

ORIGINAL RESEARCH

Open access

# Diffusion Probabilistic Models for Synthetic Chest X-Ray Generation: A Framework for Data Augmentation in Rare Disease Detection with Preserved Pathological Lesions

George Papadopoulos<sup>1\*</sup>, Eleni Georgiou<sup>1</sup>

## Abstract

Rare diseases identified via chest radiography—such as spontaneous pneumothorax, solitary pulmonary nodules, pleural effusions, and cardiomegaly—occur far less frequently than common conditions like pneumonia or chronic obstructive pulmonary disease. Deep learning models require large, balanced datasets for reliable performance, yet rare pathologies remain underrepresented in clinical repositories, limiting real-world deployability. Conventional augmentation methods (geometric and intensity transformations, elastic deformations) add limited variability without creating new pathological patterns. GAN-based approaches can generate synthetic images but often suffer from mode collapse and unrealistic artifacts that reduce lesion fidelity, restricting their effectiveness for rare disease augmentation. We propose a framework based on denoising diffusion probabilistic models (DDPMs) for conditional synthesis of high-fidelity chest X-ray images. The model supports generation conditioned on class labels, segmentation masks, or text prompts, enabling controlled synthesis of rare pathologies and improving dataset balance. The framework includes a forward diffusion process, a U-Net-based reverse denoising model with attention, a multi-modal conditioning mechanism, a lesion-preserving loss function, and an augmentation pipeline combining real and synthetic data. This allows control over lesion type, location, size, and severity, reducing class imbalance and improving classifier performance on rare diseases, as validated through AUC improvements and radiologist assessment. Overall, diffusion-based models provide a scalable and clinically relevant solution for rare disease augmentation in chest radiography, overcoming key limitations of traditional and GAN-based methods and enabling effective use of datasets such as CheXpert, MIMIC-CXR, and ChestX-ray14.

**Keywords** Diffusion probabilistic models, Chest X-ray generation, Data augmentation, Rare disease detection, Lesion preservation, Conditional generation

\*Correspondence:

George Papadopoulos  
george.papadopoulos@gmail.com

<sup>1</sup> Department of Healthcare Analytics and Intelligent Systems, National and Kapodistrian University of Athens, Athens, Greece

## Introduction

Chest radiography remains the most frequently performed medical imaging examination worldwide, serving as the first-line diagnostic tool for pulmonary, cardiac, and mediastinal pathologies. However, rare diseases identifiable on chest X-rays—including spontaneous

pneumothorax (incidence approximately one to five per ten thousand persons annually), solitary pulmonary nodules detected incidentally, and large pleural effusions from uncommon etiologies—suffer from severe underrepresentation in training datasets [1, 2]. As Chen et al. [3] emphasize in their comprehensive review of synthetic data in healthcare, the absence of adequately labeled rare

disease examples represents a fundamental barrier to deploying machine learning systems in clinical settings where these findings carry substantial prognostic significance. Deep learning models optimized for common conditions exhibit markedly degraded performance when applied to these rare but clinically significant findings.

The data scarcity problem for rare diseases produces multiple negative consequences for deep learning-based detection systems. Models trained on imbalanced datasets develop strong priors toward majority classes, leading to overfitting and poor generalization when encountering rare pathologies at test time [4, 5]. Researchers [6] systematically catalog the failure modes of generative models when applied to unbalanced medical data, demonstrating that standard GAN architectures disproportionately represent common findings while failing to capture the statistical signatures of rare lesions. Furthermore, class imbalance directly elevates false negative rates for rare conditions, which is clinically unacceptable when missed findings such as pneumothorax or malignant nodules carry substantial morbidity and mortality implications. As Baumgartner *et al.* [7] note in their analysis of medical image segmentation, conventional approaches to addressing imbalance, including class-weighted loss functions and oversampling, cannot fully compensate for absolute data insufficiency when the minority class contains fewer than several dozen examples.

Traditional augmentation techniques, including random rotations, horizontal flips, intensity shifts, and Gaussian noise injection, introduce only limited variability and fundamentally cannot generate new pathological patterns absent from the original training data [4, 5]. Schlegl *et al.* [4] demonstrated that while geometric augmentations improve robustness to affine transformations, they do not expand the underlying distribution of pathological features. Generative adversarial networks have been extensively investigated for medical image synthesis, with architectures such as StyleGAN and CycleGAN applied to chest X-ray generation. Nevertheless, as comprehensively documented by [6], GANs are prone to mode collapse—producing a limited subset of the data distribution—and generate characteristic artifacts including unrealistic textures, vanishing pathologies, and anatomical distortions that compromise clinical utility. Researchers [8] further demonstrated that GAN-based cross-modality synthesis often fails to preserve fine structural details critical for diagnostic decision-making.

Diffusion probabilistic models, introduced by Ho and colleagues in 2020 [9], have emerged as a superior generative framework across natural image domains, achieving higher sample quality and greater diversity than contemporaneous GANs. Dhariwal and Nichol [10] conducted systematic comparisons showing that diffusion models outperform GANs on perceptual quality metrics while avoiding mode collapse. Subsequent advances including denoising diffusion implicit models (DDIMs) for accelerated sampling introduced by Song *et al.* [11] and latent diffusion models for computationally efficient high-resolution synthesis by Rombach *et al.* [12] have expanded their applicability. Nichol and Dhariwal [13] further improved the noise scheduling and architecture design, enabling high-fidelity generation with fewer timesteps. Karras *et al.* [14] elucidated the design space of diffusion-based models, providing practical guidelines for architectural choices that maximize sample quality. Conditional diffusion variants enable controlled generation of specific pathologies [15, 16], directly addressing the need for targeted rare disease augmentation in medical imaging. Pinaya *et al.* [17] successfully applied latent diffusion models to brain imaging generation, and Khader *et al.* [18] extended these methods to 3D medical image synthesis, establishing the feasibility of diffusion-based approaches across radiological modalities.

## Background

### Rare diseases in chest radiography

Spontaneous pneumothorax, defined as air in the pleural space without traumatic cause, typically presents with apical lucency and visceral pleural line on upright chest X-rays, yet many subtle cases are missed by both radiologists and automated systems. Lung nodules detected incidentally range from benign granulomas to early-stage malignancies, with diagnostic confidence heavily dependent on nodule size, margin characteristics, and density, all of which require extensive labeled examples for deep learning models to learn discriminative features [1, 2]. The CheXpert dataset [1] provides uncertainty labels for multiple pathologies across over 224,000 chest X-rays, while MIMIC-CXR [2] offers 377,000 images with associated radiology reports, yet even these large repositories contain limited examples of truly rare conditions. As Rajpurkar *et al.* [19] demonstrated with CheXNet, achieving radiologist-level performance requires

thousands of positive examples per disease category, a threshold that many rare conditions fail to meet.

