Inspire Health Journal
ISSN 3110-9438, Vol. 1, No. 2, 2026
doi.org/10.65718/inspireHealth.2026.2011
Inspire Academic Press (Canada)

Research Paper OPEN ACCESS

Enhancing COVID-19 Diagnosis: Leveraging GAN-Based Image Augmentation and Deep Learning for Improved Chest X-ray Classification

Anupam Agrawal [1], Asadi Srinivasulu [2,*], Ekeshwar Aditya [1], Shreya Bharti [1]

[1] Information Technology Department, International Institute of Information Technology, Prayagraj, India
[2] crcCARE, University of Newcastle, Callaghan, NSW 2308, Australia
[*] Corresponding Author: srinivasulu.asadi@newcastle.edu.au

Abstract

The SARS-CoV-2 virus caused COVID-19, which first appeared in late 2019 and quickly spread around the world, with more than 4 million confirmed cases and 286,000 deaths reported by May 2020. It was important to make quick and accurate diagnoses, but regular RT-PCR tests take a long time and a lot of resources. Deep learning models, especially Convolutional Neural Networks (CNNs), have shown promise in finding COVID-19 in chest X-rays. However, their effectiveness is compromised by the scarcity of annotated datasets; certain public datasets contain fewer than 2,000 COVID-positive images, resulting in overfitting and inadequate generalization. To fix this, we use Generative Adversarial Networks (GANs) to add fake images to the training data. However, visual and quantitative analysis (e.g., SSIM and PSNR) reveal that not all GAN outputs are of diagnostic quality. Training CNNs on unfiltered synthetic data can degrade performance. Therefore, this study introduces a filter-ing mechanism to retain only high- quality synthetic images for training, enhanc-ing model reliability and accuracy. This approach demonstrates the potential of filtered GAN augmentation to overcome data scarcity and improve deep learning models for medical diagnosis.

Keywords: COVID-19 Detection; Generative Adversarial Networks (GANs); Deep Learning; Data Augmentation; Chest X-ray.

1. Introduction

The COVID-19 pandemic, which started in Wuhan in late 2019, put a lot of stress on healthcare systems around the world. This showed how important it is to have fast and scalable diagnostic tools. RT-PCR is still the most common test, but its problems, such as slow results and false negatives, have led to interest in medical imaging methods like chest X-rays (CXR) for early detection [1]. Convolutional Neural Networks (CNNs) and other deep learning models have shown promise for analyzing these kinds of images. However, limited access to annotated COVID-19 imaging datasets hampers their performance, frequently leading to overfitting [2].

There have been a few reasons why such data has not been widely available during COVID-19, such as privacy laws, uneven data sharing, and the need for experts to annotate imaging studies [3]. Because of this, many COVID-

Academic Editor:
Jaafar Alghazo

Received: 17/10/2025
Revised: 21/12/2025
Accepted: 29/01/2026
Published: 18/03/2026

Citation
Agrawal, A., Aditya, A. S. E., & Bharti, S. (2026). Enhancing COVID-19 Diagnosis: Leveraging GAN-Based Image Augmentation and Deep Learning for Improved Chest X-ray Classification. *Inspire health Journal*, 1(2), 120-129.

19 imaging datasets are small and not evenly distributed, which makes it more likely that deep models will overfit. When CNNs are given only a few samples, they are more likely to memorize the training data than to learn generalizable patterns. This is because they have millions of trainable parameters.

Figure 1 shows a chest X-ray of a patient who tested positive for COVID-19. This X-ray was taken from the training dataset and was used as an example input for training the deep learning models that help with diagnosis.

Basic augmentation methods like rotation, flipping, and scaling only make small changes and don't do much to fix the problem of not having enough data [4]. To solve this problem, we combine synthetic image generation using Progressive Growing GAN (PG-GAN) with traditional methods and a quality filtering process to make a training dataset that is more varied and reliable [5].

An important part of our method is an AI-powered filtering system that selects only high-quality synthetic images to be used for training the model, helping to improve its accuracy and reliability. This pipeline uses autoencoders to reconstruct images and calculates Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to assess image fidelity [6]. A dynamic thresholding mecha-nism filters out low quality outputs before feeding the refined dataset to the CNN classifier on synthetic data, and the image filtering significantly improves model ac-curacy and generalization, paving the way for more robust AI-driven COVID-19 diagnosis [7].

