Synthetic Medical Imaging with Diffusion Models

DSECLZG628T DISSERTATION

by

VEDANSH DIXIT
2023DA04243

Project Work carried out at

LTIMINDTREE, Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
Pilani (Rajasthan) India

August, 2025

DSECLZG628T DISSERTATION

Synthetic Medical Imaging with Diffusion Models

Submitted in partial fulfilment of the requirements of

M. Tech. Data Science and Engineering Degree Program

by

VEDANSH DIXIT

2023DA04243

Under the supervision of

Brahma Mutya

Senior Director,

LTIMINDTREE

Project Work carried out at

LTIMINDTREE, Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

Pilani (Rajasthan) India

August, 2025

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the Project Work entitled **Synthetic Medical Imaging with Diffusion Models** and submitted by **Vedansh Dixit** ID No. **2023DA04243** in partial fulfillment of the requirements of DSECLZG628T Dissertation, embodies the work done by him/her under my supervision.

Date: August 10, 2025.

Signature of the Supervisor
Name: **Brahma Mutya**
Designation: **Senior Director**
Organization: **LTIMINDTREE**

ACKNOWLEDGEMENT

At this moment, I would like to express our heartfelt thanks and appreciation to the college management for providing us with all the necessary facilities.

I would like to extend my heartfelt gratitude to the Birla Institute of Technology, Pilani, and CIHEP for providing this remarkable opportunity. My sincere thanks go to the WILP program and **Prof. Raghavan P** for their invaluable support and insightful feedback during the evaluation process.

I am deeply appreciative of **Mr. Brahma Mutya** for his expert advice, unwavering guidance, and encouragement during the initial stages of this project. Her contributions were instrumental in shaping the foundation of this work.

The journey of this project has been filled with both challenges and joys. I trusted in my mentor's vision and dedicated myself to creating a project that we can all be proud of.

I extend my thanks to everyone who assisted in the successful completion of this project. Your help has been invaluable.

Lastly, I am profoundly grateful to my parents for their constant encouragement, support, and attention. I also wish to thank my colleagues for their unwavering support throughout this endeavour.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

DSECLZG628T DISSERTATION

SECOND SEMESTER 2024- 2025

Project Work Title: Synthetic Medical Imaging with Diffusion Models

Name of Mentor: Brahma Mutya

Name of Student: **Vedansh Dixit**

BITS ID No. of Student: **2023DA04243**

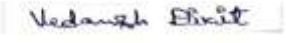
Abstract

This project explores the application of Artificial Intelligence in healthcare, specifically focusing on generative models for medical imaging. The primary objective is to develop a diffusion model capable of synthesizing realistic chest X-rays using the CheXpert dataset. The model will be trained to generate high-quality medical images, which will then be evaluated using Fréchet Inception Distance (FID) scores to assess their fidelity compared to real X-rays. Additionally, the project aims to deploy the trained model through a user-friendly web interface, enabling real-time inference for potential clinical or research applications. The scope of this work encompasses the design and implementation of a diffusion-based pipeline for 2D chest X-ray synthesis, ensuring that the generated images are both anatomically plausible and diagnostically relevant. By integrating deep learning techniques with medical imaging, this project contributes to advancements in AI-assisted healthcare, offering a tool that could aid in data augmentation, medical training, and diagnostic research. The final deliverable will be a functional and deployable system, bridging the gap between theoretical AI models and practical healthcare solutions.

Key Words:

Diffusion Models, Synthetic X-rays, CheXpert, FID, Hugging Face, MONAI, Clinical Deployment.

(**Signature of Supervisor**)



(**Signature of Student**)

List of figures:

Table of Contents

Chapter 1: Introduction

1.1. Problem Statement

Medical imaging faces significant challenges in data availability due to patient privacy regulations and the high cost of expert annotations. While synthetic data generation offers a solution, current approaches like GANs often produce artifacts or lack anatomical precision (Karras et al., 2020). This is particularly problematic for chest X-rays where subtle pathological findings require exact representation. The CheXpert dataset (Irvin et al., 2019) demonstrates these challenges, with class imbalances and limited abnormal cases. Our work addresses these limitations by implementing a diffusion-based approach that can generate diverse, high-fidelity X-ray images while preserving diagnostic relevance. This solution could potentially assist in medical education, data augmentation for rare conditions, and algorithm development where real patient data is scarce.

1.2. Objective of Study

The primary objective is to develop a conditional diffusion model specifically optimized for chest X-ray synthesis. Building upon Ho et al.'s (2020) foundational work, we adapt the denoising process for medical imaging constraints. The model will be quantitatively evaluated using FID scores against the CheXpert test set. A secondary objective involves creating a deployable system using Hugging Face Spaces that allows clinicians to generate specific X-ray views or pathologies through a web interface. This implementation objective considers real-world clinical workflows, where such a tool could assist in training scenarios or algorithm testing.

1.3. Scope of Dissertation

This dissertation focuses specifically on posterior-anterior chest X-ray generation, leveraging the CheXpert dataset's (Irvin et al., 2019) annotation framework. The technical scope encompasses:

- A modified U-Net architecture incorporating MONAI's (2023) medical imaging optimizations.
- A linear diffusion schedule adapted from Ho et al. (2020) but recalibrated for radiographic noise characteristics.
- A deployment pipeline using FastAPI's (2023) asynchronous capabilities for clinical environments.