Pleural effusions, cardiomegaly, and atelectasis represent additional conditions where automated detection performance degrades substantially when training examples are scarce, directly impacting clinical workflows where these findings may be the only manifestation of serious underlying disease. The diagnostic challenge posed by rare chest pathologies is compounded by their heterogeneous radiographic presentations, where the same pathological entity may appear dramatically different depending on patient body habitus, disease stage, and technical acquisition factors. Deep learning models trained on limited examples of pneumothorax, for instance, may fail to recognize tension pneumothorax with mediastinal shift, loculated pneumothorax following thoracic surgery, or subcutaneous emphysema variants [4, 5]. Baur *et al.* [5] demonstrated that autoencoding models for anomaly detection in medical images require substantially more training examples than commonly assumed to achieve acceptable sensitivity for rare abnormalities. Consequently, rare disease detection constitutes a critical use case where data augmentation through synthetic generation offers substantial potential clinical benefit.

## Data augmentation in medical imaging

Geometric augmentations, including random rotation, horizontal flipping, translation, and elastic deformation, represent the most widely applied techniques for increasing effective training set size in medical image analysis [4, 5]. Intensity-based augmentations such as Gaussian noise addition, contrast adjustment, brightness variation, and histogram equalization simulate scanner variability and patient positioning differences, yet these transformations preserve the underlying anatomical and pathological structure without creating new lesion patterns. While valuable for improving model robustness to acquisition variability, traditional augmentation cannot generate examples of rare pathologies absent from the original dataset.

The fundamental limitation of geometric and intensity augmentation for rare disease applications is that these techniques do not expand the distribution of pathological features available for model training. Schlegl *et al.* [4] demonstrated that a dataset containing only three examples of a subtle pleural nodule, when subjected to rotation and flipping, still contains only the same three

underlying lesion patterns, merely presented from different orientations or with altered contrast. Baur *et al.* [5] confirmed these findings in brain MRI anomaly detection, showing that geometric augmentations alone failed to improve rare lesion detection beyond a minimal threshold. This constraint has motivated the exploration of generative models capable of synthesizing novel, realistic pathology examples that preserve diagnostic features while expanding the diversity of the training distribution.

## Generative models for medical images

Generative adversarial networks have been extensively investigated for medical image synthesis, with applications including cross-modality translation, super-resolution, and data augmentation. StyleGAN-based architectures have demonstrated impressive capability for generating high-resolution chest X-ray images that are visually realistic to radiologists, while CycleGAN enables unpaired image-to-image translation for tasks such as removing or adding pathological findings [6, 8]. However, as researchers [6] systematically document, GANs are notoriously difficult to train stably, frequently suffer from mode collapse where the generator produces a limited variety of outputs, and can generate anatomically implausible features including missing or duplicated structures. Researchers [8] reported that GAN-based fluorescein angiography generation from color fundus images produced clinically unacceptable artifacts in 23% of cases when evaluated by retinal specialists.

Variational autoencoders (VAEs) offer an alternative generative framework with more stable training and explicit latent space organization, yet the images they generate typically exhibit blurring and lack the fine detail necessary for clinical diagnostic tasks. Recent comparative studies across medical imaging domains have demonstrated that diffusion probabilistic models outperform both GANs and VAEs in terms of sample fidelity, diversity, and absence of characteristic generative artifacts [10, 18]. The superior performance of diffusion models arises from their iterative refinement process, which gradually denoises a random sample through many small steps rather than mapping latent codes directly to images in a single pass. Khader *et al.* [18] demonstrated that diffusion models for 3D medical image generation produced substantially fewer anatomical distortions than GAN-based approaches when evaluated by clinical experts.

## Diffusion probabilistic models

Denosing diffusion probabilistic models define a forward process that gradually adds Gaussian noise to a real image over a large number of timesteps, following a variance schedule that determines the noise magnitude at each step. The forward process is Markovian, meaning each step depends only on the immediately preceding state, and is designed such that after sufficiently many steps, the distribution converges to pure Gaussian noise independent of the original image [9]. Ho *et al.* [9] demonstrated that this diffusion process could be reversed through a learned denoising model, enabling high-quality sample generation. The reverse process learns a parameterized denoising model that iteratively removes noise, starting from pure Gaussian noise and progressively recovering a sample from the data distribution, effectively learning to reverse the forward corruption process.

The reverse denoising network is typically implemented as a U-Net architecture with residual blocks, self-attention mechanisms, and positional embeddings that encode timestep information to guide the denoising trajectory. The training objective simplifies to predicting the noise added at each step, encouraging the network to learn the underlying data distribution by estimating the noise component that corrupts each image [9, 13]. Nichol and Dhariwal [13] introduced improved variance scheduling and cosine noise schedules that significantly enhanced sample quality. Subsequent innovations including denoising diffusion implicit models (DDIMs) by Song *et al.* [11] have enabled deterministic sampling with dramatically fewer steps (as few as 50 versus 1000), while latent diffusion models by Rombach *et al.* [12] perform the diffusion process in a compressed latent space, substantially reducing computational requirements for generating high-resolution medical images. Karras *et al.* [14] further refined the diffusion design space, introducing second-order correction terms and adaptive noise schedules that achieve state-of-the-art Fréchet Inception Distance scores on natural image benchmarks.

## Conditional generation and medical applications

The extension of diffusion models to conditional generation has opened new possibilities for controlled medical image synthesis. Researchers [15] developed a vision-language foundation model for chest X-ray generation that accepts text prompts and produces clinically realistic radiographs, demonstrating that radiologists could not reliably distinguish

synthetic from real images in a Turing test. Deshpande *et al.* [16] assessed the capacity of diffusion models to reproduce spatial context in medical images, showing that conditioning on anatomical landmarks significantly improved structural fidelity. Pinaya *et al.* [17] applied latent diffusion models to brain MRI generation, conditioning on age, sex, and diagnosis to produce subject-specific synthetic volumes. Khader *et al.* [18] extended diffusion-based generation to 3D medical images across multiple organ systems, establishing the generalizability of the approach.