Our work contributes in three key areas: (i) demonstrating the effectiveness of GAN-generated images in mitigating data limitations in COVID-19 datasets (ii) quantifying the improvement in CNN performance when augmented with syn-thetic data, and (iii) image filtering significantly improves model accuracy and gen-eralization, paving the way for more robust AI-driven COVID-19 diagnosis.

Figure 1. Chest X-ray of a COVID-19 Positive Patient from Training Dataset [16].

2. Related Work

Waheed et al. [8] were some of the first to do this kind of research. They came up with CovidGAN, an Auxiliary Classifier GAN (AC-GAN) model that makes fake chest X-ray images of COVID-19 patients. Their method showed that adding GAN-generated data to the training pipeline of CNN classifiers made performance metrics like accuracy, F1 score, and SSIM much better. But their model was limited by the specific dataset and the fact that the images it made weren't very good, which made people worry about how well it would work in the real world.

Ali et al. [9] proposed a Wasserstein GAN (WGAN) methodology to address challenges associated with data scarcity and instability in GAN training. WGANs, which are known for their more stable loss functions and ability to make high-quality images, were shown to lower training variance and increase data diversity. Nonetheless, issues like class imbalance, training time, and computational overhead remained, showing that using GANs on a large scale has its limits in practice.

Ali and Shah [10] provided a more comprehensive viewpoint in their scoping review regarding AI-enhanced GAN pipelines for COVID-19 diagnosis. They pointed out that there aren't any standardized augmentation protocols, federated learning models for privacy-preserving collaboration, or CNNs that tend to overfit to synthetic features. Their review suggested that combining GANs with traditional augmentation and self-supervised learning methods could make models more stable.

Several studies have investigated innovative GAN architectures and hybrid models to tackle particular diagnostic issues. Fedoruk et al. [11] compared augmentation based on Style-GAN2 to traditional methods, showing that the synthetic images were of higher quality and the classification accuracy was higher [12]. These methods show that GAN-based systems could improve medical imaging pipelines, but they also show how important it is to do thorough benchmarking and validation by outside sources.

A review of 25 recent peer-reviewed studies from literature reviews and top publishers like IEEE, Elsevier, and Springer shows that GAN-based data augmentation, deep learning classification, and transfer learning are all coming together [13]. These techniques have mainly utilized chest X-rays and CT scans obtained from open-access datasets. The most popular tools were TensorFlow, PyTorch, and Keras. Performance was measured using a mix of quantitative metrics, with Accuracy, AUC, SSIM, F1-score, and IoU being the most important [14].

In conclusion, GANs have greatly helped to reduce the lack of data and improve the performance of CNN-based diagnostics, but to achieve strong generalization and real-world use, we still need to deal with synthetic bias and computational problems [15].

Recent research has underscored the increasing significance of generative models in improving medical image analysis, especially for chest X-ray datasets characterized by limited and imbalanced samples. Karimi et al. [16] suggested a GAN-based framework for detecting out-of-distribution chest radiographs by modeling latent feature distributions, showing greater robustness when dealing with unseen data. Their research underscores the necessity of aligning synthetic data distributions with actual data to prevent the deterioration of classifier performance. Alam et al. [17] also came up with a multiscale attention-based generative model for making chest X-ray lesions. They showed that attention mechanisms make the images look more real and anatomically consistent. Their results show that adding structural awareness to generative models makes synthetic data more useful in the clinic and makes diagnostic tasks easier. Chen et al. [18] recently looked into generative models in low-data situations. They showed that adding high-quality synthetic images to datasets can greatly improve classification performance and cut down on overfitting in deep learning models. These studies indicate that contemporary generative methodologies, especially GAN-based frameworks, are essential for mitigating data scarcity, augmenting dataset diversity, and enhancing the generalization efficacy of medical image classification systems.

