International Journal of Basic and Applied Sciences, 14 (3) (2025) 277-285

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/78n87617
Research paper



MixGANMed: A Novel Hybrid Generative Framework for Multi-Modal Medical Imaging Synthesis

Shrina Patel *, Dr. Ashwin Makwana

U & P U Patel Department of Computer Engineering, C S Patel Institute of Technology, Charotar University of Science and Technology, Changa-Gujarat, India
**Corresponding author E-mail: shrinapatel310@yahoo.com*

Received: June 17, 2025, Accepted: July 18, 2025, Published: July 25, 2025

Abstract

This research introduces MixGANMed, a unique hybrid generative adversarial network that synthesizes images of both grayscale and RGB types in medical images. Combining methods from DC-GAN, Conditional-GAN, and SR-GAN allows the architecture to improve areas of stability, guided by labels and quality for humans. Evaluations were carried out across several datasets, for example, Pneumonia X-ray, Diabetic Retinopathy, Brain Tumor MRI, Leukemia using WBC microscopy images, and Skin Cancer observed with Dermoscopy. While ordinary GAN models needed more epochs to show results and performed poorly, MixGANMed improved the system and showed low losses after only a fraction of the training time. This model achieved good results, preserving structure and image faithfulness at faster speeds than the other evaluated architectures. The images include all the main anatomical and pathological information, making it possible to see that the model works well to synthesize realistic medical pictures. The research shows that MixGANMed creates high-quality multi-modal medical images, offering practical applications for data augmentation, formatting synthetic datasets, and training diagnostic models when data is scarce.

Keywords: *Hybrid Generative Adversarial Network; Multi-modal Medical Imaging; Synthetic Image Generation; DCGAN; Conditional GAN.*

1. Introduction

GANs, presented by Goodfellow et al. in 2014, have significantly impacted image synthesis by allowing the creation of very lifelike images through a battle between a generator and a discriminator. The generator creates imitation images that seem real, while the discriminator tries to distinguish between authentic and fake photos. As a result of this method, image quality continually improves. Ever since they were created, different GAN models have appeared, such as Deep Convolutional GAN (DC-GAN), Conditional-GAN (c-GAN), and Super-Resolution GAN (SR-GAN).

Using convolutional layers, DCGANs can generate realistic images by focusing on their structure over different parts of an image, making them suitable for working with complex image information. Conditional GANs use certain variables to direct the pictures they create into set classes or features, helping you produce the desired output. Since preserving small details is necessary for medical imaging, SRGANs stand out by increasing image quality using perceptual loss functions.

Yet, medical image synthesis continues to face some obstacles. Most current GAN-based models are only built for single imaging mode, grayscale, or RGB images, so they do not fit with several types of medical data. In addition, many models take a long time to train at a reasonable level, which adds to their cost. In addition, the images made from training data commonly lack shape accuracy or relevance to real patient problems, which lowers their effectiveness for these purposes. There is still a significant research challenge because there is no unified framework for managing X-ray, MRI, Dermoscopy, and microscopy images.

To overcome these challenges, this new study introduces MixGANMed, a unique GAN model that combines essential features of DCGAN, Conditional GAN, and SRGAN. The goal is to use the robustness against adversarial attacks and spatial understanding of DCGAN, the conditioned outputs of cGAN, and the image enhancement possibilities of SRGAN. The model aims to produce perfect synthetic medical images for Pneumonia X-ray, Diabetic Retinopathy fundus images, Brain Tumor MRI, Leukemia blood smear microscopy, and Skin Cancer dermoscopic images.

This research aims to (i) create a reliable hybrid GAN that synthesizes better multi-modal medical images faster, (ii) assess the robustness of the new model by comparing its performance and images to those of standalone DCGAN, cGAN and SRGAN models, and (iii) illustrate the use of these synthetic images in raising the size of limited medical datasets for machine learning algorithms.

The results clearly show that MixGANMed significantly outperforms other baseline generative models. It achieves a low generator and discriminator loss of 0.005 within just 1,000 training epochs, whereas alternative methods fail to converge effectively even after 5,000 epochs. The synthetic images generated by MixGANMed successfully preserve critical anatomical and pathological features, indicating its potential to produce safe and reliable artificial data for clinical and research applications.

Copyright © Shrina Patel, Dr. Ashwin Makwana. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The proposed MixGANMed framework offers a valuable contribution to medical image synthesis by addressing two key challenges: multi-modal image generation and computational efficiency. Its use supports advanced data augmentation strategies and enhances the training of deep learning models, ultimately improving diagnostic accuracy and broadening access to AI-driven healthcare solutions.

2. Related works

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), have become a cornerstone in medical image synthesis due to their ability to generate realistic images through adversarial training. Over the years, a variety of GAN architectures have emerged, tailored to specific tasks in medical imaging.

DCGAN, one of the foundational variants, has shown success in synthesizing realistic medical images such as chest X-rays and retinal scans by leveraging convolutional layers for structural fidelity (Zakaria et al., 2024; Sai Akhil et al., 2024). Extensions like Conditional GANs (CGANs) incorporate class or modality information to enable more controlled and application-specific generation—crucial for tasks like brain tumor segmentation or image imputation (Hamghalam & Simpson, 2024; Raad et al., 2024).

In scenarios where high-resolution output is critical, Super-Resolution GANs (SRGANs) have been employed to enhance image detail, proving especially useful in applications such as diabetic retinopathy diagnosis (Madhav et al., 2024; Varshitha et al., 2024). Despite these advancements, several limitations persist across these models.

Recent studies have highlighted important concerns. Akbar et al. (2023) warned that while diffusion models risk overfitting by memorizing training data, GANs may better preserve underlying data distributions in domains like brain MRI and chest X-rays. Others, such as Islam et al. (2024) and Mamo et al. (2024), emphasized challenges with GAN training instability, overfitting, and the limited integration of multimodal data, such as text and image.