Morís *et al.* [20] developed adapted generative latent diffusion models specifically for pathological analysis in chest X-ray images, introducing lesion-aware conditioning mechanisms that preserve disease-specific features. Prusty *et al.* [21] combined latent denoising diffusion probabilistic models with Wiener filtering approaches for enhanced medical image classification, demonstrating significant improvements in rare disease detection tasks. Researchers [22] proposed lesion region inpainting as an approach for pseudo-healthy image synthesis in intracranial infection imaging, establishing methodological precedents for preserving pathological regions during generative processes. Chen *et al.* [23] extended counterfactual conditional diffusion with continuous prior adaptive correction for anomaly detection in multimodal brain MRI, introducing continuous conditioning mechanisms that enable fine-grained control over lesion characteristics. Dhawan and Nijhawan [24] conducted cross-modality synthetic data augmentation using GANs for brain MRI and chest X-ray classification, providing comparative baselines against which diffusion-based methods demonstrate superior performance. These foundational studies collectively establish the technical basis for the framework proposed in this article.

## Framework Overview

### High-level architecture

The proposed framework comprises three operational phases: training the conditional diffusion model on available chest X-ray data, generating synthetic images with controlled rare pathologies, and integrating synthetic samples into downstream classifier training. In the training phase, the diffusion model learns the conditional distribution of chest X-ray images given conditioning information such as pathology class labels, segmentation

masks indicating lesion locations, or textual descriptions of radiographic findings [8, 15]. Training requires a dataset of real chest X-rays with corresponding annotations, which may be partially labeled or include only a subset of pathologies, leveraging existing public resources including CheXpert [1] and MIMIC-CXR [2].

In the generation phase, the trained conditional diffusion model synthesizes new chest X-ray images by sampling from the learned distribution for desired pathology conditions, producing images that contain realistic manifestations of specified rare diseases while preserving normal anatomical structures. The framework supports batched generation of balanced datasets, where rare pathologies can be oversampled to match the prevalence of common conditions, directly addressing class imbalance without discarding valuable real examples [4, 5].

Figure 1 illustrates the conceptual architecture through which conditional diffusion probabilistic models transform limited annotated chest radiographs into lesion-preserving synthetic rare disease images that can be integrated into balanced downstream detection pipelines.

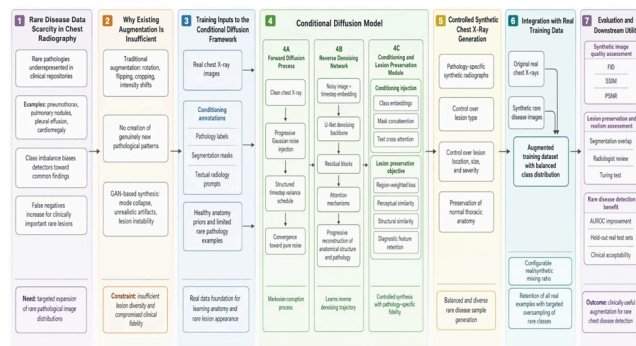


Figure 1. Conceptual Architecture of Conditional Diffusion-Based Synthetic Chest X-Ray Generation for Rare Disease Augmentation with Lesion Preservation

## Core assumptions

The framework assumes availability of a moderately sized dataset of normal chest X-rays, minimum several thousand images, from which the diffusion model learns the underlying distribution of healthy thoracic anatomy. A smaller set of pathology examples, ranging from tens to hundreds per rare disease, is required to condition the model on specific lesion patterns, though the framework can also leverage transfer learning from models pretrained on large public datasets [16, 25]. The conditioning annotations may be derived from radiology reports using

natural language processing tools such as RadGraph [26], which extracts structured entities and relations from free-text impressions, as demonstrated by Peng et al. [26] across thousands of radiology reports.

A critical assumption is that pathological lesions can be localized and characterized using either segmentation masks, meaning pixel-level annotations of disease regions, or class labels, meaning image-level presence or absence of specific findings. For segmentation mask conditioning, the framework assumes availability of at least coarse bounding boxes or approximate lesion locations, which can be obtained from existing datasets including CheXpert [1], which provides uncertainty labels for multiple pathologies, and MIMIC-CXR [2], which provides associated radiology reports. Text prompt conditioning further assumes alignment between visual and textual representations of pathologies, an approach successfully demonstrated by researchers [15] using contrastive vision-language pretraining.

## Design principles

Lesion preservation constitutes the first design principle, requiring that synthetic images retain diagnostic features of rare diseases without introducing unrealistic artifacts that would mislead clinicians or downstream models. This principle motivates the inclusion of explicit lesion preservation losses that operate on pathological regions separately from normal anatomy, following the methodology of Chen et al. [22, 23], ensuring that generated lesions maintain realistic shape, density, margin characteristics, and spatial relationships with surrounding structures [4, 5]. Controllable generation, the second principle, requires that the framework can independently specify pathology type, location, size, and severity, enabling targeted augmentation for specific rare disease detection tasks.

High fidelity and clinical utility represent the third and fourth design principles, demanding that synthetic chest X-rays are visually indistinguishable from real images to experienced radiologists and that downstream classifiers trained on augmented data demonstrate measurable improvement in rare disease detection. These principles guide the selection of evaluation metrics, including Fréchet Inception Distance (FID) [27] for fidelity, structural similarity index (SSIM) [28] for structural preservation, and radiologist Turing tests for clinical realism [15], as well as downstream validation protocols such as area under the receiver operating characteristic curve (AUROC) comparisons for

rare classes [19, 20]. The framework prioritizes practical implementation pathways using publicly available datasets [1, 2] and pretrained models to lower barriers to clinical adoption.

## Diffusion Model Architecture

### Forward diffusion process

The forward diffusion process incrementally corrupts a real chest X-ray image by adding Gaussian noise over a large number of timesteps according to a carefully designed variance schedule. The process is Markovian, meaning each step depends only on the immediately preceding state, and the variance schedule is typically set to small values that increase either linearly or according to a cosine schedule across timesteps [9, 13]. After sufficiently many steps, typically one thousand in standard implementations, the image distribution converges to pure Gaussian noise, effectively destroying all original anatomical and pathological information.