3. Research Methodology

To address the identified challenges, this research implements a multi-stage augmentation and evaluation pipeline. Figure 2 shows the overall workflow of the PG-GAN-based augmentation and filtering framework. It shows how synthetic chest X-ray images are made, checked for quality, and then added to the training dataset in a way that improves the model's accuracy and robustness. The proposed methodology consists of the following components.

3.1. Experimental Dataset

This study utilizes a publicly accessible chest X-ray dataset sourced from Kaggle, consisting of 6,432 images classified into three categories: COVID-19 (576 images), Pneumonia (4,273 images), and Normal (1,583 images) [19]. The dataset is split into training and testing folders, and each folder has subfolders for each class to make supervised learning easier.

3.2 Data Augmentation and Filtering

Conventional augmentation techniques, including rotation, horizontal flipping, cropping, and scaling, are applied to increase dataset variability and improve model generalization. To ensure the quality of augmented data, a two-stage filtering mechanism is introduced with image quality filtering based on Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) thresholds are applied to eliminate low-quality images with poor structural integrity or excessive noise. We also introduced feature consistency filtering as an encoder-based consistency check is employed to identify and remove images with abnormal feature embeddings, thereby preserving anatomical realism and clinical relevance.

3.3 Baseline Model Development

To set a standard for performance, a baseline Convolutional Neural Network (CNN) model is trained and tested on the original dataset without any changes. This step sets a standard for judging how well future augmentation strategies work by testing a standard deep learning model on data that is limited and unbalanced.

3.3 GAN-based Data Augmentation

To fix the problem of class imbalance and the fact that there aren't many COVID-19 and pneumonia images available, a Generative Adversarial Network (GAN)-based augmentation method is used. The goal is to make high-quality fake images that add variety to the dataset and make classification work better.

For making images, we use Progressive Growing of GANs (PG-GAN). PG-GAN starts training with low-resolution images and adds network layers to gradually raise the resolution. This gradual learning method keeps the training process stable and makes it possible to make chest X-ray images that are very clear and realistic.

Figure 2 gives a complete picture of the PG-GAN-based augmentation and filtering framework. It shows how synthetic chest X-ray images are made, checked for quality, and then added to the training dataset in a selective way to make the model more accurate and reliable. Progressive Growing of GANs (PG-GAN) is different from regular Generative Adversarial Networks (GANs) because it uses a gradual training strategy in which both the generator and discriminator networks grow over time. This method lets the model learn features from coarse to fine over time, which makes training more stable, mode collapse less likely, and image fidelity better.

In the proposed framework, PG-GAN learns from a dataset that is balanced across classes and was made using traditional augmentation methods like rotation, flipping, and changing the contrast. This step of pre-balancing makes sure that each class is better represented and helps create more varied and realistic synthetic samples. The progressive learning mechanism makes convergence more stable and creates high-quality chest X-ray images that look very much like real data. To make the dataset even stronger, a combined dataset strategy is used. The final training set is made up of original images, traditionally augmented samples, and filtered PG-GAN-generated images. Deep learning models, specifically ResNet50 and InceptionV3, are then trained on this combined dataset. The results are then systematically compared to those from earlier experiments to find the best way to add more data.

To make sure that the generated and augmented data is reliable, an AI-based filtering layer is added as an important preprocessing step. The main goal of this filtering system is to automatically check the quality of images before they are used to train classification models. This way, only images that are structurally meaningful and diagnostically useful are kept. This process cuts down on noise and stops low-quality or unrealistic samples from hurting how well the model works.

