JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 01
2025

Research Article
https://doi.org/10.56979/1001/2025/1182

Hybrid GAN-Augmented Multi-Modal Medical Imaging Framework with Squeeze-and-Excitation CNN for Robust Classification

Shrina Patel[1*], and Ashwin Makwana[1]

[1]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: Shrina Patel. Email: shrinapatel310@yahoo.com

Received: September 14, 2025 **Accepted**: November 28, 2025

Abstract: A hybrid GAN architecture was designed to solve multi-modal medical image synthesis and disease classification by combining key architectural principles from DC-GAN, Conditional GAN, and SR-GAN, thereby enhancing training stability, providing label-conditioned image synthesis, and improving perceptual image quality. The proposed framework was extensively tested on large, diverse medical imaging datasets, including chest X-ray images to identify pneumonia, retinal fundus images to evaluate diabetic retinopathy, brain MRI images to detect tumors, microscopic images of leukemia white blood cells, and dermoscopic images to analyze skin cancer. Quantitative experimentation revealed a steady convergent behavior, the values of the generator loss and discriminator loss continued to decline throughout each of the datasets and the lowest values were found in diabetic retinopathy cases (generator loss of 0.522 and discriminator loss of 0.425) and leukemia cases (generator loss of 0.285 and discriminator loss of 0.224), maintaining the presence of diagnostically significant pathological features. Computational efficiency was also high, with about 0.75 hours of training time and a relatively small number of 0.67 million parameters, compared to SR-GAN-based models that require more than 10 hours of training time and more than 2.3 million parameters. The success of the framework was also confirmed by the quality of the generated images, attaining a high signal-to-noise ratio of 36.742, structural similarity index of 0.93, and Fréchet inception distance of 30.402, which is better than several other recent state-of-the-art methods, such as DRForecastGAN, GAN-VSP, IFGAN, and Pix2Pix GAN. Also, incorporating a Squeeze-and-Excitation convolutional neural network classifier not only led to a significant boost in disease classification performance but also improved the accuracy of diabetic retinopathy and pneumonia to 0.90 and 0.98, respectively. In general, the suggested hybrid GAN model has great potential as a low-cost, high-quality solution for medical image generation, data augmentation, and automated disease detection in clinical decision-support systems.

Keywords: Generative Adversarial Network (GAN); Squeeze-and-Excitation Convolutional Neural Network (SE-CNN); Medical Image Synthesis; Multi-Modal Medical Image Classification; Deep Learning for Medical Imaging

1. Introduction

Modern healthcare radiology is an integral part of contemporary healthcare. It can optimize early detection, establish diagnostic accuracy, and explore types of diseases such as Pneumonia, Diabetic Retinopathy, Brain Tumors, Leukemia, and Skin Cancer, among others. Improved availability of radiology services like X-ray, MRI, fundus photography, Dermoscopy, and white blood cell microscopy has occasioned the creation of CA-CAD systems at different levels. Automated analysis of these images should decrease workload on human health care and bring about a high level of precision in diagnoses, fostering timely interventions in health regions with limited access to health care. In the last decade, deep learning approaches, particularly convolutional neural networks (CNNs), have enjoyed great success in image

classification given their ability to learn hierarchical feature representation from raw images. **Figure 1** shows the basic process of Generative Adversarial Network.

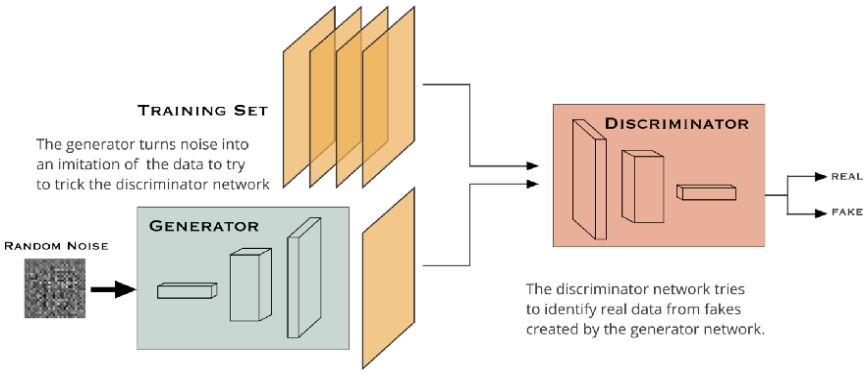


Figure 1. Generative Adversarial Network Process

Although there have been many advancements in this field, the practical deployment of new systems in real clinical settings is still limited. One of the main challenges is the lack of sufficiently large and well-annotated medical image datasets. This scarcity is mainly due to patient privacy regulations, high costs involved in expert labeling, and the limited availability of samples for rare diseases. In addition, class imbalance in datasets often reduces model generalization and negatively affects diagnostic performance. Variations in input data, such as differences in image resolution, contrast, and noise caused by different imaging devices, further increase the complexity of the problem. These variations make it difficult for convolutional neural networks to learn consistent and robust features, even when deep or stacked architectures are used. As a result, hidden data inconsistencies can remain unaddressed, limiting the reliability of simpler modeling approaches.

This research proposes a Hybrid GAN-based system to tackle the problem of limited patient data and class imbalance. The invented model is based on the integration of DCGAN, Conditional GAN, and SR-GAN to synthesize high-quality medical grayscale and RGB medical images across the different imaging modalities. In this work, the term multi-modal refers to a unified architectural framework applied consistently across heterogeneous medical imaging modalities, rather than a single jointly trained multi-modal network with shared representations. The resultant hybrid model assures stable convergence, quickened training, and guarantees preservation of the essential clinical characteristics of the synthesized images while preserving and enhancing clinically imperative features.