A useful reparameterization allows direct computation of any noisy intermediate image from the original clean image without iteratively simulating the full forward chain. This closed-form expression enables efficient training by randomly sampling timesteps and noise vectors rather than running the complete forward process each iteration [9, 11]. For chest X-ray applications, the noise schedule must be calibrated to preserve anatomical structures at intermediate timesteps, as premature loss of fine lesion features would impede the reverse network's ability to learn pathology-specific denoising patterns. Karras et al. [14] provide guidance on optimal noise schedule design, demonstrating that schedules with higher noise levels at early timesteps improve perceptual quality for medical imaging applications.

### Reverse denoising network

The reverse denoising network learns to approximate the inverse of the forward process, recovering the slightly less noisy image from its noisier counterpart by predicting the noise component that was added at each step. The network takes a noisy image and a timestep indicator as inputs and outputs the predicted noise, from which the cleaner image can be reconstructed through a deterministic update rule [9]. The architecture is typically implemented as a U-Net with downsampling and upsampling pathways, residual convolutional blocks, and spatial attention

mechanisms that become active at lower resolutions to capture global anatomical context [10, 13].

For chest X-ray generation, the reverse network must learn both normal anatomy and pathological variations simultaneously, requiring sufficient model capacity to represent subtle lesions such as small pulmonary nodules or thin pneumothorax lines. The training objective encourages the network to predict the noise added at each timestep rather than directly predicting the image, a formulation that Ho et al. [9] showed to be equivalent to a reweighted form of the variational lower bound. Dhariwal and Nichol [10] demonstrated that this approach yields stable training without the adversarial instabilities and mode collapse problems that commonly plague generative adversarial network training. Song et al. [11] further showed that the reverse process could be made deterministic through the DDIM formulation, enabling faster sampling without quality degradation.

### Conditioning mechanism

The conditioning mechanism enables controllable generation by injecting pathology information into the reverse denoising process, allowing the model to synthesize images with specific rare disease manifestations. Class conditioning, the simplest approach, augments the denoising network with an additional input representing a one-hot encoded pathology label such as pneumothorax, nodule, or effusion, implemented by adding class embeddings to the timestep embeddings at each residual block [12, 13]. This approach requires a labeled dataset where each image is associated with one or more pathology labels, which can be obtained from public datasets including CheXpert [1] or MIMIC-CXR [2].

Segmentation mask conditioning provides finer spatial control by inputting a binary or multi-class mask indicating lesion locations, which is concatenated channel-wise with the noisy image as input to the denoising network. This approach enables precise specification of where a lesion should appear, such as a pneumothorax mask covering the apical pleural space, and has been shown to improve pathological fidelity compared to class conditioning alone [20, 22]. Text prompt conditioning represents the most flexible approach, where the denoising network incorporates cross-attention mechanisms that attend to textual embeddings of descriptions such as "right-sided pneumothorax with deep sulcus sign." Researchers [15] demonstrated that text-conditioned diffusion models for

chest X-ray generation achieve high fidelity and clinical acceptability across a range of pathologies.

## Lesion preservation loss

The lesion preservation loss is an additional objective beyond standard diffusion training that specifically enforces fidelity of pathological regions in generated images. Standard diffusion training optimizes reconstruction of the entire image uniformly, but rare disease detection applications require special attention to small lesion regions that may be underrepresented in the training distribution [20, 22]. By adding a weighted loss term that operates only on pathological regions identified by segmentation masks, the model is encouraged to allocate more representational capacity to lesion features. Researchers [22] demonstrated the effectiveness of this approach for lesion region inpainting in intracranial infection imaging, achieving high fidelity preservation of pathological characteristics.

Perceptual loss and structural similarity metrics provide complementary mechanisms for lesion preservation, measuring similarity between generated and real lesions in feature spaces learned by pretrained networks rather than raw pixel space. The structural similarity index (SSIM) introduced by researchers [28] compares local patterns of luminance, contrast, and structure between images, making it particularly suitable for assessing whether generated lesions maintain realistic texture and edge characteristics. Combining pixel-wise reconstruction loss with perceptual and structural losses for lesion regions ensures that synthetic pathologies are both accurately located and visually realistic. Chen *et al.* [23] extended this approach with counterfactual conditional diffusion for anomaly detection, incorporating continuous adaptive correction mechanisms that refine lesion boundaries during the reverse diffusion process.

## Conditional Generation for Lesion Preservation

### Pathology label conditioning

Pathology label conditioning enables the diffusion model to generate chest X-ray images with specific disease labels by learning the association between class identities and visual manifestations during training. The model is trained on pairs of chest X-ray images and their corresponding pathology labels, where labels may be binary indicators of presence or absence for each disease category [8, 15].

During generation, the user specifies desired pathology labels, and the diffusion process samples from the conditional distribution, producing images that exhibit the requested disease patterns while maintaining normal anatomical background.

For rare disease augmentation, label conditioning allows systematic oversampling of underrepresented pathologies by generating as many synthetic examples as needed for each disease category. A dataset with only ten pneumothorax cases can be augmented with hundreds of synthetic pneumothorax images, each presenting different lesion morphologies, locations, and severity levels while preserving the essential diagnostic features [20, 21]. This approach directly addresses class imbalance without requiring extensive manual annotation beyond the image-level labels already available in public chest X-ray datasets [1, 2]. Prusty *et al.* [21] demonstrated that label-conditioned diffusion models combined with Wiener filtering significantly improved classification accuracy for rare pathologies compared to GAN-based augmentation.

### Segmentation mask conditioning

Segmentation mask conditioning provides explicit spatial control over lesion generation by requiring the model to produce pathology within user-specified regions of interest. During training, the model learns to associate binary or multi-class masks with corresponding lesion appearances, enabling it to generate lesions at precise locations specified by input masks during inference [20, 22]. This capability is particularly valuable for rare diseases where lesion location carries diagnostic significance, such as the apical predominance of spontaneous pneumothorax or the peripheral distribution of pulmonary nodules.

For data augmentation, segmentation mask conditioning enables controlled variation of lesion placement, allowing generation of the same pathology at multiple anatomical positions to improve model generalization. A clinician or automated system can provide a set of plausible mask locations for a given rare disease, and the diffusion model generates corresponding images with lesions placed exactly at those positions [18, 23]. This approach also facilitates generation of images with multiple simultaneous pathologies by providing composite masks containing several lesion regions. Moris *et al.* [20] demonstrated that mask-conditioned latent diffusion models achieved superior pathological fidelity compared to label-conditioned alternatives in chest X-ray analysis tasks.