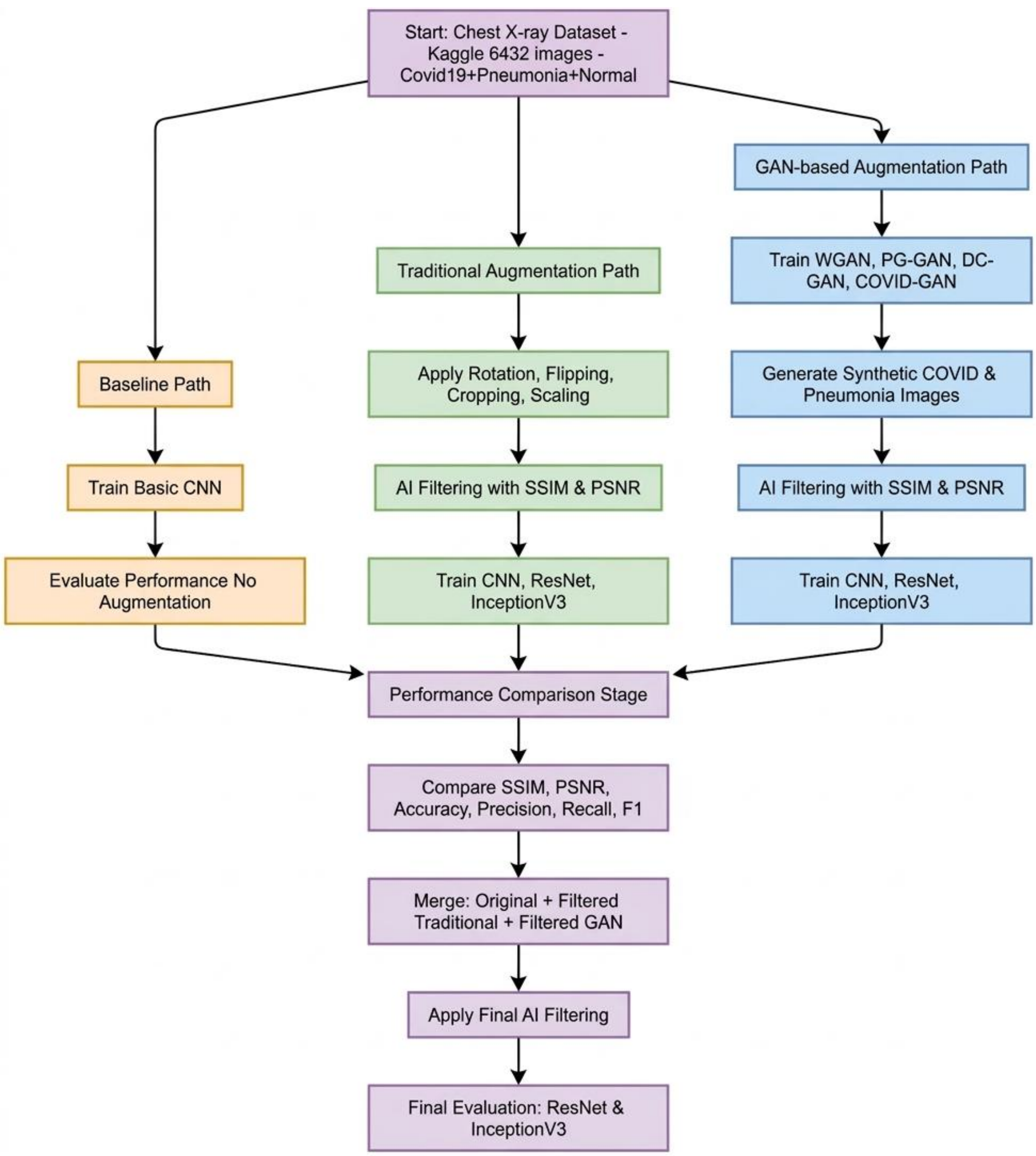

Figure 2. PG-GAN Augmentation and Filtering Workflow

The first step in the filtering process is to choose a threshold based on two common measures of image quality: the Structural Similarity Index Measure (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). SSIM checks how similar the structure of an image is to its reference, while PSNR checks how clear and noisy the signal is. To find the right thresholds, all of the training images go through a metric evaluation process. In this process, SSIM and PSNR values are calculated against a lightly smoothed version of each image, which acts as a fake reference. Using histogram visualization, we can see how these values are spread out. The SSIM values are mostly between 0.25 and 0.45, with a few lower-quality outliers. The PSNR values, on the other hand, form two distinct groups that correspond to low- and high-quality images. Based on this empirical analysis, threshold values are chosen to find a balance between keeping data and quality control. This makes sure that only images with good structural integrity and noise levels are kept.