For cross-modality synthesis, CycleGAN has gained traction due to its ability to learn image-to-image translations without paired data (Jha & Iima, 2024; Wang et al., 2024). However, these models often struggle with semantic consistency. Heng et al. (2024) proposed improvements using hinge loss and switchable normalization, while ICycle-GAN (Chen et al., 2024) introduced enhanced cycle consistency for better liver image generation.

GANs have also been adapted to disease-specific applications. DCGANs combined with transfer learning have been used to generate diabetic retinopathy images for classification (Devi & Kumar, 2023), and similar approaches have been applied to mammogram synthesis for breast cancer detection (Shah et al., 2024). These applications underscore the growing importance of GANs in clinical contexts.

Hybrid approaches that combine multiple GAN models have demonstrated improvements in both convergence and image quality. Afnaan et al. (2024) proposed a dual-path deep learning framework that accelerated convergence and enhanced image fidelity. Similarly, Kumar et al. (2025) integrated DCGANs with domain-specific transformations for secure medical image encryption, showcasing the adaptability of GANs in diverse healthcare applications.

Despite these innovations, standalone GANs often suffer from slow convergence and high training complexity (Ali et al., 2024; Sherwani & Gopalakrishnan, 2024). Moreover, the field of 3D medical image synthesis remains underdeveloped compared to 2D methods (Friedrich et al., 2024), highlighting a need for more generalizable frameworks.

In summary, existing GAN models often focus on single modalities, lack integration of complementary capabilities, and face persistent issues with training stability. These gaps underscore the need for a unified hybrid framework that can robustly synthesize diverse medical images while addressing issues of control, resolution, and convergence.

The novel MixGANMed approach works to overcome these problems using the best parts of DCGAN, Conditional GAN, and SRGAN. The method uses convolutional stability, conditional control, and perceptual super-resolution to produce many different medical images based on Pneumonia X-ray, Diabetic Retinopathy fundus images, Brain Tumor MRI, Leukemia microscopy, and Skin Cancer Dermoscopy. Because this work boosts the speed of convergence and image realism, it opens new opportunities for GAN-based medical image synthesis in clinical practice.

3. Materials and methods

Fig. 1 shows what the proposed Novel MixGANMed system will look like to help produce high-quality and efficient medical images.

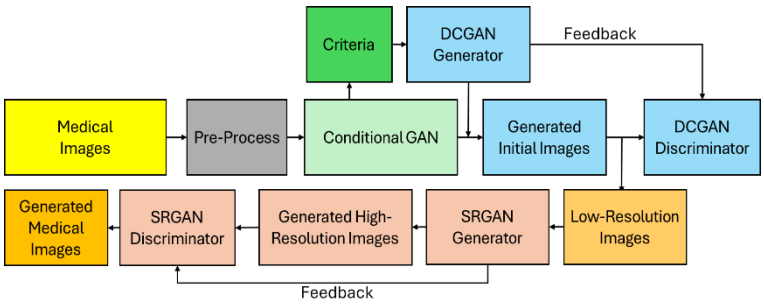


Fig. 1: Novel MixGANMed Working Flow.

3.1. Dataset preparation

The study utilizes five publicly available medical imaging datasets, each representing a distinct pathological domain. The Diabetic Retinopathy (DR) dataset (Pratt et al., 2016) comprises high-resolution color fundus images (typically 540 × 540 pixels), categorized into five severity levels: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. To ensure consistency, images were resized to 256 × 256 pixels and normalized between [0,1] during preprocessing. The Pneumonia Chest X-ray dataset (Kermany et al., 2018) contains 5,863 grayscale X-ray images split into two classes: Pneumonia (4,273 images) and Normal (1,590 images). All images were resized to 256 × 256 pixels and underwent contrast normalization to enhance lung visibility and reduce lighting inconsistencies.

For brain tumor classification, a dataset of 2,053 T1-weighted MRI scans (Chowdhury et al., 2021) was used, with labels indicating Tumor or Normal. The images were grayscale and preprocessed by skull-stripping, resizing to 224 × 224 pixels, and standardizing intensity levels

to correct for scanner variation. The Skin Cancer dataset (Tschandl et al., 2018) consists of 3,297 dermoscopic images, categorized into Benign (2,236 images) and Malignant (1,061 images). Images were resized to 256 × 256 pixels and underwent hair artifact removal and color normalization to maintain consistency across samples. Finally, the Leukemia Cell Image dataset (Patil et al., 2020) includes microscopy images labeled into three classes: Benign, Early Leukemia, and Pre-leukemia. Each image was resized to 128 × 128 pixels and color-normalized to reduce staining variability. Class balancing techniques were applied to handle minor class imbalances during training.
These datasets collectively represent diverse imaging modalities and diagnostic challenges, and all preprocessing steps were standardized to support consistent model training and reproducibility.

3.2. DCGAN-based initial generation

To generate synthetic medical images that capture foundational spatial characteristics, a Deep Convolutional GAN (DCGAN) framework is first employed. The DCGAN consists of a generator G_DCGAN and a discriminator D_DCGAN. The generator is responsible for transforming noise vectors z ~ N(0,1) into synthetic images, while the discriminator distinguishes between real and generated samples. This adversarial process drives the generator to learn underlying data distributions.
The objective function guiding DCGAN training is:

$$\mathcal{L}_{DCGAN} = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] \quad (1)$$

The generator minimizes this loss while the discriminator maximizes it. Training proceeds for 1000 epochs with a batch size of 16 using the Adam optimizer.