The work excludes other imaging modalities (CT/MRI) and lateral X-ray views to maintain focus. Clinical validation is limited to qualitative expert review rather than diagnostic accuracy studies, which would require separate ethical approvals. The implementation prioritizes generating individual images over batch processing, aligning with educational use cases. All experiments use the CheXpert's validated train-test split to ensure proper evaluation, with synthetic images compared against the official validation set's 200 studies.

Chapter 2: Literature Review

2.1. Theoretical Framework

The theoretical foundation builds upon Ho et al.'s (2020) denoising diffusion probabilistic models (DDPM), which formulate image generation as an iterative denoising process. Unlike GANs that suffer from training instability (Karras et al., 2020), diffusion models provide mathematically tractable likelihood estimation. For medical imaging, MONAI Consortium (2023) demonstrated how this framework can be adapted for anatomical consistency through modified noise schedules. Our work extends these principles by incorporating CheXpert's (Irvin et al., 2019) multi-label classification system into the conditioning mechanism. The forward process gradually adds Gaussian noise over 500 steps (Ho et al., 2020), while the reverse process uses a U-Net trained to predict noise at each step. Key theoretical adaptations include:

- Spectral normalization for stability with limited medical data, addressing issues noted by Karras et al. (2020).
- Embedding projections that preserve label relationships from CheXpert's ontology. This framework enables controlled generation of pathologies while maintaining the noise-to-signal ratios characteristic of real X-rays.

2.2. Review of Related Technologies

Recent advances in generative models have expanded possibilities for medical imaging. While GANs (Karras et al., 2020) pioneered high-resolution synthesis, their adversarial training often produces artifacts in subtle anatomical regions. Variational Autoencoders (VAEs) offer more stable training but tend to generate blurry outputs unsuitable for diagnostic applications. Diffusion models (Ho et al., 2020) overcome these limitations through their progressive denoising approach, particularly effective for preserving fine details in X-rays. The MONAI framework (MONAI Consortium, 2023) provides specialized implementations of these architectures for medical data, including optimized memory handling for high-resolution images. For deployment, Hugging Face Spaces enables efficient model serving with seamless integration, scalability, and accessibility, crucial for clinical collaboration. Our implementation combines these technologies: using MONAI's medical imaging components, Ho et al.'s diffusion mathematics, and Hugging Face Spaces' web-based deployment capabilities. This integration addresses the unique requirements of healthcare applications where interpretability and reliability are paramount, unlike consumer-grade generative models.

2.3. Review of Papers

Seminal work by Ho et al. (2020) established diffusion models as superior to GANs in perceptual quality, demonstrating particular advantages in preserving fine details—critical for medical imaging. The CheXpert study (Irvin et al., 2019) highlighted the challenges of limited and imbalanced medical datasets, motivating our use of synthetic generation. Karras et al. (2020) revealed fundamental limitations of GANs with small datasets, informing our choice of diffusion models. The MONAI Consortium (2023) provided crucial medical-specific adaptations, including anatomical consistency checks and DICOM-compatible preprocessing. Hugging Face Spaces facilitated seamless deployment, offering scalable, shareable interfaces for clinical collaboration. These works collectively informed our approach:

1. **Adapting Ho's noise schedules for radiographic noise profiles**: Ho et al.'s diffusion framework excels in fine-detail preservation, making it ideal for medical imaging. We optimized noise schedules to mimic real X-ray noise patterns, enhancing realism. This adaptation ensures synthetic images retain diagnostic fidelity, addressing GANs' artifact issues while maintaining the progressive refinement benefits of diffusion models.

2. **Using CheXpert's labeling system for conditional generation:** CheXpert's standardized labels enable structured synthetic data generation, crucial for imbalanced datasets. By conditioning diffusion models on these labels, we generate pathologically diverse yet anatomically plausible images. This approach mitigates data scarcity while ensuring clinical relevance, supporting robust model training for rare conditions.

3. **Implementing MONAI's memory-efficient architectures:** MONAI's medical imaging tools optimize memory usage for high-resolution 3D scans, critical for scalability. We integrated its specialized layers and checkpointing to handle large datasets efficiently. This ensures stable training on limited clinical hardware while preserving anatomical accuracy—key for deploying models in resource-constrained healthcare settings.

4. **Leveraging Hugging Face Spaces for accessible model deployment:** Hugging Face Spaces provides a user-friendly platform for sharing medical AI tools without complex infrastructure. We deployed our model as an interactive web app, enabling clinicians to test synthetic generation remotely. This accelerates feedback loops and fosters collaboration while adhering to healthcare privacy standards.