Along with filtering based on metrics, a second phase of structural-aware filtering uses an encoder-based consistency check. A lightweight convolutional encoder is made to learn small feature representations of real chest X-ray images. The encoder has several convolutional layers, then batch normalization and ReLU activation, and finally a global average pooling layer to make embedding vectors of a fixed length. Cosine similarity is used to compare the feature embedding of each candidate image to the mean embedding of a set of real images. Images that do not meet a certain similarity threshold (0.80) are thrown away because they are structurally inconsistent. This method works well with SSIM and PSNR because it captures global anatomical structures and gets rid of images that are not realistic or medically plausible, which pixel-level metrics alone might not be able to do.

As a useful preprocessing step, the whole AI-based filtering pipeline is set up. If necessary, each image is first changed to black and white, then changed to a [0,1] range, and finally resized to a resolution of 128×128 pixels. After that, SSIM and PSNR values are calculated, and images are only accepted if they meet the requirements of SSIM > 0.22 and PSNR > 11. Images that don't meet these standards are thrown away. This automated filtering system keeps most of the high-quality samples and gets rid of images that are clearly degraded or noisy. This cuts down on the need for manual inspection and makes the dataset more consistent. The improved dataset should make the classification models better at generalizing.

InceptionV3 and ResNet50 are two of the best deep learning architectures for classification because they have been shown to work well for analyzing medical images. InceptionV3 uses multi-scale processing to quickly capture both local and global features. It uses factorized convolutions to make the calculations easier, asymmetric convolutions to make feature extraction more efficient, and auxiliary classifiers to improve gradient flow and act as regularizers. Batch normalization is also used a lot to make training more stable, and techniques for reducing grid size are used to make the best use of memory. These design features make InceptionV3 a great choice for dealing with complicated spatial patterns in chest X-ray images.

ResNet50, on the other hand, uses residual learning to let deep neural networks be trained without losing performance. The main part of it, the residual block, learns a residual mapping that looks like this: $H(x)=F(x)+x$, where x is the input and $F(x)$ is the learned transformation. Skip connections help gradients move through the network more easily, which helps with the vanishing gradient problem. ResNet50 gets stable and efficient training when you use batch normalization and ReLU activation functions together. Because it can keep up its high performance even when the depth increases, it works especially well for medical imaging tasks that have little labeled data and a lot of variability, like detecting COVID-19 from chest X-rays.

4. Results and Discussions

This section presents a comparative evaluation of multiple training configura-tions using original, traditionally augmented, and Progressive Growing GAN (PG-GAN) generated chest X-ray images. Our goal is to assess the influence of data filtering, augmentation, and class balance on classification performance across different deep learning models.

4.1. Filtered PG-GAN + Traditional Augmentation + Original Data

Applying a filtering pipeline to PG-GAN-generated images before combining them with traditionally augmented and original data resulted in the best classifi-cation performance across all model configurations. This approach ensured the inclusion of high-quality synthetic samples and improved the overall distribution of the training data. Table 1 shows that both models performed strongly when trained with the filtered PG-GAN-augmented dataset, with InceptionV3 delivering better results than Res-Net50 in every metric. This suggests that InceptionV3 is more effective in vali-dation accurately identifying COVID-19 from chest X-ray images.

Table 1. ResNet50 vs. InceptionV3 with Filtered PG-GAN + Traditional + Original Data

Model	Accuracy	Precision	Recall	F1-score
ResNet50	96.20%	0.9635	0.9586	0.9605
InceptionV3	97.39%	0.9761	0.9705	0.9726

4.2. Unfiltered PG-GAN + Traditional Augmentation + Original Data

Training with unfiltered PG-GAN data led to a noticeable drop in classification accuracy and class-wise recall, particularly for the NORMAL class. This suggests that quality control of GAN outputs is critical for ensuring effective learning. Table 2 reveals that using unfiltered PG-GAN-generated images led to reduced performance for both ResNet50 and InceptionV3, emphasizing the need to filter out low-quality synthetic data to ensure consistent validation accuracy and de-pendable model outcomes.

