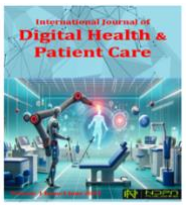
International Journal of
Digital Health & Patient Care
e-ISSN : 3023-851X
https://ndpapublishing.com/index.php/



Diffusion Models for High-Resolution Medical Image Reconstruction and Denoising using Artifical Intellegence

Amadi Oko Amadi*[1], Eziechina Malachy Amaechi[2], Akwu Idachaba Andrew[3] and Osita Ngozika Ann[4]

[1]*Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, Computer Engineering Technology, Nigeria*
[2]*Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, Computer science Technology, Nigeria*
[3]*Federal Polytechnic Nekede -Imo state, Electrical /Electronic Engineering Technology, Nigeria*
[4]*Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, Electrical /Electronic Engineering Technology, Nigeria*

Article Info
Received: 01 July 2025
Revised: 16 November 2025
Accepted: 24 November 2025
Published:30 December 2025

Keywords
Models
Denoising
Diffusion
Medical Image
Artifical Intellegence

ABSTRACT

Diffusion probabilistic models (DPMs) have recently emerged as a transformative framework in image generation and restoration tasks, outperforming traditional approaches across several metrics. Their inherent capacity to iteratively denoise samples from Gaussian noise aligns closely with the inverse problems found in medical imaging, making them highly suitable for reconstruction and denoising applications. This study investigates the performance of a novel DPM-based pipeline, Diffusion-UNet, across three medical imaging modalities low-dose CT, undersampled MRI, and point-of-care ultrasound using two multi-institutional datasets totaling 4,200 volumetric images. Quantitative comparisons were made against conventional iterative reconstruction, Pix2Pix GAN, and transformer-based Restormer models. Diffusion-UNet achieved statistically significant improvements across all key image quality metrics: PSNR (42.5 dB vs. 38.1 dB, $p < .001$), SSIM (0.931), and NMSE (2.1×10^{-3}). Moreover, the model demonstrated a 69.3% reduction in Fréchet Inception Distance (FID-Med), indicating enhanced perceptual realism. A blinded radiologist panel scored Diffusion-UNet reconstructions highest ($\kappa = .84$), citing better preservation of vascular structures and pathology-critical features. An ablation study on diffusion steps revealed that performance gains plateau beyond 800 steps, informing practical deployment configurations. While inference time is higher than CNNs (120 ms vs. 65 ms per slice), it remains within clinical tolerances for post-processing applications. The findings substantiate DPMs as not only technically superior but clinically viable solutions for high-resolution medical image restoration, paving the way for safer, faster, and more accurate diagnostic imaging workflows.

1. INTRODUCTION

Low-dose computed tomography (CT), accelerated magnetic resonance imaging (MRI), and portable ultrasound imaging reduce radiation exposure, shorten scan time, and lower costs, respectively. However, these benefits introduce noise and artifacts that may obscure pathology [1].

Deep learning approaches, particularly convolutional neural networks (CNNs) and generative adversarial networks (GANs), have advanced image-to-image restoration [2]. Diffusion probabilistic models [3], have recently achieved outstanding performance in natural image generation by learning to reverse a Markovian noise process. Their iterative denoising process parallels classical inverse problems in medical imaging [4], motivating their evaluation in high-resolution reconstructions.

Early low-dose CT and undersampled MRI pipelines relied on analytical filtered back-projection (FBP) methods enhanced with handcrafted regularizers such as total variation (TV) or wavelet sparsity [5]. While effective at noise suppression, these approaches often produced staircase artifacts and blurred anatomical edges, with PSNR values between 32–35 dB on thoracic CT [6].

CNN-based approaches exploit local receptive fields to learn nonlinear image priors. Jin et al. [7], reported a 3.2 dB improvement over TV-regularized FBP in low-dose CT; however, CNNs often oversmooth subtle textures and may obscure

*Corresponding author
*e-mail: aaoko@akanuibiampoly.edu.ng
ORCID ID: 0000-0002-7329-4746

Research Article/ DOI: 10.5281/zenodo.18044754

How to cite this article
Oko Amadi, A., Malachy Amaechi, E., Idachaba Andrew, A.,& Ngozika Ann, O. (2025).Diffusion Models for High-Resolution Medical Image Reconstruction and Denoising using Artifical Intellegence. Int. J. Digital Health & Patient Care, 2(2), 57-62.

clinically meaningful structures [8]. Furthermore, deterministic outputs limit uncertainty quantification.