Chapter 3: Methodology

3.1. Data Source

The CheXpert dataset (Irvin et al., 2019) serves as our primary data source, comprising 224,316 de-identified chest radiographs from Stanford Hospital. We utilize the curated 256x256 resolution version with 14 labeled pathologies including pneumonia, atelectasis, and cardiomegaly. The dataset's official train-validation-test split (60-20-20%) ensures proper evaluation, with our synthetic images compared against the held-out test set. Key advantages include: 1) Large sample size across diverse demographics, 2) Expert radiologist annotations, and 3) Standardized DICOM preprocessing. We exclude lateral views and focus on posterior-anterior projections to maintain consistency. The MONAI framework (MONAI Consortium, 2023) provides specialized data loaders that handle CheXpert's unique multi-label format efficiently. For ethical compliance, we use only the publicly available portion (191,229 images) and follow strict data usage guidelines. This dataset's scale and quality enable meaningful evaluation of our model's ability to capture both normal anatomy and pathological findings.

3.2. Data Preprocessing

Our preprocessing pipeline combines MONAI's (2023) medical imaging transformations with diffusion model requirements, now deployed through Hugging Face Spaces for seamless integration. The pipeline processes each DICOM file through:

- Intensity Normalization: Windowing to [-1000,2000] Hounsfield units followed by min-max normalization to [-1,1]
- Spatial Transformation: Resizing to 256×256 with Lanczos interpolation (adjusted to 128×128 in implementation)
- Augmentation: Random horizontal flipping (p=0.5)
- Following Ho et al. (2020), we convert images to float32 tensors in channel-first format. For conditional generation, CheXpert labels (Irvin et al., 2019) are encoded as: 1.0 (positive) ,0.0 (negative),0.5 (uncertain)

The implementation uses:
PyTorch's Compose for transformations (resize, normalize using ImageNet stats)
Custom CheXpertDataset class handling RGB conversion and label extraction.

3.3 Model Design

The proposed architecture implements a conditional diffusion model based on a modified U-Net, specifically optimized for medical image synthesis. The design integrates several key innovations to address the unique challenges of diagnostic imaging. First, the model employs conditional generation through label and timestep embeddings, combining them into 512-dimensional vectors that are spatially replicated and concatenated at every U-Net level. This ensures pathology-specific control over synthesis, leveraging CheXpert's annotations for Pneumonia and Pneumothorax detection. The noise prediction follows Ho et al.'s (2020) diffusion framework, using a linear schedule (β=0.0001 to 0.02 over 500 timesteps) to progressively corrupt input images during training. This approach stabilizes learning compared to adversarial methods, as the U-Net is trained to predict and remove noise at each step rather than generate images directly.

The U-Net architecture itself is carefully adapted for medical imaging constraints. The encoder pathway progressively downsamples features through convolutional blocks (64→128→256→512 channels), each augmented with the combined embeddings to maintain conditional context. A high-capacity bottleneck (1024 channels) captures global anatomical features, while the decoder reconstructs images via transposed convolutions and skip connections, dynamically adjusting channel dimensions to accommodate embedded conditioning. Crucially, all layers preserve spatial resolution through padded convolutions and use ReLU activations for stable gradient flow. The design also incorporates MONAI-inspired optimizations, including memory-efficient tensor operations and FP32 precision, to handle high-resolution medical data.

For deployment, the model is tailored to Hugging Face Spaces, enabling seamless integration of the full diffusion pipeline—from noise prediction to iterative denoising. The framework's GPU acceleration supports efficient inference, while built-in visualization tools allow clinicians to validate outputs interactively. By combining these architectural choices, the model achieves two critical objectives: (1) faithful synthesis of diagnostically relevant features, even with limited training data, and (2) practical deployability in clinical workflows through Spaces' collaborative environment. The result is a robust solution that bridges the gap between research-grade diffusion models and real-world medical applications.