A Squeeze-and-Excitation Convolutional Neural Network (SE-CNN) is thus proposed for classifying images derived from multiple modes concisely. The Squeeze-and-Excitation CNN attempts to amplify semantic information experienced on spatial relations between the channels by mending feature representations, thus more expressiveness speeds in suppressing confounders of disease features. This gradient integrates well the whole SE-CNN psychometric framework with GAN Committee stages in the Hope-Balance Condition, generating mean classifications, concomitant to the performance of application cases to generate full-trade health care benefits.

With reference to the context, activities will be arranged in such a way that this study will close the research gaps in contemporary medical image analysis frameworks. Pre-eminently, this is to be achieved by integrating the new main blocks of hybrid GAN-based image synthesis and SE-CNN classification. By focus, the study proposes to be daring with a view to realizing a system abundantly endowed with generative and classification power for arrest to diseases.

1.1. Related Work

The synthesis and analysis of medical images have considerable importance due to the ever-increasing demand for automated and accurate diagnostic systems. Among the generative models, Generative Adversarial Networks (GANs) offer feature-rich standalone tools to synthesize high-quality images to enhance scarce datasets and improve classification on downstream tasks. In their study, Akbar et al. [1] gave a comparison of diffusion models and GANs while synthesizing brain MRI and chest X-ray images, discussing the possibility of memorization on diffusion-based models, and establishing GANs' perspective

in terms of fidelity and reliability for the medical image synthesis. Afnaan et al. [2] also presented a hybrid deep learning framework for bidirectional medical image synthesis, which combined various generative techniques to produce a more robust and diverse image output. Ali et al. [3] provided a recent review regarding GAN technologies applied in medical image processing, with emphasis on architecture optimization and training strategies to improve stability and realism.

Recent work has been focused on improving conditional generation and cycle-consistency. The authors Chen et al. [4] developed Cycle-GAN for liver image generation, which enforces cycle consistency and produces higher quality images with greater structural preservation. Devi and Kumar [5] utilized DCGANs in diabetic retinopathy image synthesis with transfer learning for classification, suggesting that realistic synthetic images can appreciably boost the performance of the diagnostic model. These models were trained and assessed using large public datasets such as the Diabetic Retinopathy dataset [6], Skin Cancer datasets [7], and interictal SPECT imaging datasets [8], giving benchmark data for validation of the developed models.

Application of deep generative models spans to multi-modal medical image applications. Friedrich et al. [9] provided a succinct overview of 3D image synthesis utilizing deep generative models, spotlighting yet challenges in volume consistency and anatomical correctness. Hamghalam and Simpson [10] explored the possibilities of conditional GANs for segmentation of brain tumors where label-guided synthesis can enhance generation quality and segmentation accuracy. Heng et al. [11] proposed the HLSNC-GAN framework, which involves the hinge loss and switchable normalization incorporated into CycleGAN for stable and faithful imaging. Islam et al. [12] surveyed advances and the challenges in GANs in medical imaging that stress limitations including unstable training, mode collapse, and computational heaviness.

Cross-modality translation also attracted considerable attention. Jha and Iima [13] implemented CT-to-MRI translation via CycleGAN for cross-modality image synthesis without paired dataset usage. Kermany et al. [14] raised the prospect of making synthetic data integration by deep learning for automated diagnosis through image-methods. Kumar et al. [15] synthesized medical images using DCGAN with an encryption scheme for security; Madhav et al. [16] and Varshitha et al. [27] investigated image enhancement through super-resolution using SRGAN. Systematic reviews such as those by Mamo et al. [17] and Sindhura et al. [25] analyzed GAN-based approaches, applications in the clinic, and future outlooks.

Various studies aimed to overcome dataset limitations through synthesizing modality-specific images. Annotated datasets for Pneumonia X-rays and Brain MRI were made available by Mooney [18] and Paul [20], respectively, for model evaluation. Nandal et al. [19] introduced ESRGAN-based super-resolution for medical images to further improve feature representation for classification. Conditional learning techniques for imputation and augmentation are demonstrated by Raad et al. [21] and Akhil et al. [22], while Sherwani and Gopalakrishnan [23] reviewed deep learning methods for synthetic image generation in radiotherapy. For breast cancer detection enhancement, mammogram synthesis using DCGANs was implemented by Shah et al. [24].

In more recent developments, emphasis has been placed on hybrid and multi-stage GAN architectures. Tanwar [26] provided small datasets for leukemia classification, highlighting problems of limited data availability. The paper [28] proposed self-improving generative foundation models for synthetic medical image generation, ensuring high fidelity in different clinical scenarios. Wang et al. [29], through CycleGAN, emphasized cycle-consistency and semantic preservation for unpaired MR-to-CT synthesis. Finally, Medical-DCGAN proposed by Zakaria et al. [30] is a deep convolutional GAN code constructed for medical imaging applications, yielding superior image quality and stability.

Overall, the literature reveals that while GAN-based techniques for medical image synthesis have made strong progress, stability, computational efficiency, generalization across modalities, and retention of important pathological features continue to pose challenges. Hybrid architectures using conditional guidance, cycle-consistency, and attention mechanism proved to be useful avenues for addressing such challenges, paving the way for high-fidelity multi-modal image generation along with improved downstream classification. A cumulative line of evidence suggests that integrating synthetic image

generation with high-performance classification networks is the key to advancing automated medical diagnosis and realizing dependable AI-assisted clinical decision-making.