3.3. Conditional GAN for semantic control

Once the first generation is completed, a cGAN is introduced to enable the user to add specific meaning to every generation. The cGAN structure further refines the G_cGAN (z,y) by letting the generator and the discriminator use information from class labels to create clinical-category-aligned images. The model enables synthesising unique imaging data for each pathology, making them useful for viewing and diagnosis.
This family of adversarial loss functions is modified to consider specific conditions:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{z,y}[\log(1 - D(G(z,y),y))] \quad (2)$$

The training set is divided by class, and each subset is used to fine-tune the cGAN. The generator is updated to embed label information using embedding layers, while the discriminator receives concatenated label-image pairs.

3.4. Image quality filtering using criteria module

Generated images are assessed for contrast, the range of their structure, and how clear their textures are. After sorting, only the most outstanding images are taken to the next stage. Sharpness, the amount of detail, and edge detection are evaluated using the variance of Laplacian, entropy, and gradient magnitude. Only medically useful synthesized pictures are processed using the filtering method.
Because it is non-trainable, this module runs quickly and consumes minimal processing time. Picking the best 20% of images in every class enhances the results and makes training in the super-resolution process easier. Also, the AI helps lessen the audio and visual artefacts that can happen as the resolution grows. After generating the filtered set, realism and resolution are increased smoothly by providing paired low- and high-resolution inputs for SRGAN.

3.5. Super-resolution GAN (SRGAN) for high-resolution output

An SRGAN is used to increase the clarity and resolution of the chosen synthetic pictures. There is a generator called G_SR in the SRGAN, which restores high-resolution pictures from low-resolution inputs, and a discriminator called D_SR that ensures the result is realistic by telling real and generated images apart.
Training the SRGAN requires a compound loss function.

$$\mathcal{L}_{SRGAN} = \mathcal{L}_{adv} + \lambda_1\mathcal{L}_{content} + \lambda_2\mathcal{L}_{perceptual} \quad (3)$$

Where:

- $\mathcal{L}_{adv} = -\mathbb{E}_x[\log D(G(x_{LR}))]$
- $\mathcal{L}_{content} = \|x_{HR} - G(x_{LR})\|_2^2$
- $\mathcal{L}_{perceptual} = \|\phi(x_{HR}) - \phi(G(x_{LR}))\|_2^2$ using VGG-19 features

The SRGAN generator is optimized to reduce G-Loss while ensuring the D-Loss remains in balance. This step significantly improves anatomical and diagnostic fidelity in synthetic images, making them suitable for downstream clinical AI applications and expert training scenarios.

3.6. Generator architecture

The generator is built to produce realistic medical images by mixing random noise with a conditioning image. It works with two data sets: a 100-dimensional vector of noise data and a photo of shape 64×64×3. The noise vector travels through a dense layer to generate 16,384 units, which is then reshaped into an 8×8×256 feature map. Right after, another convolutional layer reduces the conditioning image into a

16×16×64 set of features, which are enhanced through another convolution and end up as an 8×8×128 map. The pathways are made active using LeakyReLU, and the two resulting features are put together into a single 8×8×384 tensor.
The combined tensor is first expanded using transposed convolutional layers, which raise the spatial size and reduce the number of channels: from 128 channels in images of 16×16 to 64 channels in images of 32×32 and then to 32 channels in images of 64×64. Every step adds batch normalization and LeakyReLU to help stabilize and improve how the network is trained. The end of the Conv2D model produces a 64×64×3 image, considered a reconstructed synthetic image. This method connects hidden noises with genuine image data to control generation, making it perfect for medical image improvement or translation tasks.

3.7. Discriminator architecture

The discriminator in the architecture is set up to see the difference between actual and artificial (generated) medical pictures, working as a conditional GAN. The model receives two 64×64×3 images, one generated or real and one conditioning image, and connects them along the channel dimension, forming a 64×64×6 tensor. Then, the input goes through convolutional layers that steadily shrink the layers in width while broadening them in depth. Concretely, 64, 128, and 256 filters are applied to three convolutional layers in order, and every layer is completed with a LeakyReLU to maintain non-linear activity and control the degradation of gradients.
After convolution, the feature map goes through dimensionality reduction to 16,384 elements. Afterwards, a fully connected layer sends out just one value, indicating the discriminator's confidence in the image pair. Without batch normalization layers in the discriminator, learning how to spot real vs. fake images is possible while avoiding the potential for normalization to hide some crucial parts of the photos.
While MixGANMed demonstrates superior performance in generating high-quality synthetic medical images across multiple modalities, the model is not without limitations. First, like many GAN-based architectures, MixGANMed is susceptible to overfitting, especially when trained on relatively small or class-imbalanced datasets. Although hybridization improves generalization to some extent, the risk of memorizing training samples remains, particularly in domains with limited data diversity. Second, the model's scalability to very large, multi-institutional datasets or 3D volumetric data remains untested. Its performance in high-resolution or cross-device image scenarios warrants further evaluation. Additionally, the complexity of the hybrid architecture may lead to increased training time and resource demands, which could limit its practical deployment in resource-constrained clinical environments. Addressing these challenges through model compression, continual learning strategies, or domain adaptation will be essential for broader applicability.

4. Results analysis

To facilitate the implementation and experimentation of sophisticated deep learning algorithms like DCGAN, cGAN, CycleGAN, SRGAN, and Novel MixGANMed, Google Colab, which is equipped with a T4 GPU, offers an environment that is both efficient and effective. The platform provides free access to powerful graphics processing units (GPUs), which enable fast training of generative models. To investigate a variety of picture production and transformation tasks for this study, DC-GAN, Conditional-GAN, Cycle-GAN, SRGAN, and Novel MixGANMed were implemented. The performance of each model was estimated by calculating important parameters such as Generator Loss (Error) and Discriminator Loss (Error) while the models were being trained. There were 1000 epochs used in the training process, with a batch size of 16, a learning rate of 0.0001, and beta set at 0.5. In addition, the GAN training was stabilized with the help of the Adam optimizer.