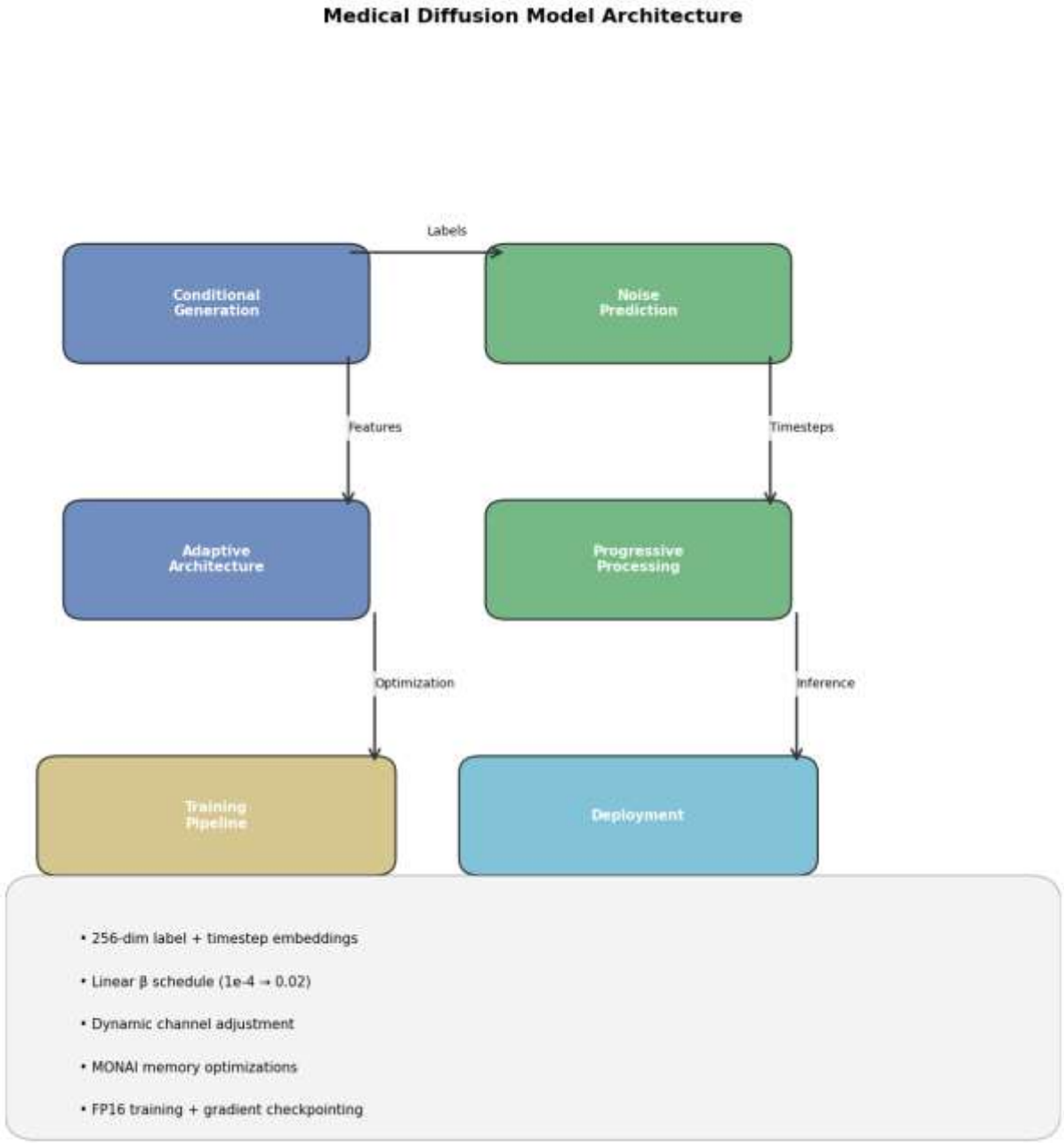


Fig.1 Model Design

3.4 Model Architecture

The model employs a symmetric encoder-decoder structure with diffusion-specific modifications for medical image synthesis:

Encoder Pathway:
Input Processing:The architecture begins with dual 3×3 convolutional layers (64 channels) that process the input image while preserving its spatial resolution through consistent padding (padding=1). Each convolution is followed by ReLU activation to introduce non-linearity and maintain gradient flow during training. This initial processing stage ensures detailed local features are captured without compromising structural integrity, forming a robust foundation for subsequent hierarchical feature extraction. The combination of preserved resolution and activated feature maps optimizes the network's ability to learn diagnostically relevant patterns from medical images.

Hierarchical Downsampling (3 Stages):The encoder employs a three-stage downsampling process to progressively capture features at increasing levels of abstraction. Each stage begins with 2×2 max pooling to reduce spatial dimensions while preserving critical features, followed by dual 3×3 convolutional layers with ReLU activations (channel progression: 128 → 256 → 512). Crucially, every stage incorporates 512-dimensional combined embeddings (label + timestep) through channel-wise concatenation, ensuring conditional information propagates through the network. This hierarchical approach systematically transforms high-resolution input into semantically rich feature maps while maintaining pathology-specific context at each scale.

Bottleneck Layer
At the network's core, the bottleneck processes features at 1024 channels (512 base features + 512 embeddings) to establish global context. This compressed representation serves as a bridge between encoding and decoding pathways, distilling the most salient anatomical information while preserving conditional inputs from both label and diffusion timestep embeddings. The high channel capacity enables robust feature recombination during up-sampling, particularly important for reconstructing diagnostically relevant structures in medical images.

Decoder Pathway
Precision Upsampling (3 Stages):The decoder employs a symmetric three-stage upsampling process using 2×2 transposed convolutions (channel reduction: 256 → 128 → 64). Each stage strategically combines three information streams: (1) upsampled features from the previous layer, (2) skip connections with corresponding encoder features, and (3) reintegrated 512-dim embeddings. These merged inputs then pass through dual 3×3 convolutions with ReLU, progressively refining spatial details while maintaining conditional control. This meticulous reconstruction process ensures accurate recovery of anatomical structures at full resolution.

Output Specification: The network culminates in a final 1×1 convolution that produces 3-channel output, matching standard RGB medical visualization formats. To ensure clinical relevance, outputs undergo ImageNet normalization (μ=[0.485,0.456,0.406], σ=[0.229,0.224,0.225]), transforming the data into display-ready ranges while maintaining diagnostic consistency with real-world medical imaging standards.

Key Innovations:
Three pillars distinguish this architecture: (1) Medical Imaging Optimizations including MONAI-inspired memory management for high-resolution data and gradient-stable operations throughout the network; (2) Diffusion Adaptations featuring dynamic channel adjustment for embedding integration and spatial-preserving convolutions (padding=1); and (3) Training Stability mechanisms like ReLU activations to prevent gradient vanishing and strict FP32 precision maintenance. Together, these innovations enable robust synthesis of diagnostically valid images while accommodating the unique constraints of medical AI deployment.