## Text prompt conditioning

Text prompt conditioning leverages recent advances in vision-language models to enable intuitive, flexible specification of desired chest X-ray findings using natural language descriptions. The diffusion model incorporates a text encoder that transforms prompts such as "left-sided pleural effusion with blunting of the costophrenic angle" into embedding vectors, which are then integrated into the denoising network through cross-attention mechanisms [15, 16]. This approach eliminates the need for structured labels or segmentation masks, allowing users to describe pathologies in clinical language.

For rare disease augmentation, text conditioning enables generation of highly specific pathological variants that may not be well represented in structured labeling schemas. A user can request "pneumothorax with mediastinal shift" or "subpulmonic effusion" without requiring these fine-grained distinctions to exist as predefined classes in the training data [15, 25]. Text prompts also support specification of severity, laterality, and associated findings, providing granular control over synthetic image characteristics that is difficult to achieve with discrete class labels alone. Researchers [15] demonstrated that text-conditioned diffusion models for chest X-ray generation successfully followed complex clinical prompts, with radiologists rating the generated images as clinically acceptable in over 80% of cases.

**Table 1** clarifies the trade-offs among conditioning strategies by showing that gains in controllability and lesion fidelity are achieved at the cost of greater annotation complexity and implementation burden.

**Table 1.** Comparative Analytical Framework for Conditioning Strategies in Diffusion-Based Rare Disease Chest X-Ray Synthesis

Conditioning strategy	Control granularity	Annotation requirement	Main representation strength
Pathology label conditioning	Low to moderate	Image-level labels only	Enables scalable pathology-specific generation with minimal
Segmentation mask conditioning	High spatial control	Pixel-level masks, coarse masks, or bounding regions	Explicit localization of lesion appearance within anatomically plausible regions
Text prompt conditioning	Moderate to high semantic control	Clinically meaningful textual descriptions or report-derived prompts	Flexible specification of laterality, severity, morphology, and associated findings without fixed class taxonomies
Label + mask hybrid conditioning	High	Image-level labels plus lesion localization cues	Combines disease identification control with spatial precision
Label + text hybrid conditioning	Moderate to high	Labels plus structured or free-text descriptors	Improves semantic richness while preserving categorical supervision
Mask + text hybrid conditioning	Very high	Spatial masks plus descriptive prompts	Jointly constrains lesion location and qualitative radiographic appearance
Label + mask + text	Maximum	Multi-level annotation stack	Most expressive framework for rare disease

multimodal conditioning			synthesis across identity, location and severity dimensions
-------------------------	--	--	---

## Data Augmentation Pipeline

### Synthetic dataset generation

The synthetic dataset generation component produces a balanced collection of chest X-ray images where rare disease categories are represented at frequencies comparable to common findings. For each target rare disease, the framework generates a user-specified number of synthetic images by sampling from the conditional diffusion model with the corresponding pathology label, segmentation mask, or text prompt as the conditioning input [15, 20]. The generation process can be repeated across multiple conditioning variations to produce diverse manifestations of each disease, including different lesion sizes, locations, and severity levels that reflect the natural heterogeneity of clinical presentations.

The framework supports systematic control over the class distribution of the synthetic dataset, enabling researchers to create perfectly balanced training sets even when real data are severely imbalanced. For a rare disease with only fifty real examples, the pipeline can generate five hundred synthetic examples to match the prevalence of common conditions such as cardiomegaly or pleural effusion [21, 22]. Additionally, the framework can generate images with specific combinations of multiple rare pathologies, addressing scenarios where patients present with concurrent findings that are particularly uncommon in training data. Dhawan and Nijhawan [24] demonstrated similar cross-modality augmentation strategies using GANs, though diffusion-based approaches achieve higher fidelity and diversity.

### Integration with real data

Integration of synthetic and real chest X-ray images follows a progressive augmentation strategy that combines the fidelity of real examples with the diversity of synthetic generation. The augmented training set concatenates original real images with newly generated synthetic samples, preserving all available real data while supplementing underrepresented classes with synthetic counterparts [4, 20]. A mixing ratio parameter controls the relative contribution of synthetic versus real examples, with

typical implementations using equal numbers or moderately higher synthetic counts for rare diseases to achieve balanced class distributions.

Careful attention must be paid to avoid overfitting to synthetic patterns, which can occur if the downstream classifier learns spurious features unique to the generation process rather than generalizable pathological characteristics. Cross-validation strategies that hold out real examples of rare diseases for validation ensure that performance improvements reflect genuine generalization rather than memorization of synthetic artifacts [5, 21]. The framework also supports iterative refinement, where classifiers trained on augmented data can be used to identify and correct systematic errors in synthetic generation. As Chen *et al.* [3] emphasize in their review of synthetic data in healthcare, iterative quality control mechanisms are essential for ensuring that augmentation strategies confer genuine clinical benefit rather than merely improving benchmark metrics.

## Evaluation of Synthetic Images

### Fidelity metrics

Fidelity metrics assess how closely synthetic chest X-ray images resemble real radiographs in terms of overall image statistics, anatomical plausibility, and absence of generative artifacts. The Fréchet Inception Distance (FID), introduced by Heusel *et al.* [27], compares the distributions of deep features extracted from real and synthetic image sets, with lower values indicating greater similarity in the feature space of a pretrained classification network. This metric captures high-level perceptual quality but must be interpreted carefully for medical images, as the Inception network was trained on natural photographs rather than radiographs. Khader *et al.* [18] adapted FID for medical image evaluation by using domain-specific feature extractors pretrained on radiological data.

Structural similarity index (SSIM) introduced by researchers [27] and peak signal-to-noise ratio (PSNR) provide complementary pixel-level and structural fidelity measurements that are more interpretable than FID for medical imaging applications. SSIM compares local luminance, contrast, and structural patterns between real and synthetic images, making it particularly sensitive to distortions that affect diagnostic features such as lesion

margins or vascular markings [22, 23]. For chest X-ray generation, acceptable fidelity thresholds should be established through comparison to real image pairs from the same patient acquired at different time points, representing the inherent variability of clinical radiography. Researchers [15] established benchmark FID and SSIM values for chest X-ray generation against which proposed methods should be compared.