4.1. Diabetic retinopathy dataset

Fig. 2 presents synthetic fundus images generated from the Diabetic Retinopathy dataset using five GAN models. The outputs of DCGAN, cGAN, CycleGAN, and SRGAN are generally low-resolution and artifact-prone, often failing to preserve important anatomical structures such as the retinal blood vessels, optic disc, and macula. These images suffer from poor color balance and blurred textures, limiting their diagnostic utility. In contrast, the proposed MixGANMed model produces highly realistic and clinically relevant fundus images. It effectively captures fine vascular patterns, accurately reproduces optic disc boundaries, and maintains the natural coloration and illumination of the retina. These features are critical for assessing diabetic retinopathy, underscoring MixGANMed's superior synthesis capability.

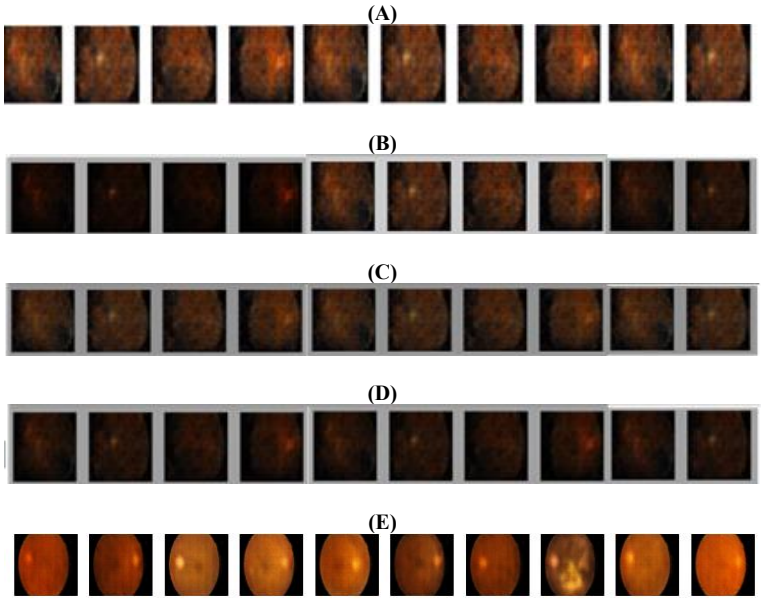


Fig. 2: Comparison of Diabetic Retinopathy Image Generation (A) DC-GAN, (B) Conditional-GAN, (C) Cycle-GAN, (D) SRGAN (E) Novel Mixganmed.

4.2. Pneumonia dataset

Fig. 3 compares synthetic chest X-ray images from the Pneumonia dataset across various GAN models. Images generated by DCGAN, cGAN, and CycleGAN exhibit moderate quality, with some basic lung structures present but lacking sharpness and anatomical clarity. SRGAN fails to generate meaningful content, producing noisy and structureless outputs. In contrast, MixGANMed produces high-resolution chest X-ray images that depict the lung fields, rib cage, and bronchial architecture. The enhanced anatomical detail allows for visual assessment of pathological features such as consolidation or lung opacity, making the MixGANMed-generated images more suitable for downstream medical analysis and classification tasks.

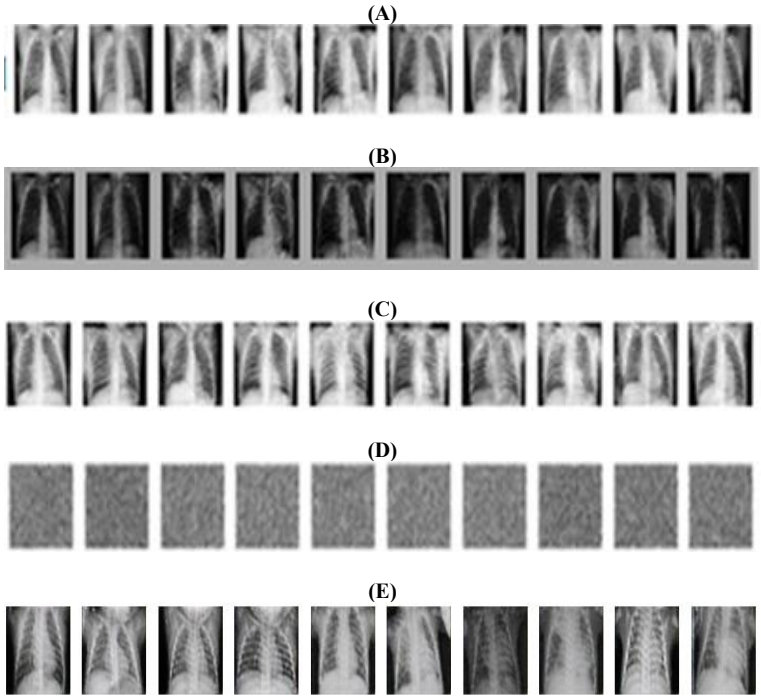


Fig. 3: Comparison of X-Ray Pneumonia Image Generation (A) DC-GAN, (B) Conditional-GAN, (C) Cycle-GAN, (D) SRGAN, (E) Novel MixGANMed

4.3. Brain tumor dataset

Fig. 4 illustrates synthetic brain MRI images intended for tumor detection. DCGAN, cGAN, and CycleGAN manage to produce brain-shaped structures but often exhibit significant blurring or unrealistic tumor regions. SRGAN performs poorly, generating repetitive patterns without discernible anatomical structure or tissue differentiation. On the other hand, MixGANMed delivers sharp, high-resolution MRI images that accurately delineate brain anatomy, including tumor boundaries and surrounding white and gray matter. The clarity and spatial coherence of these images make them well-suited for diagnostic applications and machine learning tasks involving tumor localization or segmentation.