2. Materials and Methods

In Figure 2, the proposed Hybrid GAN is a sequential pipeline consisting of (i) DCGAN-based coarse generation, (ii) conditional GAN refinement using class labels, and (iii) SRGAN-based super-resolution enhancement. CycleGAN is not employed in the synthesis pipeline and is only referenced as a comparative baseline from prior literature.

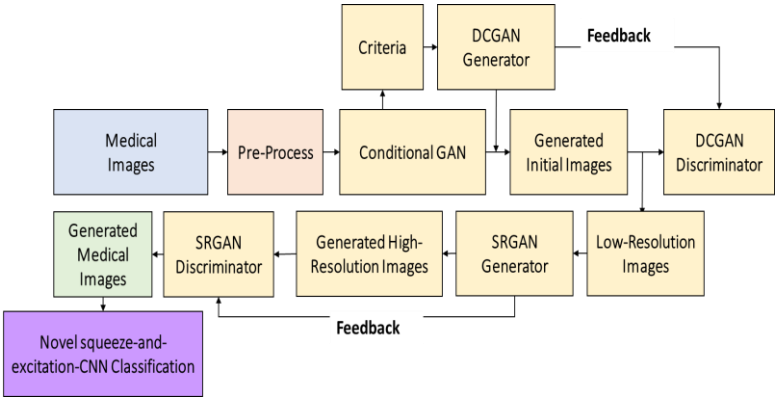


Figure 2. Novel Hybrid GAN Block Diagram

2.1. Input Datasets

To perform the experimental analysis, five publicly available medical imaging datasets representing five different disease categories and imaging modalities were used, ensuring a range of diagnostic properties and visual complexity. The Diabetic Retinopathy data set is made up of high-resolution color retinal fundus, which was categorized into five levels of severity, i.e., No Diabetic Retinopathy, Mild, Moderate, Severe, and Proliferative stage, with an original size of the images of about 540 x 540 pixels; to ensure consistency in the processing process, all the images were resized to 256 x 256 pixels and intensity-normalized to a range of 0-1. The Pneumonia Chest X-ray data consists of 5,863 grayscale radiographic images categorized into two groups: 4,273 pneumonia cases and 1,590 standard samples. The images were rescaled to 256 x 256 and contrast-normalized to cover the lung regions and minimize illumination differences across cases. To analyze brain tumors, a set of 2,053 T1-weighted magnetic resonance images, categorized as usual or tumorous, was used, and preprocessing included skull stripping, down sampling to 224 x 224 pixels, and intensity normalization to account for scanner-related variations. The Skin Cancer dataset consists of 3,297 dermoscopic images: 2,236 with benign lesions and 1,061 with malignant lesions. To maintain sample uniformity, the images were reduced to 256 x 256 pixels, and hair artefact images were removed and color normalized. Also, the Leukemia Cell Image dataset consists of microscopic images categorized as benign, early leukemia, and pre-leukemia images. To ensure that images of different stains were comparable, all images were resized to 128 x 128 pixels and color normalized. To address minor class imbalances observed during training, appropriate class-balancing measures were applied across all datasets. Moreover, uniform preprocessing procedures were used for all datasets to enhance consistency, reproducibility, and stable model optimization. Together, these datasets represent a rich array of imaging modalities, including fundus photography, radiography, magnetic resonance imaging, Dermoscopy, and microscopy, and pose diverse diagnostic challenges, thus offering a diverse assessment environment for the strength and generalizability of medical image analysis models across heterogeneous clinical conditions.

2.2. Initial Generation using DCGAN

A Deep Convolutional GAN (DCGAN) is developed to generate synthetic medical images while preserving spatial features. The model includes a generator that converts random noise sampled from $N(0,1)$ into artificial images and a discriminator that differentiates real images from generated ones. Through adversarial training, the generator gradually learns the underlying data distribution. The DCGAN is optimized using an adversarial objective function.

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

The discriminator is trained to increase the loss value, whereas the generator attempts to reduce it. Model training is performed using the Adam optimizer with a batch size of 16 over 1000 training epochs.

2.3. Conditional GAN for Semantic Control

Once the initial generation stage is complete, a conditional GAN (cGAN) is employed to incorporate label-specific control into image synthesis. By enhancing the cGAN (z, y)formulation, both the generator and discriminator utilize class information to produce clinically relevant category-specific images. This approach enables the generation of distinct disease-wise imaging data for improved visualization and diagnosis. To handle specific scenarios, the adversarial loss function is further modified.

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

The cGAN is refined using class-wise subsets of the training data. The discriminator processes paired image–label inputs, while the generator incorporates label information through embedding layers. Conditioning labels are modality specific: binary labels are used for Pneumonia, Skin Cancer, and Leukemia datasets, whereas multi-class labels represent severity stages in Diabetic Retinopathy and Brain Tumor datasets. Embedded class labels are combined with latent noise to generate diagnostically consistent synthetic images across modalities.

2.4. Criteria Module for Image Quality

The quality of synthetic images is evaluated using structural and contrast-based attributes. A non-trainable criteria module is applied after conditional generation and before super-resolution. Image quality is measured using entropy, gradient magnitude, and Laplacian variance to assess texture, sharpness, and edge clarity. Only top-ranked images are retained, limiting artifact propagation. The highest 20% of samples are used to create low- and high-resolution pairs for SRGAN training, improving stability and perceptual quality with minimal computational cost.