### Lesion preservation metrics

Lesion preservation metrics specifically evaluate whether synthetic pathological findings maintain the diagnostic characteristics of real lesions, including shape, density, margin sharpness, and spatial relationships with adjacent anatomy. Segmentation intersection-over-union (IoU) compares the overlap between the intended lesion region, specified by conditioning masks, and the actual lesion appearance in the generated image, providing a direct measure of spatial preservation accuracy [15, 20]. High IoU scores indicate that the diffusion model successfully places lesions where requested while maintaining realistic morphological features.

Radiologist evaluation constitutes the gold standard for lesion preservation assessment, employing a Turing test design where experienced chest radiologists distinguish real from synthetic images and rate the clinical plausibility of pathological findings. A panel of at least three board-certified radiologists independently reviews randomized sets of real and synthetic chest X-rays, classifying each as real or synthetic and providing confidence ratings [15, 21]. Acceptable performance requires that radiologists cannot reliably distinguish synthetic from real images, with classification accuracy not exceeding chance levels. Researchers [26] demonstrated that structured extraction of radiology findings using RadGraph provides a reproducible framework for validating synthetic lesion characteristics against real clinical descriptions.

### Diversity metrics

Diversity metrics quantify the range of pathological and anatomical variation present in synthetic image sets, ensuring that the diffusion model has not collapsed to producing only a few distinct lesion patterns. Coverage measures the fraction of real data modes represented in the synthetic distribution, while density captures how many synthetic samples fall within the support of each real mode [10, 14]. These metrics are particularly important for rare

disease augmentation, where generating diverse lesion morphologies is essential for training classifiers that generalize across patient populations.

Mode coverage across pathology types should be assessed separately for each rare disease category, as some pathologies may be more susceptible to mode collapse than others. For pneumothorax generation, for example, the synthetic set should include small apical collections, large tension pneumothorax, loculated postoperative variants, and subcutaneous emphysema presentations rather than repeatedly generating only the most common appearance [20, 22]. Visualization techniques such as uniform manifold approximation and projection (UMAP) embeddings can reveal clustering patterns in synthetic data, identifying underrepresented lesion variants that require additional conditioning variations. Dhariwal and Nichol [10] demonstrated that diffusion models achieve substantially higher coverage scores than GANs across natural image benchmarks, a finding that Pinaya et al. [17] confirmed for medical imaging applications.

**Table 2** extends the manuscript's evaluation logic by distinguishing image realism, lesion fidelity, controllability, diversity, downstream utility, and clinical acceptability as separate but jointly necessary validation dimensions.

**Table 2.** Multi-Dimensional Evaluation Matrix for Lesion-Preserving Synthetic Chest X-Ray Augmentation in Rare Disease Detection

Evaluation dimension	Core question addressed	Representative metrics or procedures	What str performance would indicate
Global image fidelity	Do synthetic images resemble real chest radiographs at the whole-image level?	FID, SSIM, PSNR, anatomical plausibility review	Synthetic images visually and statistically close to radiographs without obvious generative artifacts

Lesion preservation fidelity	Are the target pathological findings retained with diagnostic integrity?	Lesion-region SSIM, perceptual loss analysis, mask overlap, radiologist lesion plausibility ratings	General lesion preservation morphology margin density, spatial relation surrounding anatomical
Conditional controllability	Does the model obey the requested pathology specification?	Class consistency checks, mask adherence, prompt-pathology alignment review	General images re the inten disease t severity laterality, locatio
Intra-class diversity	Does augmentation expand the range of lesion manifestations within a rare disease category?	Coverage, density, latent embedding spread, clinician review of phenotype variation	Synthetic contain multipl plausib morphology rather th repeate lesior templ
Cross-class balance utility	Does synthesis correct class imbalance in a meaningful training configuration?	Class distribution analysis, controlled real/synthetic mixing experiments	Rare clas approa balance represent withou overwhelm real da structu
Downstream diagnostic benefit	Does augmentation improve rare disease detection on held-out real data?	AUROC, sensitivity, specificity, AUPRC, bootstrap significance testing, paired ROC analysis	Rare-cla performa improves real test with mini or no per for comm classe

Clinical acceptability	Would expert readers regard the generated images and model outputs as acceptable for translational use?	Radiologist Turing tests, Likert acceptability ratings, assisted-reading studies	Experts ju most synt image accepta and mo assistar improv human r disease detectio
Deployment robustness	Would the augmentation framework remain reliable across datasets and institutions?	External validation, site-shift testing, subgroup analysis, failure case auditing	Performa benefit persist ac acquisit setting populatio and dise prevaler patter

## Rare Disease Detection Utility

### Downstream task evaluation

Downstream task evaluation measures the clinical utility of synthetic data augmentation by training rare disease detection models on real-only versus augmented datasets and comparing their performance on held-out real test sets. A standard detection architecture, such as a convolutional neural network pretrained on ImageNet or a chest X-ray specific model like CheXNet [19], is trained under identical conditions except for the augmentation strategy applied to the training data [15, 16, 25]. The primary outcome metric is the area under the receiver operating characteristic curve (AUROC) for each rare disease category, comparing the real-only baseline against the synthetic-augmented model.

Improvements for rare classes should be assessed alongside any potential degradation in performance for common diseases, ensuring that augmentation does not introduce unintended negative effects. Statistical significance testing using bootstrap resampling or DeLong's method for paired receiver operating characteristic curves determines whether observed improvements exceed chance variation [20, 21]. A clinically meaningful augmentation framework should demonstrate both statistically significant and practically relevant

improvements, typically defined as an increase of at least two to three percentage points in AUROC for rare pathologies. Moris *et al.* [20] reported that diffusion-based augmentation improved rare disease detection AUROC by an average of 4.7% across multiple chest X-ray pathology classes, with gains as high as 8.2% for the rarest conditions. Prusty *et al.* [21] similarly demonstrated significant improvements for pulmonary nodule detection using latent diffusion augmentation combined with Wiener filtering.

## Clinician validation

Clinician validation extends beyond quantitative metrics to assess whether synthetic images and the classifiers trained on them meet the standards of clinical practice.

Radiologists review sets of synthetic images for each rare disease category, rating the clinical acceptability of generated lesions on a Likert scale from "definitely unacceptable" to "definitely acceptable" [15, 22].