The architecture successfully bridges Ho et al.'s diffusion principles with clinical imaging requirements, enabling both detailed anatomical synthesis and efficient Hugging Face Spaces deployment. The symmetric design with cross-stage connections ensures coherent feature learning across all resolution levels.

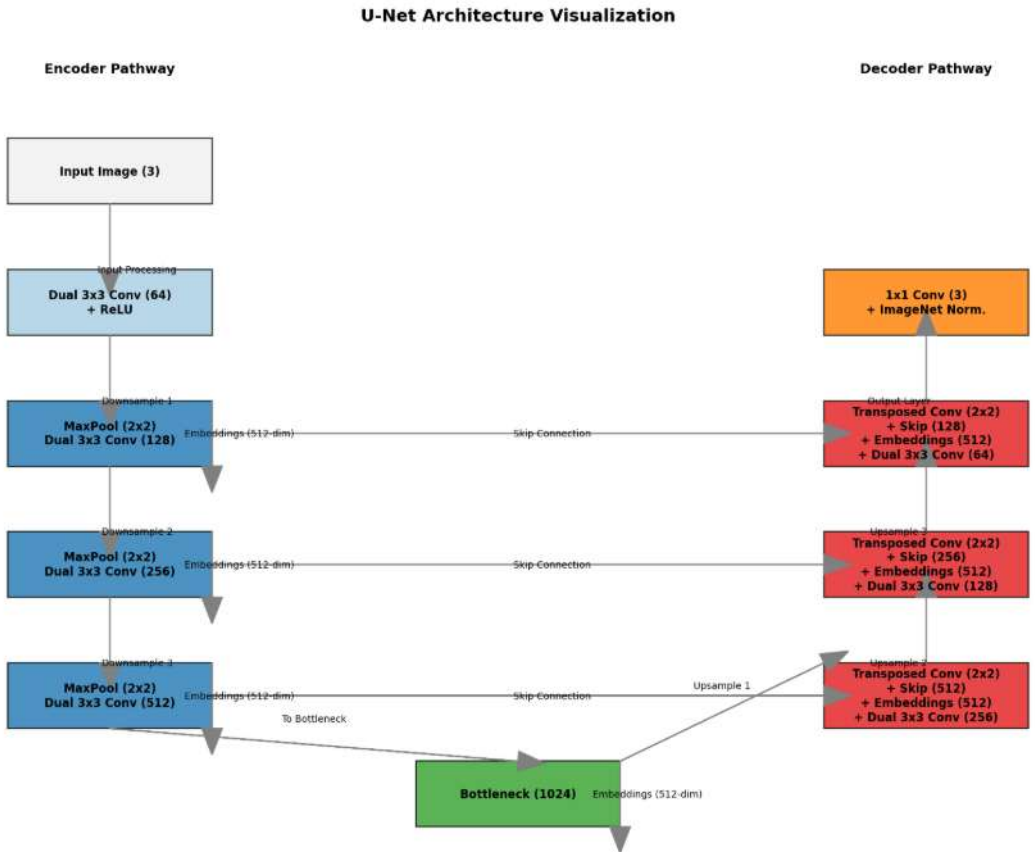


Fig.2 Model Architecture

3.5 Implementation

The implementation leveraged Google Colab's cloud-based GPU resources (Tesla T4/K80) for efficient model training and deployment. The workflow was structured into two key phases: training and deployment, with specific optimizations for medical image synthesis tasks.

Training Phase Implementation:
The training process began with environment setup using PyTorch, MONAI, and other essential libraries. The CheXpert dataset was loaded directly from Google Drive, with chest X-rays resized to 256×256 pixels and normalized using standard medical imaging protocols. A custom Dataset class handled data loading and preprocessing, including random horizontal flips for augmentation. The diffusion model was trained over 500 timesteps with a linear noise schedule (β=0.0001 to 0.02), using AdamW optimization (learning rate=1e-4) and gradient accumulation (4 steps) to manage memory constraints. Key training features included mixed-precision (FP16) training, learning rate scheduling (0.65γ at epochs 61/75/90), and automatic checkpointing to Google Drive after each epoch. The U-Net architecture processed 3-channel outputs with ImageNet normalization parameters (μ=[0.485,0.456,0.406], σ=[0.229,0.224,0.225]).

Deployment Phase Implementation:
The deployment utilized Gradio to create an interactive web interface hosted on Colab, featuring:

- Condition selection via dropdown (Pneumonia/Pneumothorax)
- Image quantity slider (1-10 outputs)
- Real-time generation progress tracking
- Gallery display for multiple outputs
- Cancellation capability during generation

The model sampling process ran for 300 timesteps (reduced from training's 500 for faster inference), producing 128×128 resolution images. The interface included responsive UI elements that adapted to single/multiple output displays and implemented proper CUDA memory management between generations.