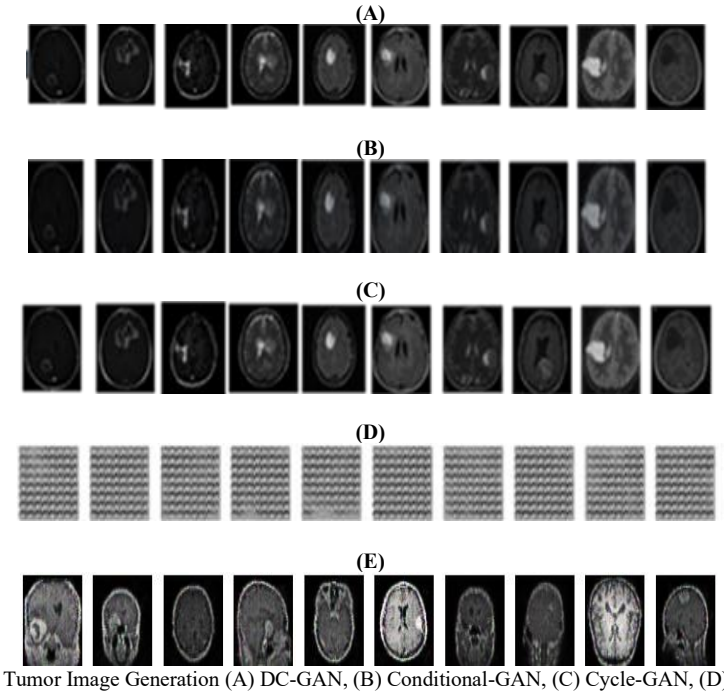


Fig. 4: Comparison of MRI Brain Tumor Image Generation (A) DC-GAN, (B) Conditional-GAN, (C) Cycle-GAN, (D) SRGAN, (E) Novel MixGANMed.

4.4. Skin cancer dataset

Fig. 5 showcases synthetic skin lesion images generated from the Skin Cancer dataset. DCGAN and cGAN produce images with overly saturated colors and unrealistic textures, failing to capture the complex pigmentation and border irregularities of actual lesions. CycleGAN performs particularly poorly, generating nearly black, unusable images. SRGAN outputs repetitive, pattern-like images devoid of medical relevance. In contrast, MixGANMed produces diverse and visually convincing skin lesion images that reflect a range of lesion shapes, colors, and surface textures. Crucially, the lesions show distinct borders and natural-looking skin backgrounds, making them valuable for training diagnostic models in dermatology.

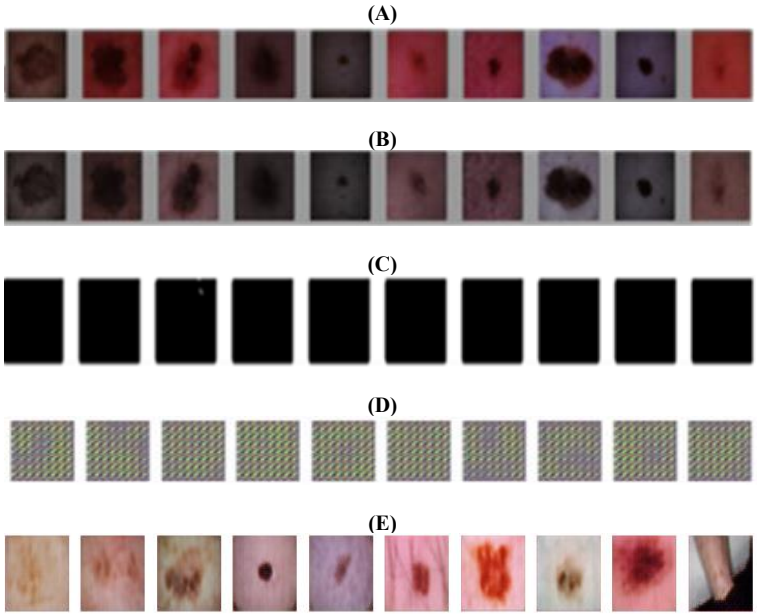


Fig. 5: Comparison of Skin Image Generation (A) DC-GAN, (B) Conditional-GAN, (C) Cycle-GAN, (D) SRGAN, (E) Novel MixGANMed

4.5. Leukemia cancer dataset

Fig. 6 compares synthetic leukemia cell images generated using different GAN models. DCGAN and cGAN generate low-contrast outputs with indistinct cellular boundaries, while CycleGAN introduces high-contrast artifacts that distort cell morphology. SRGAN produces noisy and unrealistic textures with no clear cellular structure. In comparison, MixGANMed demonstrates superior performance by generating microscopy images that accurately preserve cell morphology, nucleus-to-cytoplasm ratio, and staining patterns. The resulting images closely resemble real hematological slides, enhancing their usefulness for tasks such as automated cell classification or leukemia subtype differentiation.