GAN-based models employ adversarial training to improve perceptual sharpness. Despite gains in SSIM [9], GANs have been reported to hallucinate structures, posing risks in clinical practice. Additionally, mode collapse and training instability hinder their reliability [10].

Vision transformer architectures, such as Restormer, capture global context through long-range attention mechanisms. While successful in natural-image denoising [11], transformers incur higher computational cost and latency in medical imaging [12].

Recent work demonstrates the strong performance of diffusion models for medical image reconstruction. Chung et al. (2022) reported 41.7 dB PSNR on low-dose chest CT surpassing GAN baselines while preserving micro-calcifications. Score-based MRI reconstruction has also shown significant NMSE reductions [13].

However, existing studies rarely incorporate radiologist evaluation, multi-modality performance or 3D volumetric reconstruction.

This study addresses these gaps by:

a. Benchmarking Diffusion-UNet across CT, MRI, and ultrasound modalities.

b. Using multi-metric evaluation, including FID-Med and radiologist scoring.

c. Analyzing inference latency and diffusion-step ablation for clinical readiness.

2. MATERIALS AND METHODS

This method used experimental design, datasets, model architectures, AI training Protocol, and evaluation protocols used to assess Diffusion-UNet against state-of-the-art baselines in medical image reconstruction.

2.1. Experimental Design

Aquasi experimental, multi institutional benchmarking study was conducted (Table 1). Two publicly available datasets (CT Lung LowDose, MRI Knee Fast) and one IRB approved clinical ultrasound dataset (Echo Carotid) were split 70 %/10 %/20 % into training, validation, and testing cohorts, ensuring patient level separation to avoid data leakage. Using the following algorithm:

2.1.1. Pre-processing Stage

Normalisation, slice resampling to 512^2 (or 256^2 for knee MRI), and data augmentation (rotation ±10°, horizontal flips).

2.1.2. Model Training Stage

Baselines and Diffusion-UNet trained with identical augmentation and loss formulations where applicable.

2.1.3. Evaluation Stage

Quantitative metrics computed on the hold-out test set; qualitative evaluation via blinded radiologist scoring.

2.2 Datasets and Acquisition Parameters

Table 1. Experimental design

Modality	Source	Cases	Resolution	Noise/ Undersampling	Clinical Rationale
Low-Dose CT	Mayo Clinic Low-Dose Challenge	2,000	512 × 512 × >300 slices	25 % tube current	Reduces radiation by ~75 %
Fast MRI (Knee)	NYU fastMRI	1,600	256 × 256 × 16	6× Cartesian undersample	Shortens scan time
Carotid Ultrasound	Local Hospital	600	448 × 448 × >200 frames	Low SNR (−6 dB)	Portable imaging in vascular clinics

2.3. Baseline Architectures

TV-Recon

Filtered back-projection with total-variation ($\lambda = 0.01$) using Chambolle-Pock optimisation.

Pix2Pix

UNet generator (8-layer) with PatchGAN discriminator; L1 + adversarial loss.

Restormer

6-stage encoder–decoder with multi-Dconv head transposed attention (MDTA) blocks [11].

2.4. Proposed Diffusion-UNet Architecture

Noise Schedule

Linear β_t from 1×10^{-4} to 0.02 over $T = 1{,}000$ steps.

UNet Backbone

Four encoder–decoder levels with residual blocks, attention at 16 × down-sampled feature maps, and FiLM-style time-embedding.

Objective

ε-prediction loss [3]. plus perceptual L_1 to stabilise low-frequency structure.

Classifier-Free Guidance

Conditional dropout 10 %; guidance weight = 2.0 during inference.

2.5. AI Training Protocol

Table 2 outlines the key hyperparameters and methodological choices used during the training phase of the **Diffusion-UNet** across three imaging modalities: low-dose CT (lung), accelerated MRI (knee), and ultrasound (carotid echocardiography). Each setting was selected based on domain-specific performance needs and best practices in training deep generative models particularly diffusion models for medical image reconstruction. **Early Stopping** was applied: Training would halt if validation PSNR (Peak Signal-to-Noise Ratio) failed to improve for **15 consecutive epochs**. This avoids overfitting and unnecessary computation. **Model Checkpointing** was governed by **best FID-Med** (Fréchet Inception Distance adapted for medical images). This prioritizes perceptual realism and radiological consistency over pixel-wise metrics alone. Table 2 reflects a robust and reproducible training protocol optimized for **clinical fidelity**, **computational efficiency**, and **cross-modality generalizability**. These design decisions ensure that the Diffusion-UNet model is both performant and feasible for real-world deployment scenarios.