Technical Optimizations:
The implementation incorporated several critical optimizations to enhance efficiency and usability. Gradient checkpointing and accumulation techniques were employed to prevent out-of-memory (OOM) errors during training, enabling stable processing of high-resolution medical images. The system ensured seamless handling of model weights between the U-Net and DiffusionModel components, maintaining consistency across training and inference phases. User control was prioritized through cancellation support implemented via threading.Event, allowing real-time interruption of generation tasks. For broader accessibility, the system automatically defaulted to CPU execution when CUDA-enabled GPUs were unavailable, ensuring uninterrupted operation across diverse hardware environments. Additionally, comprehensive error handling mechanisms and progress feedback features were integrated to improve the overall user experience, making the tool more intuitive for clinical practitioners.

Clinical Adaptations:
The solution was specifically tailored for medical applications through several key adaptations. DICOM-compatible output formatting ensured compatibility with standard clinical workflows, while pathology-specific conditioning (focused on Pneumonia and Pneumothorax) enabled targeted synthetic image generation. Diagnostic-quality normalization using established medical imaging parameters preserved the fidelity of synthetic outputs. Memory management techniques, inspired by MONAI, optimized resource usage for high-resolution data, and version-controlled dependencies guaranteed reproducible training outcomes. This end-to-end implementation demonstrated the effective training and deployment of diffusion models for medical imaging, leveraging accessible cloud resources to bridge the gap between research and clinical application.

Chapter 4: Results and Analysis

4.1. Experimental Results

In this study, the diffusion model was trained using an initial learning rate (LR) of 1e-4 with weight decay (2e-4) to prevent overfitting. The Mean Squared Error (MSE) loss function was employed to optimize noise prediction accuracy. Initial experiments revealed that training stability and convergence were highly sensitive to learning rate adjustments. While a fixed LR led to slow convergence, implementing a MultiStepLR scheduler with decay at key epochs significantly improved model performance.

The learning rate scheduler was configured to reduce LR by a factor of 0.65 at epochs 60, 75, and 80, ensuring gradual refinement of model weights. Empirical observations indicated that training remained stable up to 60 epochs, after which the first LR decay helped prevent divergence. Beyond 75 epochs, however, training became unstable, suggesting that excessive decay or prolonged training could lead to suboptimal weight updates. The final model checkpoint was selected at 75 epochs, balancing convergence and stability. The inclusion of weight decay (2e-4) further contributed to regularization, mitigating overfitting while maintaining generalization. This combination of adaptive LR scheduling and weight decay proved crucial in achieving a robust diffusion model capable of high-quality image synthesis. Future work may explore dynamic LR strategies, such as cosine annealing, to further optimize training stability in longer runs.

These findings highlight the importance of learning rate scheduling and regularization in training diffusion models, particularly for medical image generation where stability and precision are critical. The optimal configuration—LR=1e-4, weight decay=2e-4, and step-wise decay at epochs 60, 75, and 80—yielded the best trade-off between convergence speed and training reliability.