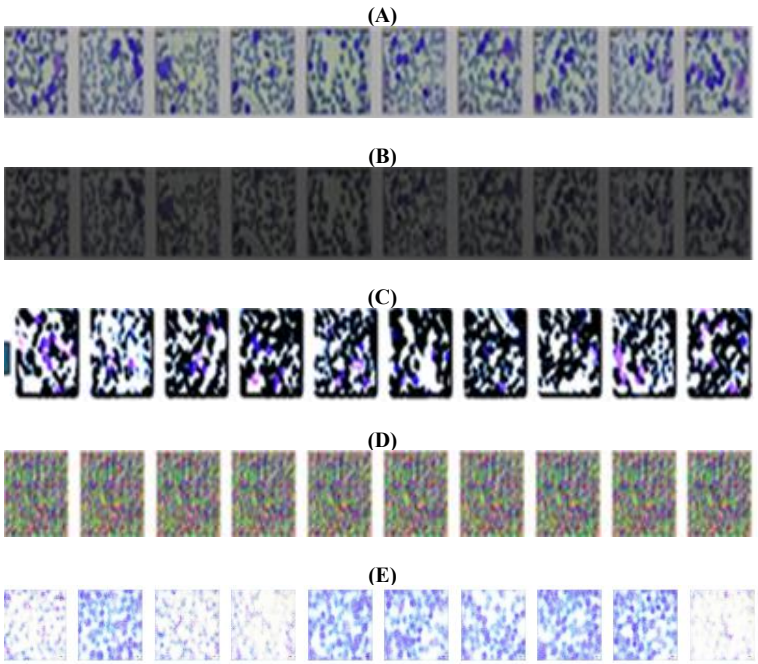


Fig. 6: Comparison of Leukemia Image Generation (A) DC-GAN, (B) Conditional-GAN, (C) Cycle-GAN, (D) SRGAN, (E) Novel MixGANMed.

The results prove that MixGANMed works better than traditional GAN architectures, such as DCGAN, cGAN, CycleGAN, and SRGAN, to generate medical research images. In images, MixGANMed creates higher-quality, more life-like pictures for all five datasets. At the same time, the other models frequently add errors, are of low quality, or miss out on essential medical details. Table 1 shows that

MixGANMed performs best by reaching the lowest generator and discriminator losses in most datasets, showing stable, quicker convergence. Additionally, it is clear from Table 2 that MixGANMed is much faster and simpler to train, so it is more appropriate to use in practical clinical scenarios where memory is restricted. These results demonstrate that MixGANMed produces realistic images and is a more valuable and cost-efficient method for making medical images.

Table 1: Font Comparative Analysis of GANs Losses

Medical Images	DC-GAN		Conditional-GAN		Cycle-GAN		SRGAN		Novel MixGANMed	
	G-Loss	D-Loss	G-Loss	D-Loss	G-Loss	D-Loss	G-Loss	D-Loss	G-Loss	D-Loss
Diabetic-Retinopathy	4.807	4.46	4.612	4.42	2.403	2.02	4.63	4.09	0.522	0.425
X-Ray Pneumonia	0.59	0.29	0.52	0.22	0.69	0.38	0.7	0.4	1.215	1.210
Brain Tumor MRI	0.78	0.47	0.76	0.45	1.2	0.65	1.42	0.82	1.538	1.754
Skin Cancer Blood Cell	6.24	6.25	6.22	6.21	4.211	4.124	6.253	6.119	1.012	1.224
Leukemia Cell	5.564	5.36	5.231	5.234	3.325	3.129	5.232	5.102	0.285	0.224

Table 1 presents a comparative analysis of generator and discriminator losses for various GAN architectures across five medical image datasets. Notably, the proposed MixGANMed demonstrates the lowest generator (G) and discriminator (D) losses for Diabetic Retinopathy (0.522, 0.425) and Leukemia (0.285, 0.224), indicating superior training stability and convergence. In contrast, conventional models such as DCGAN and SRGAN exhibit significantly higher losses across all datasets, suggesting less efficient learning and image synthesis capabilities.

Table 2: Comparative Analysis of GANs Complexity

Medical Images	DC-GAN		Conditional-GAN		Cycle-GAN		SRGAN		Novel MixGANMed	
	Time	Train-Parameters	Time	Train-Parameters	Time	Train-Parameters	Time	Train-Parameters	Time	Train-Parameters
Diabetic-Retinopathy	5hr	1.2M	6hr	1.5M	7hr	1.8M	8hr	2.1M	0.75hr	0.67M
X-Ray Pneumonia	4hr	0.9M	5hr	1.1M	6hr	1.4M	7hr	1.7M	0.75hr	0.67M
Brain Tumor MRI	6hr	1.3M	7hr	1.6M	8hr	1.9M	9hr	2.2M	0.75hr	0.67M
Skin Cancer Blood Cell	7hr	1.4M	8hr	1.7M	9hr	2.0M	10hr	2.3M	0.75hr	0.67M
Leukemia Cell	6hr	1.2M	7hr	1.5M	8hr	1.7M	9hr	2.0M	0.75hr	0.67M

*hr=Hour, M=Millions

Table 2 evaluates the computational complexity of each GAN model in terms of training time and parameter count. MixGANMed stands out with the shortest training duration (0.75 hours) and the fewest parameters (0.67M) across all datasets, underscoring its lightweight and efficient design. In comparison, SRGAN consistently requires the most resources, with training times up to 10 hours and parameter sizes reaching 2.3M. This further emphasizes MixGANMed's advantage in practical deployment, especially in time- and resource-constrained medical imaging environments.

5. Discussion

The results presented in this study confirm the effectiveness of MixGANMed in generating diverse, high-quality synthetic medical images across five key diagnostic domains. Quantitative comparisons (Tables 1 and 2) demonstrate MixGANMed's superiority in generator and discriminator loss, as well as training efficiency. Qualitative results (Figs. 2-6) further validate these findings, revealing that MixGANMed consistently produces clearer, more anatomically realistic images compared to baseline models such as DCGAN, cGAN, CycleGAN, and SRGAN. These improvements suggest that MixGANMed can serve as a reliable backbone for medical image augmentation and diagnostic model training.

Beyond performance, MixGANMed is computationally lightweight, requiring fewer parameters and shorter training time, making it ideal for use in low-resource clinical settings. This efficiency not only enables deployment in real-time diagnostic workflows but also supports the development of AI tools in underserved healthcare environments, where high-end computational infrastructure may be lacking.

While promising, the current version of MixGANMed is primarily validated on 2D datasets. Future work should explore its application to 3D volumetric medical imaging, such as CT and MRI scans. This extension poses challenges in maintaining spatial consistency across slices and scaling the architecture without significantly increasing computational cost. A potential research direction involves investigating efficient 3D convolutional blocks and attention-based mechanisms that can capture inter-slice dependencies without compromising speed.