Table 2. ResNet50 vs. InceptionV3 with Unfiltered PG-GAN + Traditional + Original Data

Model	Accuracy	Precision	Recall	F1-score
ResNet50	94.76%	0.9509%	0.9476	0.9474
InceptionV3	93.55%	0.9355%	0.9311	0.9346

4.3. Balanced Dataset Using Traditional Augmentation

To mitigate class imbalance, we constructed a dataset with 8000 images per class via traditional augmentation. This approach improved performance com-pared to training on original-only data and also outperformed configurations with unfiltered GAN data. Although slightly behind the filtered GAN approach, this result validates that resolv-ing class imbalance contributes meaningfully to performance gains. Table 3 shows that using traditional augmentation to balance the dataset led to notice-able performance gains for both models, with InceptionV3 achieving better results than ResNet50 in all evaluation metrics. This highlights how balancing class distributions can positively impact diagnostic validation accuracy.

Table 3. Balanced Dataset Using Traditional Augmentation (8000 Samples per Class).

Model	Accuracy	Precision
ResNet50	91.08%	0.9203
InceptionV3	93.04%	0.9376

4.4. Original Data Only

Training models solely on the original dataset resulted in the lowest performance. This limitation is primarily due to the small dataset size and class imbalance. Table 4 highlights that both models performed well during training but struggled to maintain validation accuracy on validation data, indicating overfitting caused by the limited size and imbalance of the original dataset.

Table 4. Training on Original Dataset Only

Model	Training Accuracy	Validation Accuracy
ResNet50	96.15%	87.34%
InceptionV3	94.46%	89.36%

4.5. Baseline CNN Performance

A simple CNN model trained on the original data set served as a baseline and achieved the weakest generalization performance. Table 5 reveals that while the baseline CNN achieved decent accuracy and low loss, its overall performance lagged behind deeper models, highlighting the limitations of using simpler architectures on limited, non-augmented datasets.

Table 5. Baseline CNN on Original Dataset

Metric	Train Accuracy	Val Accuracy	Loss
CNN (Baseline)	92.40%	88.12%	0.2077

4.6. Comparison of Existing vs. Proposed System Performance

The comparison clearly shows that the proposed system performs better than the existing one, achieving a higher accuracy of 97.39% compared to 94.76%, thanks to the use of GAN-based image augmentation and deep learning. Improvements in precision, recall, and F1-score further demonstrate the system's strong ability to accurately detect COVID-19 cases, supported by balanced data and the use of explainable AI for greater clinical trust. Table 6. outlines how the dataset was carefully expanded to ensure each class had an equal number of samples for effective model training. Since the COVID-19 category originally had fewer images, it was enhanced using both traditional augmentation and high quality synthetic images generated by PG-GAN, whereas the Pneumonia and Nor-mal categories were balanced using only traditional augmentation methods to reach 8000 images each.

Table 6. Performance Comparison Between Existing vs Proposed System Models

Parameters	Existing System	Proposed System
Accuracy	94.76%	97.39%
Precision	0.9509	0.9761
Recall	0.9476	0.9705
F1-Score	0.9474	0.9726

Table 7 outlines how the dataset was carefully expanded to ensure each class had an equal number of samples for effective model training. Since the COVID-19 category orig-inally had fewer images, it was enhanced using both traditional augmentation and high-quality synthetic images generated by PG-GAN, whereas the Pneumonia and Normal cat-egories were balanced using only traditional augmentation methods to reach 8000 images each.

Table 7. Dataset Summary After Traditional and PG-GAN Based Augmentation

Category	New images added using Original Dataset	New images added using Traditional Augmentation	PG-GAN Augmentation (Filtered)	Final Image Count (Balanced)
COVID-19	576	6848	576	8000
Pneumonia	4273	3727	0	8000
Normal	1583	6417	0	8000

5. Conclusions

The experimental results show that adding carefully filtered PG-GAN-generated chest X-ray images to the training dataset, along with traditional augmentation methods and original data, greatly improves the accuracy of classification and the strength of the model. The addition of an AI-based filtering system was very important because it got rid of low-quality and structurally inconsistent synthetic samples, which stopped performance from getting worse because of artifacts or unrealistic patterns. Also, even though traditional augmentation methods helped reduce class imbalance and improve generalization, their effects were not as strong when used alone. On the other hand, adding high-quality GAN-generated data made the data more diverse and better at representing features, which led to bigger performance gains. The suggested multi-stage augmentation and filtering framework is a reliable and scalable way to boost deep learning performance in medical image classification tasks, especially when the datasets are small and unbalanced. Future endeavors will concentrate on investigating more sophisticated generative models, including StyleGAN and CycleGAN, to further augment the realism and variability of synthetic images. We will also look into adding more advanced filtering methods and training on higher-resolution, clinically refined datasets using advanced computing resources to make diagnostics more accurate and useful in the real world.