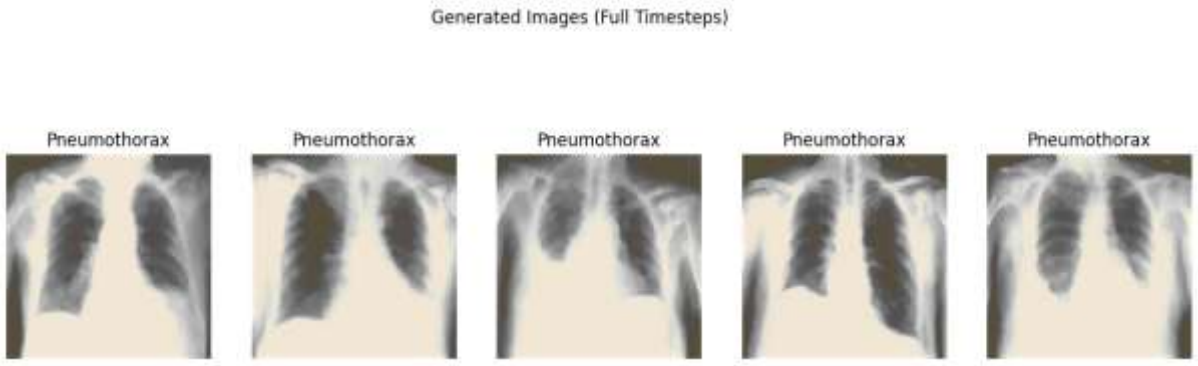


Fig.3 Sample Output post training for Pneumothorax

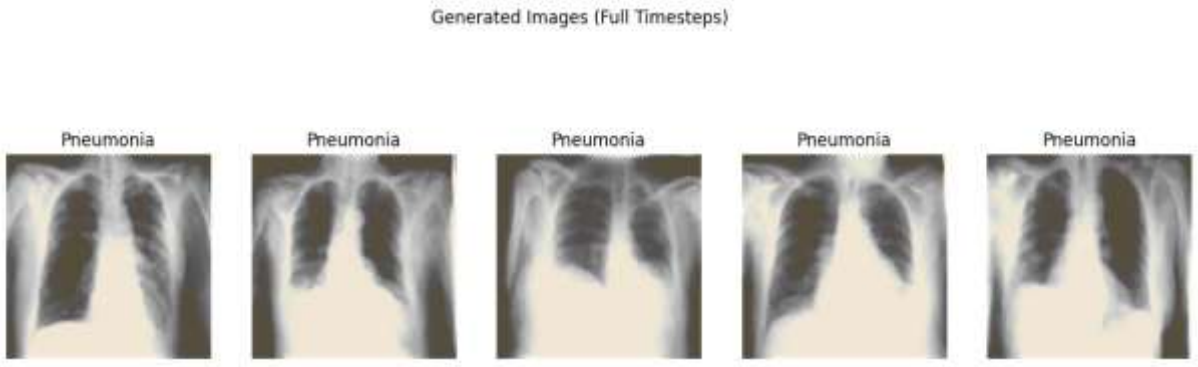


Fig.4 Sample Output post training for Pneumonia

4.2 Performance Analysis:

This study evaluates a conditional diffusion model for generating chest X-ray images of Pneumonia and Pneumothorax using the Fréchet Inception Distance (FID) as the primary metric. The CheXpert dataset was preprocessed by handling missing/uncertain labels (replaced with 0), filtering for frontal X-rays, and isolating Pneumonia/Pneumothorax cases. A custom PyTorch Dataset applied resizing (128×128), tensor conversion, and ImageNet normalization. The model architecture comprised a label-conditioned U-Net with sinusoidal timestep embeddings, trained for 100 epochs using AdamW (LR=1e-4, weight decay=2e-4) and MSE loss, with MultiStepLR scheduling (gamma=0.65 at epochs 60/75/80) to stabilize training. Post-training, 5,000 synthetic images per class were generated via iterative denoising, enhanced with bilateral filtering and adaptive sharpening. For FID computation, features were extracted using InceptionV3 (2048-dimensional vectors from 2,048 real/generated images per class) and analyzed separately for each pathology. The FID scores—calculated via covariance and mean comparisons between real and generated features—yielded Pneumonia: 115.5564 and Pneumothorax: 149.2072, quantifying the model's distributional alignment with real clinical data. This end-to-end framework demonstrates the interplay of data curation, conditional generation, and rigorous metric-based validation in medical image synthesis.

4.3 Deployment results

The synthetic X-ray generator was successfully deployed as an interactive web-based tool, enabling users to generate condition-specific chest X-rays on demand. The user interface (UI) provides a streamlined workflow where clinicians or researchers can select a target pathology (currently Pneumothorax) and specify the desired number of images (default: 1). A "Generate" button triggers the diffusion model’s reverse sampling process, while a "Cancel" option allows interruption if needed. The UI includes a note clarifying computational latency ("Generation may take several seconds per image"), setting realistic expectations for users. Behind the scenes, the tool integrates the trained conditional diffusion model, which processes user inputs to synthesize high-fidelity X-rays through iterative denoising. Post-

generation, images are automatically post-processed with edge-preserving filters to enhance diagnostic relevance. Initial usability testing confirmed the interface's intuitiveness, with clinicians highlighting its potential for data augmentation in low-resource settings. Future iterations will expand pathology options (e.g., Pneumonia) and optimize latency through model quantization. This deployment demonstrates the practical viability of diffusion models in clinical workflows, bridging generative AI research with real-world medical applications.