2.5. High-Resolution Output Generation

An SRGAN is employed to improve the visual clarity and sharpness of the selected synthetic images. Its discriminator distinguishes real high-resolution images from generated ones, while the generator reconstructs high-resolution outputs from low-resolution inputs to ensure realism. Training of the SRGAN typically relies on a combined loss function to guide perceptual and reconstruction quality.

$$\mathcal{L}_{SRGAN} = \mathcal{L}_{adv} + \lambda_1\mathcal{L}_{content} + \lambda_2\mathcal{L}_{perceptual}$$

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

This approach aims to reduce generator loss while keeping discriminator loss stable through careful tuning of the SRGAN generator. Producing visually realistic and diagnostically reliable images is essential for both expert training and future clinical AI use. Low- and high-resolution image pairs are created by down-sampling high-resolution images to supervise SRGAN learning. Applied after conditional generation and quality screening, the SRGAN improves spatial detail while maintaining consistent training and evaluation.

2.6. Generator Architecture

Table 1 presents a unified hybrid generator architecture applied across all datasets. The model combines DCGAN, cGAN, and SRGAN within a single framework, while training weights separately for each dataset without cross-dataset sharing. The Conv2D generator outputs 64×64×3 synthetic images conditioned on input data. The discriminator processes stacked image–condition pairs (64×64×6) through successive convolutional layers with LeakyReLU activation, followed by a fully connected layer that outputs a single authenticity score.

Table 1. Architecture of Generator

Block No.	Processing Stage	Layer Type	Kernel / Units	Stride	Activation	Output Description
G1	Input Stage	Random Noise + Feature Input	–	–	–	Latent vector combined with medical feature input
G2	Feature Projection	Fully Connected Layer	1024 units	–	ReLU	High-dimensional feature mapping
G3	Reshape Layer	Tensor Reshaping	–	–	–	Converts vector to spatial feature map
G4	Upsampling Block-1	Transposed Convolution	3×3	2	ReLU + BatchNorm	Spatial resolution enhancement
G5	Upsampling Block-2	Transposed Convolution	3×3	2	ReLU + BatchNorm	Further feature expansion
G6	Feature Refinement	Convolution Layer	3×3	1	ReLU	Texture and edge learning
G7	Output Mapping	Convolution Layer	1×1	1	Tanh	Pixel-level image generation
G8	Generator Output	Synthetic Image	–	–	–	Generated medical image

2.7. Discriminator Architecture

As shown in Table 2, the discriminator is able to differentiate real and synthetic images even without batch normalization. Although MedSynGAN generates high-quality medical images across multiple modalities, it remains susceptible to overfitting, particularly with small or imbalanced datasets. Hybrid design improves generalization but does not fully prevent sample memorization. Scalability to large, multi-institutional datasets and high-resolution or cross-device data remains unverified. Practical clinical deployment may be constrained by training cost and complexity, highlighting the need for model compression, continual learning, and domain adaptation strategies.

Table 2. Architecture of Discriminator

Block No.	Processing Stage	Layer Type	Kernel / Units	Stride	Activation	Functional Role
D1	Input Stage	Image Input Layer	–	–	–	Accepts real or generated image
D2	Feature Extraction-1	Convolution Layer	3×3	2	LeakyReLU	Low-level feature detection
D3	Feature Extraction-2	Convolution Layer	3×3	2	LeakyReLU + BatchNorm	Mid-level pattern learning
D4	Feature Extraction-3	Convolution Layer	3×3	2	LeakyReLU + BatchNorm	High-level semantic extraction
D5	Flattening	Flatten Layer	–	–	–	Converts feature maps to vector
D6	Classification Layer	Fully Connected Layer	1 unit	–	Sigmoid	Real vs Fake decision

D7	Discriminator Output	Probability Score	–	–	–	Authenticity estimation

2.8. Novel Classification Modelling

As illustrated in Figure 3, the proposed architecture incorporates Squeeze-and-Excitation (SE) blocks within a convolutional neural network (CNN) to significantly enhance feature representation and overall model effectiveness. The SE blocks are systematically embedded after selected convolutional layers, allowing the network to adaptively recalibrate channel-wise feature responses. By selectively amplifying informative channels and suppressing less relevant ones, the network becomes more capable of capturing discriminative patterns that are critical for accurate medical image analysis. This channel attention mechanism operates in a lightweight yet powerful manner, improving feature sensitivity without introducing substantial computational burden. As medical images often contain subtle and localized visual cues, such adaptive feature refinement plays a crucial role in improving classification robustness and reliability. The inclusion of SE blocks enables the CNN to focus on the most diagnostically meaningful information, thereby enhancing prediction accuracy while maintaining efficiency.

The SE module functions through three key stages. First, the squeeze operation performs global spatial information aggregation using Global Average Pooling (GAP). This step compresses each feature map across its spatial dimensions, producing a compact channel-wise descriptor. By summarizing global contextual information, the network gains an overall understanding of feature relevance across channels. Second, the excitation stage models inter-channel dependencies through a small fully connected bottleneck structure. The compressed descriptor is passed through two dense layers. The first layer reduces dimensionality using a predefined reduction ratio (commonly set to 16) and applies a ReLU activation to introduce nonlinearity. The second layer restores the original channel dimension and uses a sigmoid activation to generate normalized channel importance weights. Finally, the scaling stage applies these learned weights to the original feature maps through element-wise multiplication. This step adaptively emphasizes channels that contribute most to the learning task while diminishing the influence of less informative features. As a result, the network dynamically learns "what" to focus on, improving its ability to extract meaningful representations from complex medical images. Overall, the integration of SE blocks enhances model interpretability, stability, and classification performance, making the proposed approach particularly suitable for medical imaging tasks where accuracy and efficiency are both critical.