Data Availability Statement
Not applicable.

Funding
This work was supported without any funding.

Conflicts of Interest
The author declares no conflicts of interest.

Ethical Approval and Consent to Participate
Not applicable.

References

[1]. Gulakalap, H., et al. (2022). CNN-GAN pipeline for COVID-19 detection. *Computers in Biology and Medicine, 147*, 107262.
[2]. Fedoruk, D., et al. (2023). Multi-class GAN with EfficientNet for COVID-19 X-ray. *Scientific Reports, 13*(1).
[3]. Fedoruk, D., et al. (2023). GAN-enhanced vs traditional augmentation for COVID-19. *Medical Vision and AI, 32*(3), 77–85.

[4]. Rounaq, S., Bulu, M. N., & Rahman, S. (2023). *Wasserstein GAN + CNN for COVID-19 data scarcity* (arXiv:2401.06397). arXiv.

[5]. Motamed, S., Arabnia, H. R., & Gholami, R. (2021). Improved DAGAN for pneumonia detection in COVID-19 X-rays. *IEEE Transactions on Medical Imaging, 40*(11), 2898–2907.

[6]. Sakib, M., Alom, M., & Hassan, A. H. (2020). Modified CNN for chest radiographs. *IEEE Access, 8*, 220178–220187.

[7]. Puttagunta, K., & Reddy, D. V. (2022). DCGAN + transfer learning for COVID-19 detection. *Procedia Computer Science, 198*, 625–632.

[8]. Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). CovidGAN: Data augmentation using auxiliary classifier GAN for improved COVID-19 detection. *IEEE Access, 8*, 91916–91923.

[9]. Ali, D., Grönlund, C., & Shah, Z. (2023). *Improved GAN for COVID-19 detection using chest X-ray* (arXiv:2304.03536). arXiv.

[10]. Ali, A., & Shah, J. (2022). AI-driven GAN augmentation for COVID-19 X-ray. *JMIR Medical Informatics, 10*(6), e37365.

[11]. Fedoruk, D., et al. (2024). *Optimized StyleGAN2 for COVID-19 augmentation* (arXiv:2401.14705). arXiv.

[12]. Chen, L., Qiao, S., & Zhang, Y. (2022). Review of GANs in medical image synthesis. *Computers in Biology and Medicine, 141*, 105382.

[13]. Fedoruk, D., et al. (2023). *GAN vs. classical methods in COVID-19 X-rays* (arXiv:2304.09067). arXiv.

[14]. Krinski, F., Silva, R., Miranda, T., & Costa, M. (2023). CNN-GAN hybrid for low CT segmentation accuracy. *Computer Methods in Biomechanics and Biomedical Engineering, 26*(2), 137–145.

[15]. Khalif, R., Hossain, T., & Hossain, M. Z. (2020). GAN-enhanced CNN for COVID-19 detection. *Information, 11*(5), 258.

[16]. Alam, M. S., Rahman, M. M., & Hossain, M. A. (2025). Multiscale attention-based generative adversarial network for chest X-ray lesion synthesis. *Scientific Reports*.

[17]. Chen, Y., Wang, L., & Zhang, H. (2025). Generative models for improved medical image classification in low-data regimes. *Nature Biomedical Engineering*.

[18]. Karimi, D., Dou, H., & Gholami, A. (2025). Out-of-distribution detection in chest X-rays using generative adversarial networks. *Medical Image Analysis*.

[19]. Prashant268. (n.d.). *Chest X-ray (COVID-19 & pneumonia)* [Dataset]. Kaggle Source: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia.