS performed adequately when applied to X-ray data, generating outputs that were visually plausible to untrained observers. The t-SNE analysis further shows that synthetic dermoscopic and CTA images from most samplers clustered more closely with real images, with synthetic dermoscopic images from all samplers aligning more closely with their real counterparts compared to X-ray data. This can be attributed to the richer feature set provided by the three-channel RGB nature of dermoscopic images, which enhances the model's ability to generate more realistic outputs. In contrast, the grayscale and anatomically specialized features of CTA and X-ray images present greater challenges for realistic synthesis.

For the quantitative evaluation of generated images, we employed Learned Perceptual Image Patch Similarity (LPIPS) [17] and Structural Similarity Index Measure (SSIM) to assess the quality of images across different classes within each imaging modality. LPIPS evaluates perceptual similarity by focusing on how closely the generated images align with human visual quality perception, with lower LPIPS values indicating greater perceived similarity. In contrast, SSIM measures similarity based on structural information, luminance, and contrast, offering a more comprehensive assessment of perceived image quality than traditional pixel-wise comparisons. Higher SSIM values reflect better structural similarity, with values approaching 1 indicating high similarity and values closer to 0 indicating greater dissimilarity. As shown in Table 2, the lowest LPIPS score and highest SSIM score for benign dermoscopic images was achieved using the Euler A method. These results demonstrate that benign samples generated using the Euler A method exhibit the highest similarity to real data, both in terms of perceptual and structural quality. For the CTA modality, the PLMS method

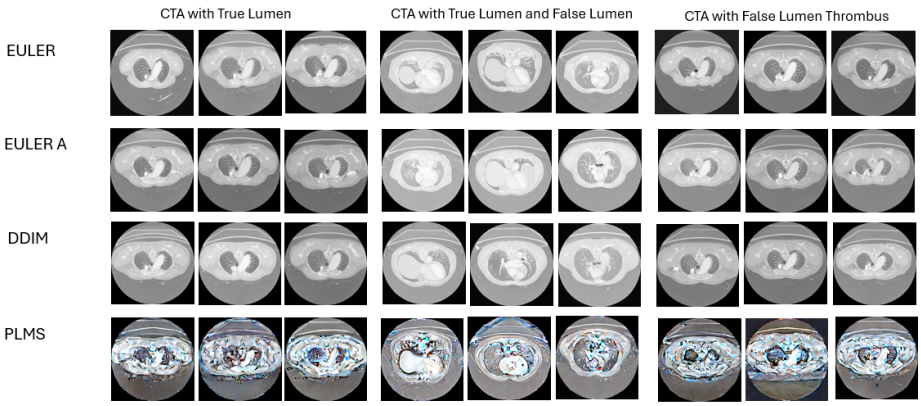

Fig. 5: Synthetic CTA images of Type B aortic dissection showing the true lumen in the first column, both the true lumen and false lumen in the second column, and the false lumen thrombus in the third column. Images were generated using four different sampling methods: Euler, Euler A, DDIM, and PLMS.

yielded the highest LPIPS score especially when rendering the TL class data, thus indicating the lowest perceptual similarity when compared to real-world CTA data. Similarly the lowest SSIM score was also achieved on all classes of CTA data using the PLMS method which means that rendered data samples have not enough similarity with real world CTA data samples when generated via PLMS sampling method.

5.2.2 Computational Efficiency and Resource Usage We assess the computational complexity of diffusion-based sampling methods by evaluating their time and space complexities and analyzing the trade-offs among various samplers. Our focus is on key parameters such as the number of sampling steps, Classifier-Free Guidance (CFG) scale, and inference time. The CFG scale modulates the influence of conditional information (e.g., a text prompt) on the model; higher CFG values place greater emphasis on the prompt, while lower values introduce more diversity and randomness. For our experiments, we set the CFG scale to 7 and investigated the effects of increasing sampling steps. In this evaluation, we utilized time required in seconds to generate single image as critical metrics to gauge the efficiency of each method. Table 3 shows the inference time per image by selecting the sampling steps ranging from 20 to 24 to render robust imaging output for each medical imaging modality. It can be observed from Table 3 that DDIM requires the highest inference time, with larger sampling steps especially in case of X-ray imaging data whereas Euler and Euler A relatively require less inference time for all the imaging modalities thus making this computationally less expensive.

		Dermascopic (n=665)		ImageTBAD (n=52)			Chest X-Ray (n=242)	
		Benign	**Malignant**	**TL**	**TL+FL**	**FLT**	**Normal**	**Pneumonia**
LPIPS	**DDIM**	0.575	0.586	0.446	0.444	0.418	0.481	0.493
	Euler	0.571	0.606	0.447	0.445	0.415	0.480	0.492
	Euler A	0.558	0.597	0.442	0.447	0.417	0.479	0.497
	PLMS	0.583	0.586	0.564	0.530	0.540	0.486	0.492
SSIM	**DDIM**	0.621	0.503	0.245	0.259	0.244	0.305	0.416
	Euler	0.631	0.463	0.234	0.261	0.259	0.310	0.422
	Euler A	0.642	0.498	0.265	0.261	0.268	0.305	0.400
	PLMS	0.581	0.556	0.144	0.171	0.153	0.291	0.404

Table 2: Quantitative comparison of sampling methods (DDIM, Euler, Euler A, PLMS) across three datasets, using LPIPS and SSIM to measure image similarity to real data for each modality. Green colored numbers represent higher perceptual similarity (lower LPIPS), while blue colored numbers indicate better structural similarity (higher SSIM) between real and synthetic images.

		Dermoscopy		CTA			X-Ray	
Sampling Steps	**Sampler**	**Benign**	**Malignant**	**TL**	**TL+FL**	**FLT**	**Normal**	**Pneumonia**
20	**Euler**	1.6	1.6	1.4	1.6	1.5	1.4	1.4
	Euler A	1.5	1.6	1.6	1.6	1.5	1.3	1.3
	DDIM	1.5	1.5	1.5	1.5	1.5	1.9	1.9
	PLMS	1.5	1.6	1.5	1.6	1.5	1.4	1.4
22	**Euler**	1.7	1.7	1.6	1.7	1.6	1.6	1.5
	Euler A	1.8	1.7	1.7	1.6	1.7	1.5	1.5
	DDIM	1.7	1.8	1.7	1.6	1.7	2.2	2.2
	PLMS	1.8	1.8	1.7	1.8	1.7	1.6	1.6
24	**Euler**	1.9	1.8	1.8	1.8	1.6	1.6	1.6
	Euler A	1.9	1.8	1.8	1.6	1.8	1.6	1.6
	DDIM	1.9	1.9	1.8	1.8	1.8	2.1	2.4
	PLMS	1.9	1.9	1.8	1.9	1.9	1.7	1.7

Table 3: Inference time (in seconds per image) across various sampling step intervals for four distinct sampling methods. The green values denote the lowest inference times achieved for each medical imaging modality, emphasizing the superior efficiency of the corresponding sampling method compared to the other methods.

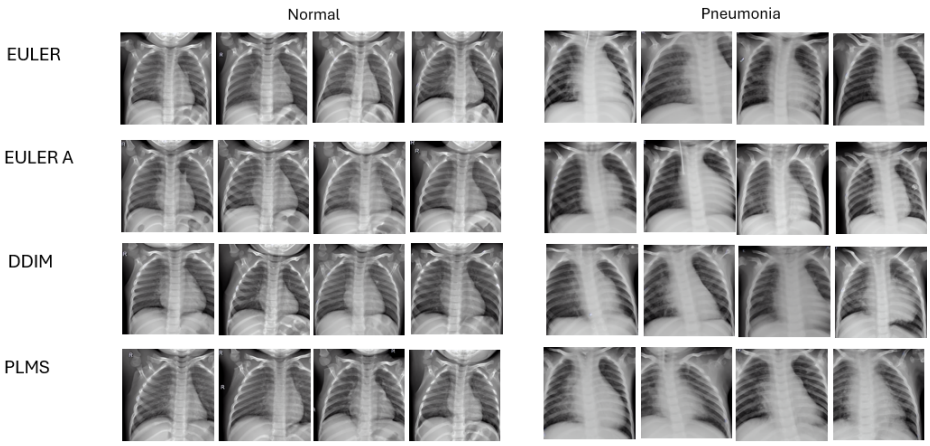

Fig. 6: Synthetic chest X-ray images of pediatric patients with and without pneumonia highlight lung conditions. The first column shows healthy lungs, while the second depicts pneumonia-related opacities. Images were generated using four sampling methods—Euler, Euler A, DDIM, and PLMS.