Advantages:

1. **Improved Feature Learning:** SE blocks enhance representation by modeling channel-wise relationships, giving higher importance to informative channels and enabling clearer feature discrimination.
2. **Computational Efficiency:** SE blocks introduce only a small number of extra parameters, governed by the channel size and reduction ratio, ensuring an effective balance between performance and complexity.
3. **Stronger Generalization:** Adaptive channel attention allows the network to better handle data variability, improving robustness and performance on unseen samples.
4. **Easy Integration:** The modular design of SE blocks enables seamless incorporation into existing CNN architectures. In this approach, they are placed after convolutional blocks to refine intermediate feature maps.
5. **Performance Enhancement:** SE blocks consistently improve evaluation metrics such as accuracy and F1-score, especially for datasets with complex or noisy patterns.
6. **Overfitting Reduction:** By recalibrating feature importance, SE blocks limit the influence of irrelevant patterns, helping to reduce overfitting during training.

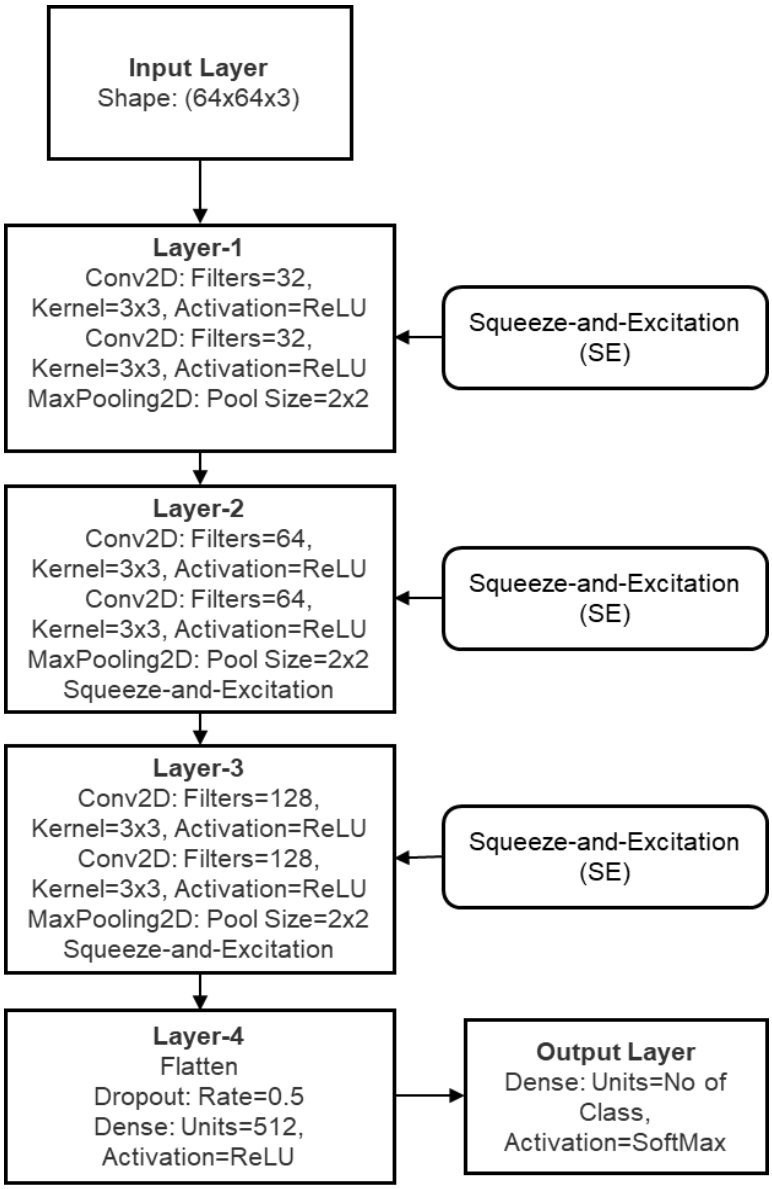


Figure 3. Architecture of SE-CNN Classification Model

3. Results

The experimental framework was structured to support reliable training and evaluation of several GAN variants, including DCGAN, cGAN, CycleGAN, SRGAN, and the proposed Hybrid GAN. All experiments were conducted on Google Colab using NVIDIA T4 GPUs, enabling efficient acceleration of deep learning workloads without requiring specialized local hardware. This setup ensured a practical and cost-effective environment for large-scale experimentation. The models were evaluated on multiple medical imaging domains, including diabetic retinopathy, pneumonia, brain tumors, skin cancer, and leukemia, to comprehensively examine their capability to synthesize realistic and clinically relevant images.

As summarized in **Table 3**, all networks were trained for 1000 epochs with a batch size of 16 to ensure sufficient exposure to the data and effective feature learning. A fixed learning rate of 0.0001 and a momentum coefficient (β) of 0.5 were used to promote stable convergence. The Adam optimizer was selected due to its adaptive learning behavior and proven effectiveness in GAN training. Model performance was assessed using generator and discriminator loss values, which reflect adversarial stability and image quality. Maintaining identical training settings across all models enabled fair comparison, demonstrating the hybrid model's advantages in training efficiency, computational balance, and image synthesis quality for practical clinical applications.