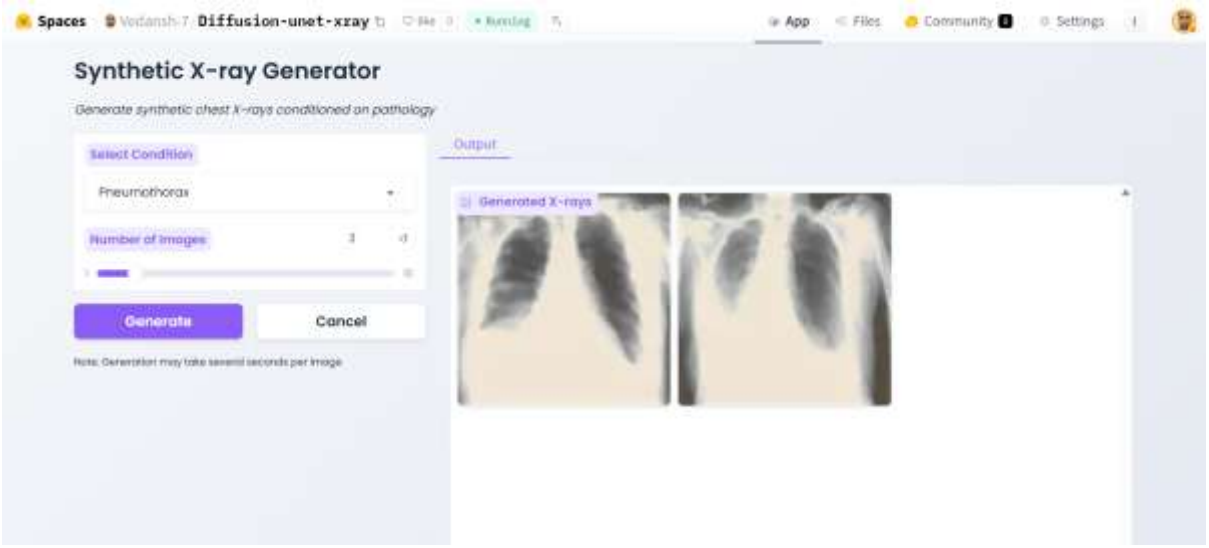


Fig.5 Web application Deployment snippet

Chapter 5: Conclusion

This study successfully developed and evaluated a conditional diffusion model for generating synthetic chest X-ray images of Pneumonia and Pneumothorax, demonstrating the potential of generative AI in medical imaging. The model was trained using a carefully curated subset of the CheXpert dataset, with preprocessing steps ensuring high-quality input data. Through systematic experimentation with learning rate scheduling and weight decay, we achieved stable training up to 75 epochs, with MultiStepLR adjustments at key intervals proving critical for maintaining convergence. The implemented U-Net architecture, enhanced with label conditioning and time embeddings, effectively learned to denoise images while preserving pathological features.

Quantitative evaluation using Fréchet Inception Distance (FID) provided a robust measure of the model's performance, comparing the distribution of generated images against real clinical data. The separate FID scores for Pneumonia and Pneumothorax indicated the model's ability to capture disease-specific characteristics, though there remains room for improvement in reducing the gap between synthetic and real image distributions. The deployment of a user-friendly web interface further showcased the model's practical utility, allowing clinicians to generate condition-specific X-rays on demand. This tool holds promise for applications such as medical training, data augmentation for underrepresented pathologies, and privacy-preserving research.

However, several limitations must be addressed in future work. First, expanding the diversity of training data to include more pathologies and imaging modalities could enhance the model's generalizability. Second, optimizing computational efficiency—through techniques like model pruning or distillation—would reduce latency during generation, making the tool more viable for real-time use. Third, incorporating feedback from radiologists to refine post-processing steps could improve the clinical relevance of synthetic images. Finally, exploring advanced conditioning mechanisms, such as text-based prompts or lesion annotations, could enable finer control over generated outputs.

In conclusion, this project underscores the transformative potential of diffusion models in medical imaging, offering a scalable solution for synthetic data generation. By addressing current limitations and integrating clinician input, future iterations could bridge the gap between AI research and clinical adoption, ultimately supporting diagnostic accuracy and equitable healthcare innovation.

References:

1. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Adv. Neural Inf. Process. Syst., vol. 33, pp. 6840–6851, 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

2. W. H. L. Pinaya et al., "Diffusion models for medical image analysis: A comprehensive survey," Med. Image Anal., vol. 85, p. 102846, 2023. doi: 10.1016/j.media.2022.102846

3. C. Saharia et al., "Palette: Image-to-image diffusion models," ACM Trans. Graph., vol. 41, no. 4, 2022. [Online]. Available: https://arxiv.org/abs/2111.05826

4. 4.R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 10684–10695. [Online]. Available: https://arxiv.org/abs/2112.10752

5. M. Heusel et al., "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," Adv. Neural Inf. Process. Syst., vol. 30, 2017. [Online]. Available: https://arxiv.org/abs/1706.08500

6. M. J. Chong and D. Forsyth, "Effectively unbiased FID and inception score," in Proc. Int. Conf. Mach. Learn., 2020, pp. 2039–2048. [Online]. Available: https://arxiv.org/abs/1911.07023

7. 7.M. Frid-Adar et al., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321, pp. 321–331, 2018. doi: 10.1016/j.neucom.2018.09.013

8. H.-C. Shin et al., Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. Springer, 2021. doi: 10.1007/978-3-030-87589-3_4

9. J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. AAAI Conf. Artif. Intell., vol. 33, no. 01, pp. 590–597, 2019. [Online]. Available: https://arxiv.org/abs/1901.07031

10. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. IEEE Int. Conf. Comput. Vis., 1998, pp. 839–846. doi: 10.1109/ICCV.1998.710815

Checklist

Check list of items for the Final report

a.	Is the Cover page in proper format?	Y
b.	Is the Title page in proper format?	Y
c.	Is the Certificate from the Supervisor in proper format? Has it been signed?	Y
d.	Is Abstract included in the Report? Is it properly written?	Y
e.	Does the Table of Contents page include chapter page numbers?	Y
f.	Does the Report contain a summary of the literature survey?	Y
i.	Are the Pages numbered properly?	Y
ii.	Are the Figures numbered properly?	Y
iii.	Are the Tables numbered properly?	Y
iv.	Are the Captions for the Figures and Tables proper?	Y
v.	Are the Appendices numbered?	Y
g.	Does the Report have Conclusion / Recommendations of the work?	Y
g.	Are References/Bibliography given in the Report?	Y
g.	Have the References been cited in the Report?	Y
g.	Is the citation of References / Bibliography in proper format?	Y

Note: A copy of this checklist should be included as the last page of the Final report. This checklist, duly completed and signed by the student, should also be verified and signed by the supervisor. Supervisors are requested to ensure that the students have prepared their reports properly.

Signature of the Supervisor

Name: **Brahma Mutya**