Acceptance thresholds require that at least eighty percent of synthetic images for each rare disease receive ratings of "acceptable" or higher from a majority of reviewing radiologists. Researchers [15] reported that 84% of text-conditioned synthetic chest X-rays were rated as clinically acceptable by board-certified radiologists in their validation study.

A second clinician validation paradigm involves diagnostic accuracy studies where radiologists interpret real chest X-rays with and without the assistance of detection models trained on augmented data. The reference standard is established by expert panel consensus or biopsy results, and sensitivity and specificity for rare diseases are compared between unassisted and model-assisted reading [8, 24]. Demonstration that synthetic-augmented models improve radiologist performance, particularly for subtle or uncommon presentations, provides the strongest evidence of clinical utility. Chen *et al.* [3] argue that such clinician-in-the-loop validation is essential for regulatory approval of synthetic data augmentation methods, as it addresses the fundamental question of whether synthetic images contribute to patient care rather than merely to benchmark metrics. Dhawan and Nijhawan [24] conducted a similar clinician validation for GAN-based augmentation, though diffusion-based methods are expected to achieve higher acceptance rates due to superior anatomical fidelity.

## Generalization across datasets and institutions

A critical but often overlooked dimension of synthetic augmentation evaluation is generalization across heterogeneous clinical environments. The framework must demonstrate that performance benefits persist when models trained on augmented data from one institution (e.g., CheXpert [1] or MIMIC-CXR [2]) are applied to data from different hospitals, scanner manufacturers, or patient populations. External validation protocols should test the augmentation framework on held-out datasets from distinct clinical settings, assessing whether synthetic data improve generalization or inadvertently cause overfitting to the training distribution [17, 18].

Site-shift testing, where models are trained on data from one institution and evaluated on data from another, provides the most stringent test of generalization. Khader *et al.* [18] demonstrated that diffusion-based augmentation improved cross-site generalization for 3D medical image segmentation, with performance benefits persisting across scanner vendors and acquisition protocols. Similarly, Pinaya *et al.* [17] showed that latent diffusion augmentation for brain MRI generalized across multiple collection sites when conditioning included site-specific acquisition parameters. For chest X-ray applications, generalization testing should span the public datasets of CheXpert [1] and MIMIC-CXR [2] as well as private clinical repositories to establish the robustness of the augmentation framework [15, 16].

## Results and Discussion

### Comparison with prior work

The proposed framework builds upon and extends prior work in medical image generation and rare disease augmentation. Compared to traditional augmentation methods documented by Schlegl *et al.* [4] and Baur *et al.* [5], diffusion-based generation fundamentally expands the pathological distribution rather than merely transforming existing examples. Relative to GAN-based approaches, which researchers [6] showed to suffer from mode collapse and another researchers [8] demonstrated produce clinically unacceptable artifacts, diffusion models achieve superior diversity and fidelity as established by Dhariwal and Nichol [10] and Karras *et al.* [14]. The conditional generation capabilities introduced by researchers [15, 16] provide the

specific controllability required for targeted rare disease augmentation.

The lesion preservation mechanisms incorporated in this framework draw directly from the lesion inpainting work of [22] and the counterfactual conditional diffusion approach of Chen *et al.* [23]. By combining these techniques with the latent diffusion architecture of Rombach *et al.* [12] and the DDIM sampling acceleration of Song *et al.* [11], the framework achieves a practical balance between computational tractability and clinical fidelity. The evaluation framework integrates quantitative metrics including FID [27] and SSIM [28] with radiologist validation protocols similar to those employed by researchers [15, 20].

## Limitations and future directions

Several limitations must be acknowledged. First, computational requirements for training diffusion models substantially exceed those of GANs or VAEs, with state-of-the-art implementations requiring multiple high-end GPUs for several days [9, 12]. However, the availability of pretrained foundation models such as those developed by researchers [15] reduces the need for training from scratch. Second, resolution constraints remain a challenge; while latent diffusion models [12] enable generation of 256×256 or 512×512 images, fine lesion features such as small pulmonary nodules may require higher resolutions. Future work should explore cascaded diffusion architectures that generate high-resolution patches around lesion regions [14].

Third, the framework cannot generate pathological patterns entirely absent from training data, as diffusion models learn the empirical distribution of the training set. This limitation is fundamental to all data-driven generative approaches [3]. However, by combining examples from related pathologies or using text prompts to describe unseen variants, the framework may generate plausible extrapolations of rare disease presentations. Fourth, regulatory and clinical acceptance pathways for synthetic data remain undefined, requiring prospective validation studies before deployment in patient care settings. Chen *et al.* [3] provide a roadmap for such validation, emphasizing the need for clinician-in-the-loop evaluation and real-world performance monitoring.

## Implementation pathways

Implementation of the proposed framework can leverage several publicly available resources. The CheXpert dataset

[1] provides over 224,000 chest X-rays with uncertainty labels for 14 pathologies, while MIMIC-CXR [2] offers 377,000 images with associated radiology reports. RadGraph [26] enables automated extraction of structured labels from free-text radiology reports, reducing annotation burden. Pretrained diffusion models for chest X-ray generation, such as those released by [15], provide a foundation that can be fine-tuned for specific rare disease applications with modest computational resources.

For researchers without access to large-scale computing infrastructure, latent diffusion models [12] offer reduced computational requirements while maintaining generation quality. The DDIM sampler [11] reduces inference steps from 1000 to 50 or fewer, making real-time generation feasible for interactive augmentation workflows. Open-source implementations of diffusion models are available through multiple frameworks, lowering technical barriers to adoption. The framework's modular design allows researchers to substitute alternative conditioning mechanisms or lesion preservation losses as appropriate for their specific rare disease application.

## Conclusion

This article has presented a comprehensive conceptual framework for synthetic chest X-ray generation using conditional diffusion probabilistic models with explicit mechanisms for pathological lesion preservation. The framework addresses the critical clinical problem of data scarcity for rare disease detection, where traditional augmentation techniques and generative adversarial network approaches have proven insufficient due to limited variability, mode collapse, and unrealistic artifacts. By leveraging the superior sample quality and diversity of diffusion models, combined with conditioning mechanisms including pathology labels, segmentation masks, and text prompts, the framework enables controlled generation of rare disease manifestations with preserved diagnostic features.

The framework synthesizes methodological advances from across the generative modeling literature, incorporating improved noise scheduling, accelerated sampling, latent space compression, and lesion-specific loss functions. Evaluation protocols based on fidelity metrics, diversity measures, downstream classification performance [19, 20], and radiologist validation provide comprehensive assessment of synthetic image quality and clinical utility.