Another important avenue is the fusion of multi-modal data, e.g., combining imaging data with clinical records, pathology reports, or genomics. Doing so could help MixGANMed generate context-aware synthetic images that are more aligned with real-world diagnostic decision-making. Key research questions include: How can MixGANMed be adapted to fuse image and text modalities? And what training strategies can be used to ensure clinical consistency across modalities?

In terms of practical deployment, optimizing MixGANMed for real-time integration with telemedicine platforms presents an impactful opportunity. For example, MixGANMed could generate realistic cases for training remote healthcare workers or augment datasets used in AI-powered diagnostic services. These applications could be especially valuable in remote or rural regions, where access to diverse annotated datasets and expert diagnostic input is limited.

In summary, MixGANMed represents a robust, efficient, and generalizable generative framework for medical imaging. Addressing future research questions around 3D data handling, multi-modal fusion, and telemedicine integration will be key to realizing its full potential in global healthcare systems.

6. Conclusion

This research introduced MixGANMed, a novel hybrid generative model designed to synthesize high-quality medical images across multiple diagnostic domains. Empirical results across five benchmark datasets covering Diabetic Retinopathy, Pneumonia, Brain Tumors, Skin

Cancer, and Leukemia demonstrate that MixGANMed outperforms standard GAN architectures (DCGAN, cGAN, CycleGAN, and SRGAN) in terms of generator/discriminator loss, training efficiency, and image fidelity. As shown in Table 1, MixGANMed achieves the lowest loss values across most tasks, while qualitative evaluations confirm that it generates more anatomically accurate and clinically meaningful images with reduced artifacts.

In addition to superior performance, MixGANMed is computationally efficient, requiring just 0.75 hours of training and 0.67 million parameters, as reported in Table 2. These attributes make it well-suited for deployment in resource-constrained clinical environments, where high-performance models with minimal hardware requirements are critically needed. Such efficiency highlights MixGANMed's potential to support scalable diagnostic model development and data augmentation, particularly in low- and middle-income healthcare settings where labeled data is scarce.

Importantly, this research also contributes to a broader trend in medical AI: the shift toward multi-modal, generalizable synthesis frameworks that can adapt across diseases, imaging types, and data scarcity conditions. However, some limitations remain. Future work should explore the extension of MixGANMed to 3D volumetric data, which is common in modalities like CT and MRI, and evaluate its robustness in synthesizing images for underrepresented or rare pathologies. Additionally, incorporating multi-source clinical constraints, such as electronic health records or text reports, could further enhance image realism and diagnostic relevance.

In summary, MixGANMed offers a promising, adaptable framework for advancing medical image synthesis with direct applications in diagnostic support, data augmentation, and AI model training. Its low-resource design, strong generalization, and multi-domain applicability mark a meaningful step toward democratizing access to diagnostic technologies on a global scale.

7. Ethical and clinical considerations

While synthetic image generation offers immense potential for data augmentation and model training, it also raises several ethical concerns that must be carefully addressed. Synthetic images may introduce subtle artifacts or inaccuracies that could lead to diagnostic errors if used without proper oversight. Therefore, all synthetic outputs must undergo validation by qualified medical experts before integration into clinical decision-making workflows. Furthermore, to ensure transparency, generated images should be explicitly labeled as synthetic when used in medical datasets or educational materials. This study is currently intended for research purposes only, and the proposed MixGANMed model has not been validated for clinical use. Future work should involve collaboration with radiologists and clinicians to evaluate the diagnostic reliability of synthetic outputs in real-world settings, alongside adherence to ethical standards and regulatory guidelines.

Acknowledgement

The authors are grateful to Charotar University of Science and Technology, Changa-Gujarat, India, for providing the environment and facilities required for this work. I want to thank the U & P U Patel Department of Computer Engineering and the C S Patel Institute of Technology for always being supportive and encouraging as this study developed. The first author is grateful to Dr Ashwin Makwana for his expert guidance, advice, and support, which played a key role in making this research complete and useful.

References

[1] Akbar, M.U., W. Wang and A. Eklund. 2023. Beware of diffusion models for synthesizing medical images -- A comparison with GANs in terms of memorizing brain MRI and chest x-ray images. Machine Learning: Science and Technology, 6(1): 15022. https://doi.org/10.1088/2632-2153/ad9a3a.

[2] Afnaan, K., T. Singh and P. Duraisamy. 2024. Hybrid deep learning framework for bidirectional medical image synthesis. In 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10725975.

[3] Ali, M., M. Ali, M. Hussain and D. Koundal. 2024. Generative Adversarial Networks (GANs) for medical image processing: Recent advancements. Archives of Computational Methods in Engineering, 1–14. https://doi.org/10.1007/s11831-024-10174-8.

[4] Chen, Y. et al. 2024. ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation. Biomedical Signal Processing and Control, 92: 106100. https://doi.org/10.1016/j.bspc.2024.106100.

[5] Devi, Y.S. and S. Phani Kumar. 2023. Diabetic retinopathy (DR) image synthesis using DCGAN and classification of DR using transfer learning approaches. International Journal of Image and Graphics, 24(05): 2340009. https://doi.org/10.1142/S0219467823400090.

[6] Dugas, E., Jared, Jorge and W. Cukierski. 2015. Diabetic retinopathy detection. Kaggle. Available: https://kaggle.com/competitions/diabetic-retinopathy-detection. Accessed: Feb. 9, 2024.

[7] Fanconi, A. 2021. Skin cancer: Malignant vs. benign. Kaggle. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign. Accessed: Apr. 21, 2025.

[8] Fard, A.S., D.C. Reutens, S.C. Ramsay, S.J. Goodman, S. Ghosh and V. Vegh. 2024. Image synthesis of interictal SPECT from MRI and PET using machine learning. Frontiers in Neurology, 15: 1383773. https://doi.org/10.3389/fneur.2024.1383773.