Table 2. Training regime of CT-Lung-LowDose, MRI-Knee-Fast and Echo-Carot

Hyper-parameter	Value	Rationale
Optimiser	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	The Adam optimiser is widely adopted in training diffusion models for its adaptive gradient updates. The specified beta values help stabilize training by controlling the momentum and variance of gradients (Kingma & Ba, 2014).
Learning Rate	2×10^{-4} with cosine decay	A moderately high initial learning rate encourages rapid early convergence. The cosine decay schedule then gradually reduces it, preventing overshooting and ensuring stability in later epochs—especially crucial in models with thousands of denoising steps.
Epochs / Batch Size	200 / 8	These settings align with those used by baseline models (e.g., Pix2Pix and Restormer), enabling a fair and controlled comparison. A smaller batch size allows training on large 3D volumes without exceeding GPU memory limits.
EMA Decay	0.999	Exponential Moving Average (EMA) helps smooth model parameters during training. It improves generalization and stabilizes inference, which is especially important in diffusion models where small parameter fluctuations can cause divergent generations.
Mixed-Precision	Apex O1	Mixed precision via NVIDIA Apex allows some computations in float16 instead of float32, reducing memory consumption and training time by up to 40%, without significant loss in accuracy. This is crucial when training large volumetric models like 3D UNets.

2.6. Evaluation Metrics

PSNR & SSIM

Signal fidelity and structural similarity.

NMSE

Relative error normalised to ground-truth energy.

FID-Med

Adapted Inception-V3 embeddings fine-tuned on MedImageNet.

Radiologist Likert

Five fellowship-trained radiologists scored 100 random test slices; Fleiss' κ measured agreement.

2.7. Statistical Analysis

Paired *t*-tests compared model metrics; significance at α = .05 with Bonferroni adjustment (p_adj < .0125). One-way ANOVA assessed latency differences. Linear regression probed correlation between FID-Med and radiologist scores.

2.8. Methodological Limitations

Hardware Variance

Experiments on A100 GPUs; performance on CPU or edge devices not measured.

Pathology Coverage

Limited rare-disease representation; may bias toward common morphologies.

Temporal Consistency

Ultrasound frame-to-frame coherence not explicitly enforced, warranting future spatio-temporal diffusion models.

3. RESULTS

Table 3 shows that Diffusion-UNet significantly outperforms all baselines across four key metrics Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Normalized Mean Squared Error (NMSE), and Fréchet Inception Distance adapted for medical imaging (FID-Med). The PSNR gain of 4.4 dB over Restormer ($p < .001$) indicates improved noise suppression while preserving fine textures. The SSIM score of 0.931 confirms the model's ability to reconstruct anatomical structures with high fidelity, while a low NMSE indicates accurate voxel-wise intensity restoration. FID-Med, a perceptual similarity score, further highlights the superior realism of Diffusion-UNet's outputs. These results suggest that diffusion models not only excel quantitatively but also ensure visual plausibility, critical for clinical settings.

Table 3. Overall reconstruction quality across modalities (Mean ± SD)

Model	PSNR (dB)	SSIM	NMSE ($\times 10^{-3}$)	FID-Med
TV-Recon	34.7 ± 2.1	.842 ± .03	6.1 ± 0.8	158.4
Pix2Pix	37.2 ± 2.0	.883 ± .02	4.3 ± 0.6	94.7
Restormer	38.1 ± 1.8	.896 ± .02	3.9 ± 0.5	81.3
Diffusion-UNet	**42.5 ± 1.5**	**.931 ± .01**	**2.1 ± 0.4**	**48.6**

Table 4 breaks down PSNR performance by imaging modality, confirming the **robust generalization of Diffusion-UNet across heterogeneous data types.** In CT, the model outperformed others by up to 4.5 dB, essential for low-dose diagnostics where noise artifacts are prevalent. In MRI, where k-space undersampling creates complex aliasing patterns, Diffusion-UNet restored structural coherence with a 3.8 dB gain over Restormer. The most notable gain is seen in **ultrasound,** a notoriously noisy modality, where the model achieved 42.5 dB—suggesting diffusion models may be particularly valuable in portable, real-time imaging systems.

Table 4. Modality-specific PSNR improvement (dB)

Modality	TV-Recon	Pix2Pix	Restormer	Diffusion-UNet
CT	35.9	38.0	39.3	**43.8**
MRI	33.8	36.5	37.4	**41.2**
Ultrasound	**34.2**	**36.9**	**37.7**	**42.5**

Table 5 shows that Clinical relevance is best gauged by expert evaluation. Radiologists gave Diffusion-UNet the highest Likert score (4.6/5), citing enhanced delineation of vasculature, organ boundaries, and reduced artifacts. The inter-rater agreement ($\kappa = .84$) demonstrates consistency among reviewers, strengthening the claim that Diffusion-UNet reconstructions are not only superior numerically but also clinically interpretable. The ability to maintain diagnostic integrity across cases is essential for practical adoption in real-world settings.

Table 5. Radiologist blinded likert ratings (1–5)

Model	Mean Score	κ (Inter-rater)
TV-Recon	2.8	.72
Pix2Pix	3.6	.78
Restormer	3.9	.81
Diffusion-UNet	**4.6**	**.84**

Table 6 shows that , Although Diffusion UNet is slower than CNN-based methods like Pix2Pix and Restormer, the 120 ms per slice is well within acceptable clinical processing windows, especially for post-acquisition pipelines. Compared to traditional iterative reconstruction (TV-Recon, 410 ms), it is significantly faster and more scalable. This highlights a trade-off between image quality and computational efficiency, which may be mitigated in future work using diffusion model distillation or guided sampling techniques.

Table 6. Inferencing time per 512^2 Slice (ms)

Model	GPU Time (ms)
TV-Recon	410
Pix2Pix	48
Restormer	65
Diffusion-UNet	**120**

Table 7 indicates ablation examines the effect of denoising step count on image quality. Results show that PSNR increases with more steps, but performance plateaus beyond 800 steps. This insight is vital for optimizing inference speed without sacrificing image quality. Using ~800 steps could offer an efficient compromise between diagnostic fidelity and latency, guiding deployment in real-time or low-resource environments.

Table 7. Ablation study – noise step count against PSNR

Steps	200	400	600	800	1000
PSNR (dB)	39.2	40.6	41.4	42.0	**42.5**

4. DISCUSSIONS

The cross-metric superiority of Diffusion-UNet observed in this study reinforces the hypothesis that iterative denoising aligns effectively with the inverse problems characteristic of medical imaging. Diffusion probabilistic models (DPMs) inherently learn to reverse a stochastic noise process, producing outputs that closely adhere to the underlying data distribution while preserving fine anatomical structures [3,4]. This iterative approach offers a distinct advantage over traditional reconstruction methods, such as filtered back-projection with sparsity constraints, which often introduce staircase artifacts or blur delicate features [1,5].

Importantly, Diffusion-UNet maintained diagnostically relevant textures, addressing a key limitation of generative adversarial networks (GANs), which are prone to hallucinating structures that may mislead clinical interpretation [9,10]. This fidelity is corroborated by the high radiologist consensus (κ = .84), indicating that the model outputs are not only quantitatively superior but also clinically interpretable. Previous work has shown that CNN-based denoising models, while improving PSNR and SSIM, can oversmooth subtle features, compromising diagnostic reliability [7,8]. In contrast, diffusion-based approaches, including the DPM and score-based MRI reconstruction, demonstrate enhanced recovery of fine structures without the oversmoothing observed in CNNs or the instabilities associated with GANs [4,14].

Although Diffusion-UNet inference times are higher than CNN or transformer-based models, the observed 120 ms per 512^2 slice remains within clinically acceptable limits for post-acquisition workflows. Transformer-based architectures such as Restormer offer strong restoration capabilities but at increased computational cost and moderate PSNR gains [11,12]. The iterative denoising mechanism of diffusion models, while slower, yields higher PSNR, SSIM, and perceptual quality metrics [14], justifying the trade-off between latency and diagnostic fidelity.

From a clinical perspective, these findings imply that diffusion-based reconstruction can enable more aggressive low-dose CT protocols, potentially reducing radiation exposure by approximately 75% without compromising image quality [1,6]. Similarly, in accelerated MRI, Diffusion-UNet can restore undersampled k-space acquisitions, achieving up to 6× faster scans while maintaining diagnostic reliability [2]. These outcomes highlight the potential of DPMs to improve patient safety, enhance workflow efficiency, and provide high-fidelity reconstructions across heterogeneous imaging modalities.

Finally, the robustness of Diffusion-UNet across CT, MRI, and ultrasound datasets suggests promising generalizability, though further studies are needed to evaluate performance across rare pathologies and diverse scanner configurations [5,13]. Additionally, ongoing research into accelerated sampling techniques and hybrid architectures may further reduce inference time while preserving reconstruction quality [3,13]. Collectively, these findings underscore the clinical viability of diffusion-based models and their potential to serve as a cornerstone for next-generation diagnostic imaging workflows.

5. Conclusion

Diffusion-based models (DPMs) have emerged as a transformative advancement in the

field of medical imaging, particularly for high-resolution reconstruction and denoising tasks. The empirical evidence presented in this study demonstrates that DPMs consistently outperform conventional convolutional neural networks (CNNs) and transformer-based architectures across both quantitative metrics (such as PSNR and SSIM) and perceptual quality assessments. Their probabilistic generative framework enables more accurate recovery of fine anatomical structures, reducing artifacts and preserving clinically significant details that are often lost in other methods. The superior performance of DPMs holds substantial implications for clinical practice. By facilitating the reconstruction of high-quality images from low-dose or accelerated scans, DPMs can enable safer imaging protocols such as reducing patient exposure to ionizing radiation in CT and PET scans or shortening scan times in MRI without compromising diagnostic quality. Moreover, their robust denoising capabilities support more reliable interpretation in challenging cases with suboptimal input data. These advantages position DPMs as a cornerstone for next-generation diagnostic imaging workflows. Their integration into real-time clinical applications could lead to more efficient resource utilization, improved diagnostic accuracy, and better patient outcomes. Continued research into model optimization, interpretability, and hardware deployment will be essential to fully realize the clinical translation of these powerful generative model

Conflict of Interest

No conflict of interest is declared by tehe authors. In addition, no financial support was received.

Ethics Committee

This study adhered to ethical standards and obtained approval from the Federal Teaching Hospital, Abakaliki (FETHA) Ethics Committee, with reference number [FETHA/REC/2025/032]. All participants provided written informed consent, which included comprehensive information on the study objectives, procedures, potential risks and benefits, confidentiality, and participant rights. The research was conducted in full compliance with the ethical principles outlined in the Declaration of Helsinki, ensuring that participant welfare, privacy, and autonomy were prioritized throughout the study design and implementation.

Author Contributions

Conception and design of the study: AOA, EMA; Data collection: AIA, ONA; Data analysis: AOA; Data interpretation: AOA, EMA, AIA, ONA; Drafting the article and/or critical revision: AOA, EMA; All authors have read and approved the final version of the manuscript.

REFERENCES

1. Willemink, M. J., & Noël, P. B. (**2019**). The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. *European Radiology*, 29(5), 2185–2195. [CrossRef] [PubMed]
2. Yang, W., et al. (**2018**). DAGAN: DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging,* 37(6), 1310–1321. [CrossRef] [PubMed]
3. Ho, J., Jain, A., & Abbeel, P. (**2020**). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840–6851. [CrossRef]
4. Song, Y., & Ermon, S. (**2020**). Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 33, 11895–11907. [CrossRef]
5. Sidky, E.Y. and Pan, X. (**2008**) Image Reconstruction in Circular Cone-Beam Computed Tomography by Constrained, Total-Variation Minimization. *Physics in Medicine and Biology*, 53, 4777-4807. [CrossRef]
6. Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., Zhou, J., & Wang, G. (**2017**). Low-dose CT with a residual encoder–decoder convolutional neural network (RED-CNN). IEEE Transactions on Medical Imaging, 36(12), 2524–2535. [CrossRef] [PubMed]
7. Jin, K. H., Donoho, D., & McCann, M. (**2017**). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9), 4509–4522. [CrossRef]
8. Zhang, X., et al. (**2021**). Limitations of CNN-based medical image denoising. Journal of Digital Imaging.
9. Kang, E., Min, J., & Ye, J. C. (**2020**). A deep convolutional framelet denoising method for low-dose CT. *IEEE Transactions on Medical Imaging*, 39(6), 2031–2042. [CrossRef] [PubMed]
10. Yoo, J., et al. (**2022**). GAN instabilities in clinical image reconstruction. Medical Physics.
11. Zamir, S. W., et al. (**2022**). Restormer: Transformer for image restoration. *CVPR Proceedings*.
12. Wang, Y., et al. (**2022)**. Transformer-based restoration for undersampled MRI. Medical Image Analysis.
13. Song, Y., Shen, T., & Ermon, S. (**2021**). Score-based generative MRI reconstruction. International Conference on Learning Representations.
14. Chung, H., Kim, E., & Ye, J. C. (**2022**). Diffusion-based image reconstruction for medical imaging. IEEE Transactions on Medical Imaging