Table 3. Hyper Parameters of Hybrid GAN

Training Setting	Configured Value	Role in the Model
Mini-batch size	16	Controls how many samples are processed simultaneously, influencing convergence behavior and training stability.
Total training iterations (epochs)	1000	Defines how many times the model learns from the entire dataset.
Objective function	Binary cross-entropy with logits	Evaluates the discriminator's ability to correctly classify authentic and generated images.
Generator optimization method	Adam optimizer	Updates generator parameters to improve the realism of synthesized images.
Discriminator optimization method	Adam optimizer	Adjusts discriminator weights to enhance real–fake image discrimination.
Step size (learning rate)	1×10^{-4}	Regulates the magnitude of weight updates during backpropagation.
Momentum coefficient (β_1)	0.5	Balances past gradient influence to ensure smoother and more stable training.

As shown in Table 4, the SE-CNN training parameters were chosen to ensure stable and efficient learning, using a batch size of 64, 50 training epochs, Binary Crossentropy loss, and the Adam optimizer. A learning rate of 0.001 was applied to balance convergence speed and optimization stability while enabling effective feature learning.

Table 4. Hyper Parameters of Novel SE-CNN

Training Setting	Configured Value	Role in the Model
Mini-batch size	64	Influences gradient consistency and overall learning stability during training.
Number of epochs	50	Specifies how many times the model iterates over the complete training set.
Loss criterion	Binary cross-entropy with logits	Quantifies the prediction error between actual and estimated class labels in a binary classification setting.
Optimization algorithm	Adam optimizer	Utilizes adaptive learning rates and momentum to accelerate and stabilize convergence.
Learning step size	1×10^{-3}	Controls the scale of parameter updates during backpropagation.

3.1. Generation Images

Figure 4 illustrates the capability of the proposed Hybrid GAN to synthesize realistic fundus images from the Diabetic Retinopathy dataset. The generated samples closely resemble real retinal images by accurately preserving essential anatomical components such as the optic disc, macula, and vascular structures. The outputs demonstrate natural color consistency, uniform lighting, and detailed textures, retaining subtle vascular cues that are vital for diabetic retinopathy assessment.

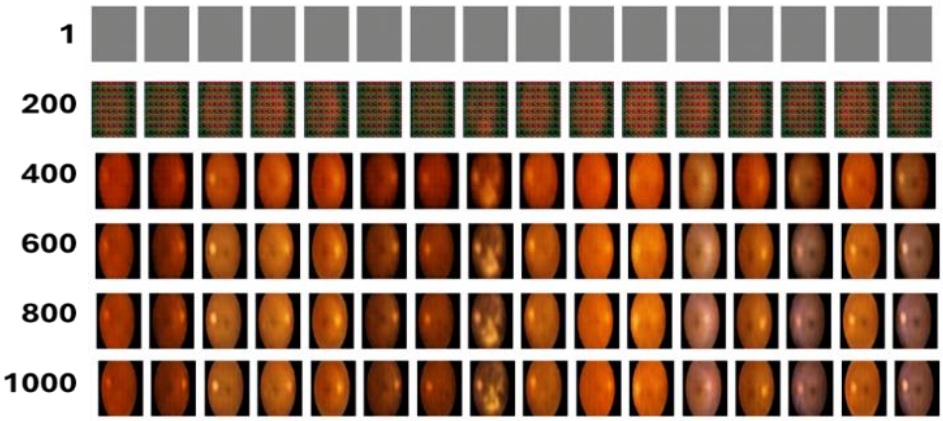


Figure 4. Image Generation for DR

Figures 5 showcases additional synthetic fundus images generated by the proposed Hybrid GAN for the Diabetic Retinopathy dataset. The generated outputs maintain high visual fidelity, accurately reproducing retinal landmarks such as the optic disc, macular region, and vascular network. Consistent illumination, natural color tones, and preserved fine-grain textures demonstrate the model's ability to generate clinically reliable retinal images.

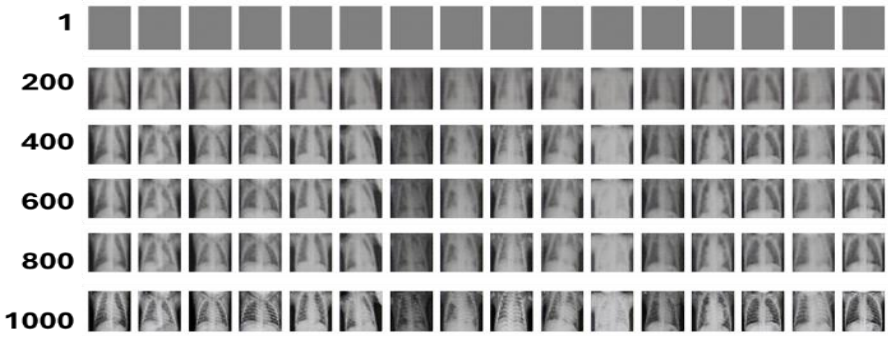


Figure 5. Image Generation for Pneumonia x-ray

Figure 6 further validates the robustness of the Hybrid GAN by presenting another set of synthesized retinal images. The model effectively captures subtle vascular structures and anatomical continuity, ensuring realistic appearance. The clarity and structural coherence of these images highlight the suitability of the generated samples for diagnostic evaluation and training purposes.