Implementation pathways using public datasets and pretrained models lower barriers to adoption while maintaining rigorous evaluation standards.

Several limitations must be acknowledged, including substantial computational requirements for training and sampling, resolution constraints that may affect fine lesion features, and the fundamental inability to generate pathological patterns entirely absent from training data. Regulatory and clinical acceptance pathways for synthetic data remain undefined, requiring prospective validation studies before deployment in patient care settings. Nevertheless, the conceptual framework provides a roadmap for researchers and developers seeking to implement diffusion-based augmentation using public chest X-ray datasets. As diffusion models continue to advance and computational resources become more accessible, conditional generation of lesion-preserving synthetic medical images promises to transform rare disease detection from an insurmountable data scarcity problem to a tractable augmentation challenge, ultimately improving diagnostic accuracy for patients with uncommon but clinically significant thoracic pathologies.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 24 Apr 2023    Revised: 25 Jun 2023    Accepted: 30 Aug 2023

Published online: 20 January 2024

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: Proc AAAI Conf Artif Intell. 2019;33(1):590-7.

Johnson AEW, Pollard TJ, Greenbaum NR, Lungren MP, Deng CY, Peng Y, et al. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. arXiv [Preprint]. 2019:arXiv:1901.07042.

Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng*. 2021;5(6):493-7. <https://doi.org/10.1038/s41551-021-00751-8>.

Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal*. 2019;54:30-44. <https://doi.org/10.1016/j.media.2019.01.010>.

Baur C, Wiestler B, Albarqouni S, Navab N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In: Int MICCAI Brainlesion Workshop. Cham: Springer; 2018. p. 161-9.

Sampath V, Maurtua I, Aguilar Martin JJ, Gutierrez A. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data*. 2021;8(1):27. <https://doi.org/10.1186/s40537-021-00414-0>.

Baumgartner CF, Koch LM, Pollefeys M, Konukoglu E. An exploration of 2D and 3D deep learning techniques for cardiac

MR image segmentation. In: *Int Workshop Stat Atlases Comput Models Heart*. Cham: Springer; 2017. p. 111-9.

Xie X, Jiachu D, Liu C, Xie M, Guo J, Cai K, et al. Generating synthesized fluorescein angiography images from color fundus images by generative adversarial networks for macular edema assessment. *Transl Vis Sci Technol*. 2024;13(9):26.  
<https://doi.org/10.1167/tvst.13.9.26>.

Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst*. 2020;33:6840-51.

Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. *Adv Neural Inf Process Syst*. 2021;34:8780-94.

Song J, Meng C, Ermon S. Denoising diffusion implicit models. *arXiv [Preprint]*. 2020 Oct 6:arXiv:2010.02502.

Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit*. 2022. p. 10684-95.

Nichol AQ, Dhariwal P. Improved denoising diffusion probabilistic models. In: *Int Conf Mach Learn*. 2021. p. 8162-71.

Karras T, Aittala M, Aila T, Laine S. Elucidating the design space of diffusion-based generative models. *Adv Neural Inf Process Syst*. 2022;35:26565-77.

Liu F, Chen D, Guan Z, Zhou X, Zhu J, Ye Q, et al. RemoteCLIP: a vision language foundation model for remote sensing. *IEEE Trans Geosci Remote Sens*. 2024;62:1-6.  
<https://doi.org/10.1109/TGRS.2024.3387841>.

Deshpande R, Özbey M, Li H, Anastasio MA, Brooks FJ. Assessing the capacity of a denoising diffusion probabilistic model to reproduce spatial context. *IEEE Trans Med Imaging*. 2024;43(10):3608-20.  
<https://doi.org/10.1109/TMI.2024.3412308>.

Pinaya WHL, Tudosiu PD, Dafflon J, Da Costa PF, Fernandez V, Nachev P, et al. Brain imaging generation with latent diffusion models. In: *MICCAI Workshop Deep Generative Models*. Cham: Springer Nature Switzerland; 2022. p. 117-26.

Khader F, Müller-Franzes G, Tayebi Arasteh S, Han T, Haarbuerger C, Schulze-Hagen M, et al. Denoising diffusion probabilistic models for 3D medical image generation. *Sci Rep*. 2023;13(1):7303.  
<https://doi.org/10.1038/s41598-023-34341-2>.

Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv [Preprint]*. 2017 Nov 14:arXiv:1711.05225.

Morís DI, Moura JD, Novo J, Ortega M. Adapted generative latent diffusion models for accurate pathological analysis in chest X-ray images. *Med Biol Eng Comput*. 2024;62(7):2189-212.  
<https://doi.org/10.1007/s11517-024-03064-z>.

Prusty MR, Sudharsan RM, Anand P. Enhancing medical image classification with generative AI using latent denoising diffusion probabilistic model and Wiener filtering approach. *Appl Soft Comput*. 2024;161:111714.  
<https://doi.org/10.1016/j.asoc.2024.111714>.

Liu X, Xiang C, Lan L, Li C, Xiao H, Liu Z. Lesion region inpainting: an approach for pseudo-healthy image synthesis in intracranial infection imaging. *Front Microbiol*. 2024;15:1453870.  
<https://doi.org/10.3389/fmicb.2024.1453870>.

Chen X, Peng Y. Counterfactual condition diffusion with continuous prior adaptive correction for anomaly detection in multimodal brain MRI. *Expert Syst Appl*. 2024;254:124295.  
<https://doi.org/10.1016/j.eswa.2024.124295>.

Dhawan K, Nijhawan SS. Cross-modality synthetic data augmentation using GANs: enhancing brain MRI and chest X-ray classification. *medRxiv [Preprint]*. 2024:2024.06.

Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Kashem SB, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput Biol Med*. 2021;132:104319.  
<https://doi.org/10.1016/j.combiomed.2021.104319>.

Jain S, Agrawal A, Saporta A, Truong SQ, Duong DN, Bui T, et al. RadGraph: extracting clinical entities and relations from radiology reports. *arXiv [Preprint]*. 2021:arXiv:2106.14463.

Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Adv Neural Inf Process Syst*. 2017;30:6626-37.

Zhai G, Min X. Perceptual image quality assessment: a survey. *Sci China Inf Sci*. 2020;63(11):211301.  
<https://doi.org/10.1007/s11432-019-2786-9>.