6 Conclusion and Future Work

This study comprehensively evaluates the performance of four distinct diffusion sampling methods for three different medical imaging modalities. By employing a combination of structural and perceptual similarity metrics along with t-SNE visualization, and inference time per image, we have provided a balanced and thorough assessment of each sampling method's effectiveness in rendering diversified medical data. The evaluation highlights differences in image quality, perceptual fidelity, and computational efficiency, ensuring a robust and fair comparison between the sampling methods. From our experimental findings, we concluded that different image sampling methods perform variably across different medical imaging modalities. The Euler A method yields the highest perceptual quality (lower LPIPS score) and highest structural similarity index measure (SSIM) score for benign dermoscopic images. In CTA imaging, the PLMS sampler introduces the most noise across all subclasses, as evidenced by its high LPIPS scores (indicating low perceptual similarity) and low SSIM scores (reflecting poor structural similarity). Notably, both Euler and Euler A require less inference time across all modalities, making them efficient for generating high-quality synthetic data. Thus we concluded that Euler sampling method performs good by generating robust quality data for most of the selected medical imaging modalities and further Euler and Euler A also requires least inference time for all the medical imaging modalities. This highlights the importance of selecting the right sampling method for balancing data quality and computational efficiency.

For future work, we aim to explore more advanced sampling techniques and integrate additional evaluation metrics, particularly those focusing on clinical

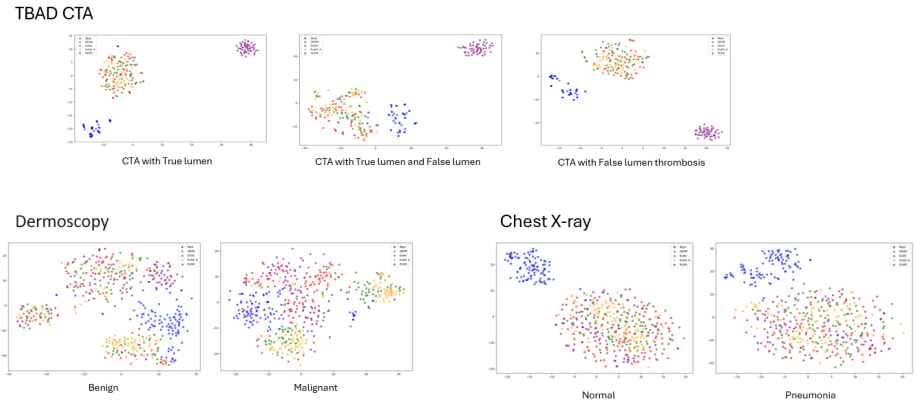

Fig. 7: t-SNE visualization comparing synthetic and real images across each imaging modality, showcasing results for all four sampling methods.

relevance and diagnostic accuracy. Expanding the study to cover more diverse medical imaging modalities and datasets will further strengthen the generalizability of the findings. Additionally, optimizing sampling methods for improved computational performance without compromising image quality remains a priority, with the goal of enhancing real-time medical image generation in clinical settings.

Acknowledgements

The first author would like to thank the research funding from the College of Science and Engineering. In addition, the research conducted in this publication was jointly supported by ADAPT - Centre for Digital Content Technology, Enterprise Ireland, Irish Research Council under grant number IRCLA/2023/1992 and with the financial support of Science Foundation Ireland under Grant Agreement No SFI/12/RC/2289_P2.

References

1. Abaid, A., Farooq, M.A., Hynes, N., Corcoran, P., Ullah, I.: Synthesizing cta image data for type-b aortic dissection using stable diffusion models. arXiv preprint arXiv:2402.06969 (2024)
2. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(9), 10850–10869 (2023)
3. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
4. Farooq, M.A., Yao, W., Schukat, M., Little, M.A., Corcoran, P.: Derm-t2im: Harnessing synthetic skin lesion data via stable diffusion models for enhanced skin disease classification using vit and cnn. arXiv preprint arXiv:2401.05159 (2024)

5. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
6. Liu, L., Ren, Y., Lin, Z., Zhao, Z.: Pseudo numerical methods for diffusion models on manifolds. arXiv preprint arXiv:2202.09778 (2022)
7. Müller-Franzes, G., Niehues, J.M., Khader, F., Arasteh, S.T., Haarburger, C., Kuhl, C., Wang, T., Han, T., Nolte, T., Nebelung, S., et al.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Scientific Reports **13**(1), 12098 (2023)
8. Pan, S., Abouei, E., Wynne, J., Chang, C.W., Wang, T., Qiu, R.L., Li, Y., Peng, J., Roper, J., Patel, P., et al.: Synthetic ct generation from mri using 3d transformer-based denoising diffusion model. Medical Physics **51**(4), 2538–2548 (2024)
9. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125 **1**(2), 3 (2022)
10. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I.: Zero-shot text-to-image generation. In: International conference on machine learning. pp. 8821–8831. Pmlr (2021)
11. Razavi, A., Van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with vq-vae-2. Advances in neural information processing systems **32** (2019)
12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
13. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22500–22510 (2023)
14. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
15. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8798–8807 (2018)
16. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys **56**(4), 1–39 (2023)
17. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)