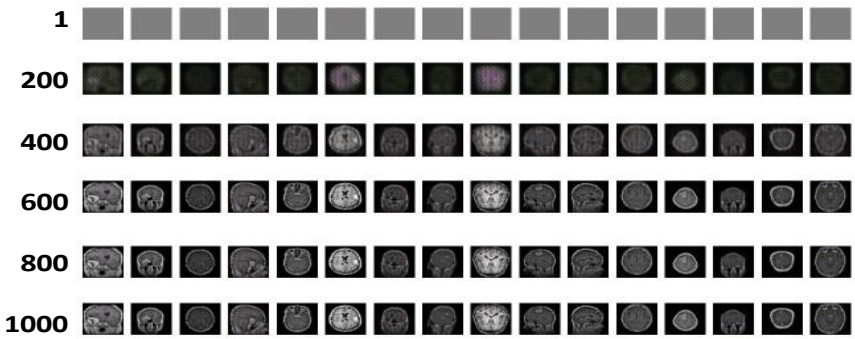


Figure 6. Image Generation for brain tumor MRI

Figure 7 Hybrid GAN continues to demonstrate stable performance, producing fundus images with well-defined anatomical features and uniform lighting conditions. The preservation of delicate vessel patterns and texture details reflects the model's strength in modeling complex retinal characteristics associated with diabetic retinopathy.

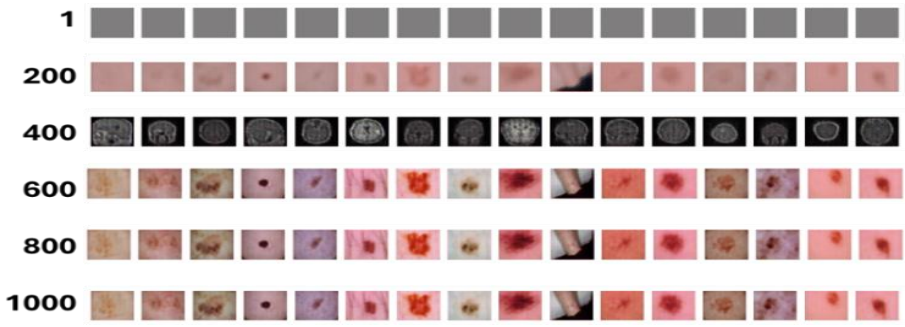


Figure 7. Image Generation for Skin cancer

Figure 8 illustrates consistent synthetic image quality across multiple generations. The Hybrid GAN successfully maintains color balance, spatial detail, and anatomical accuracy, reinforcing its reliability for large-scale retinal image synthesis and data augmentation in medical imaging workflows.

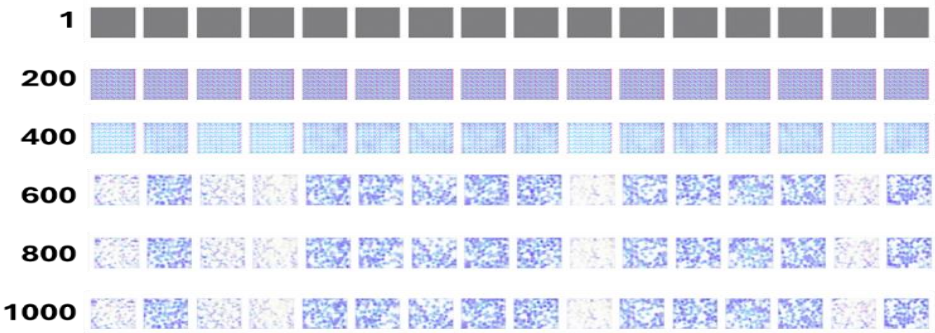
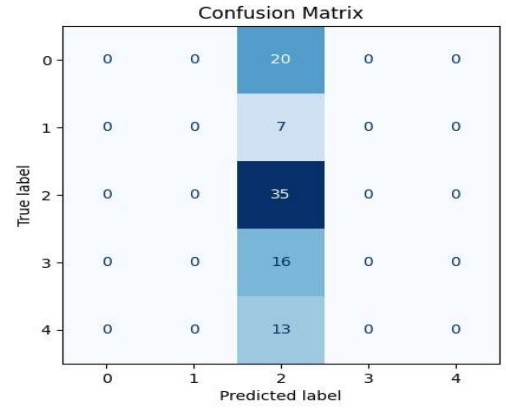


Figure 8. Image Generation for Leukemia cancer

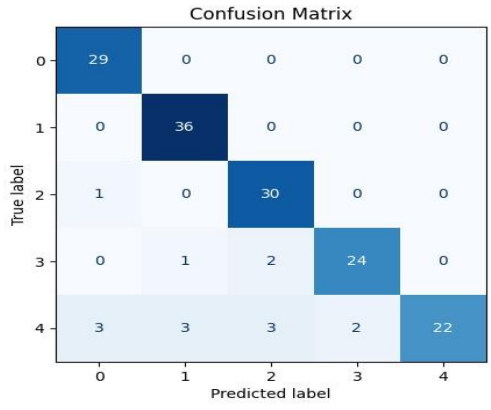
3.2. Classification Using SE-CNN

Figure 9 compares confusion matrices obtained without and with Hybrid GAN augmentation.



	precision	recall	f1-score	support
0	0.00	0.00	0.00	20
1	0.00	0.00	0.00	7
2	0.38	1.00	0.56	35
3	0.00	0.00	0.00	16
4	0.00	0.00	0.00	13
accuracy			0.38	91
macro avg	0.08	0.20	0.11	91
weighted avg	0.15	0.38	0.21	91