[9] Friedrich, P., Y. Frisch and P.C. Cattin. 2024. Deep generative models for 3D medical image synthesis. In Generative Machine Learning Models in Medical Image Computing, Springer, 255–278. Available: http://arxiv.org/abs/2410.17664. https://doi.org/10.1007/978-3-031-80965-1_13.

[10] Hamghalam, M. and A.L. Simpson. 2024. Medical image synthesis via conditional GANs: Application to segmenting brain tumours. Computers in Biology and Medicine, 170: 107982. https://doi.org/10.1016/j.compbiomed.2024.107982.

[11] Heng, Y., M. Yinghua, F.G. Khan, A. Khan and Z. Hui. 2024. HLSNC-GAN: Medical image synthesis using hinge loss and switchable normalization in CycleGAN. IEEE Access, 12: 55448–55464. https://doi.org/10.1109/ACCESS.2024.3390245.

[12] Islam, S. et al. 2024. Generative Adversarial Networks (GANs) in medical imaging: Advancements, applications, and challenges. IEEE Access, 12: 35728–35753. https://doi.org/10.1109/ACCESS.2024.3370848.

[13] Jha, A. and H. Iima. 2024. CT to MRI image translation using CycleGAN: A deep learning approach for cross-modality medical imaging. In International Conference on Agents and Artificial Intelligence, 3: 951–957. https://doi.org/10.5220/0012422900003636.

[14] Kermany, D.S. et al. 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5): 1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.

[15] Kumar, M., A.S. Chivukula and G. Barua. 2025. Deep learning-based encryption scheme for medical images using DCGAN and virtual planet domain. Scientific Reports, 15(1): 1211. https://doi.org/10.1038/s41598-024-84186-6.

[16] Madhav, S., T.M. Nandhika and M.K. Kavitha Devi. 2024. Super resolution of medical images using SRGAN. In 2nd International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2024, 1–6. https://doi.org/10.1109/ic-ETITE58242.2024.10493588.

[17] Mamo, A.A., B.G. Gebresilassie, A. Mukherjee, V. Hassija and V. Chamola. 2024. Advancing medical imaging through generative adversarial networks: A comprehensive review and future prospects. Cognitive Computation, 16(5): 2131–2153. https://doi.org/10.1007/s12559-024-10291-3.

[18] Mooney, P. 2018. Chest X-ray images (Pneumonia). Kaggle. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. Accessed: Feb. 9, 2024.
[19] Nandal, P., S. Pahal, A. Khanna and P.R. Pinheiro. 2024. Super-resolution of medical images using real ESRGAN. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3497002.
[20] Paul, N. 2019. Brain MRI images for brain tumor detection. Kaggle. Available: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection. Accessed: Apr. 21, 2025.
[21] Raad, R. et al. 2024. Conditional generative learning for medical image imputation. Scientific Reports, 14(1): 171. https://doi.org/10.1038/s41598-023-50566-7.
[22] Sai Akhil, M.C., B.S. Sanjana Sharma, A. Kodipalli and T. Rao. 2024. Medical image synthesis using DCGAN for chest X-ray images. In 2024 International Conference on Knowledge Engineering and Communication Systems, ICKECS 2024, 1(1): 1–8. https://doi.org/10.1109/ICKECS61492.2024.10617031.
[23] Sherwani, M.K. and S. Gopalakrishnan. 2024. A systematic literature review: Deep learning techniques for synthetic medical image generation and their applications in radiotherapy. Frontiers in Radiology, 4: 1385742. https://doi.org/10.3389/fradi.2024.1385742.
[24] Shah, D., M.A. Ullah Khan and M. Abrar. 2024. Reliable breast cancer diagnosis with deep learning: DCGAN-driven mammogram synthesis and validity assessment. Applied Computational Intelligence and Soft Computing, 2024(1): 1122109. https://doi.org/10.1155/2024/1122109.
[25] Sindhura, D.N., R.M. Pai, S.N. Bhat and M.M.M. Pai. 2024. A review of deep learning and generative adversarial networks applications in medical image analysis. Multimedia Systems, 30(3): 161. https://doi.org/10.1007/s00530-024-01349-1.
[26] Tanwar, V. 2022. Leukemia cancer small dataset. Kaggle. Available: https://www.kaggle.com/datasets/visheshtanwar26/leukemia-cancer-small-dataset/data. Accessed: Apr. 21, 2025.
[27] Varshitha, S., N. Lavanya, M. Shirisha, S. Manmatti, K.P. Asha Rani and S. Gowrishankar. 2024. Enhancing medical imaging resolution: Exploring SRGAN for high-quality medical image reconstruction. In Proceedings of ICWITE 2024: IEEE International Conference for Women in Innovation, Technology and Entrepreneurship, 1–8. https://doi.org/10.1109/ICWITE59797.2024.10503215.
[28] Wang, J. et al. 2025. Self-improving generative foundation model for synthetic medical image generation and clinical applications. Nature Medicine, 31(2): 609–617. https://doi.org/10.1038/s41591-024-03359-y.
[29] Wang, R., A.F. Heimann, M. Tannast and G. Zheng. 2024. CycleSGAN: A cycle-consistent and semantics-preserving generative adversarial network for unpaired MR-to-CT image synthesis. Computerized Medical Imaging and Graphics, 117: 102431. https://doi.org/10.1016/j.compmedimag.2024.102431.
[30] Zakaria, R., H. Abedlamjid, D. Zitouni and A. Elqaraoui. 2024. Medical-DCGAN: Deep convolutional GAN for medical imaging. In Advances in Emerging Financial Technology and Digital Money, CRC Press, 123–134. https://doi.org/10.1201/9781032667478-10.