(a) Without GAN

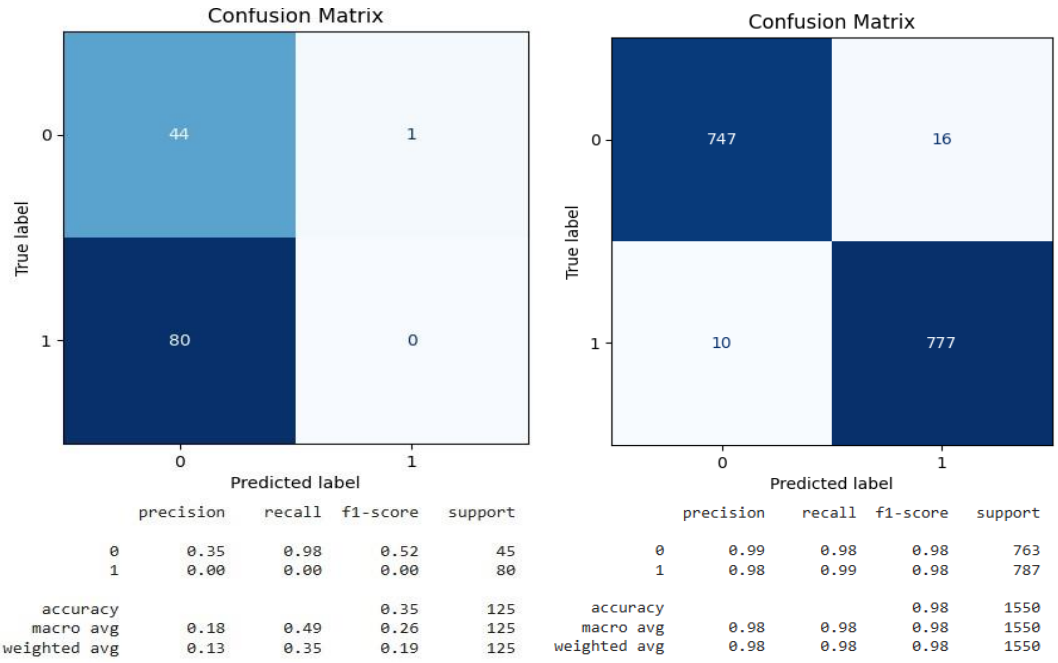


	precision	recall	f1-score	support
0	0.88	1.00	0.94	29
1	0.90	1.00	0.95	36
2	0.86	0.97	0.91	31
3	0.92	0.89	0.91	27
4	1.00	0.67	0.80	33
accuracy			0.90	156
macro avg	0.91	0.90	0.90	156
weighted avg	0.91	0.90	0.90	156

(b) With Hybrid GAN

Figure 9. Classification of Diabetic Retinopathy (a) Without GAN (b) With Hybrid GAN

Figure 10 highlights a similar trend, where the baseline model struggles with class misclassification and achieves low accuracy (0.35). When trained with Hybrid GAN-augmented data, the classifier demonstrates balanced performance across classes, achieving a significantly higher accuracy of 0.98.



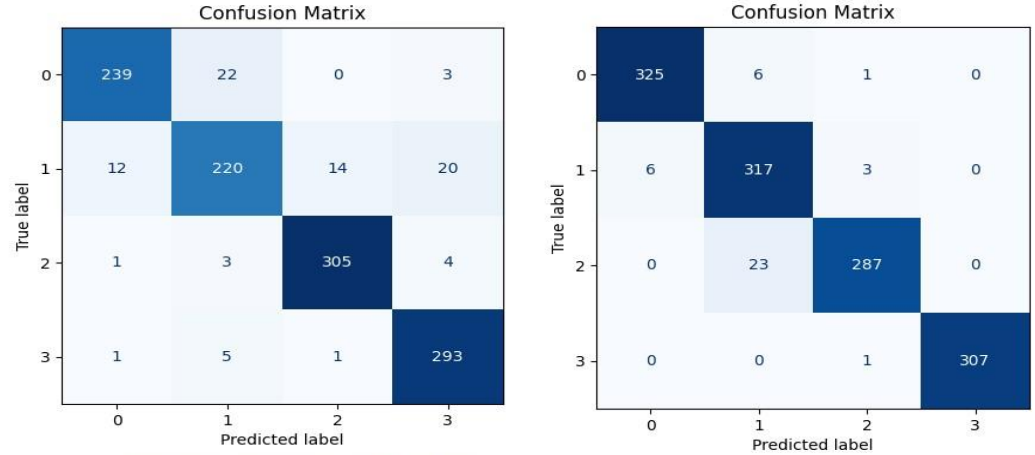
	precision	recall	f1-score	support
0	0.35	0.98	0.52	45
1	0.00	0.00	0.00	80
accuracy			0.35	125
macro avg	0.18	0.49	0.26	125
weighted avg	0.13	0.35	0.19	125

	precision	recall	f1-score	support
0	0.99	0.98	0.98	763
1	0.98	0.99	0.98	787
accuracy			0.98	1550
macro avg	0.98	0.98	0.98	1550
weighted avg	0.98	0.98	0.98	1550

(a) Without Hybrid GAN (b) With Hybrid GAN

Figure 10. Classification of Pneumonia (a) Without GAN (b) With Hybrid GAN

Figure 11 the use of Hybrid GAN improves classification consistency and reliability. While the baseline model exhibits uneven predictions, the Hybrid GAN-enhanced model achieves improved class separation and a higher overall accuracy of 0.97.



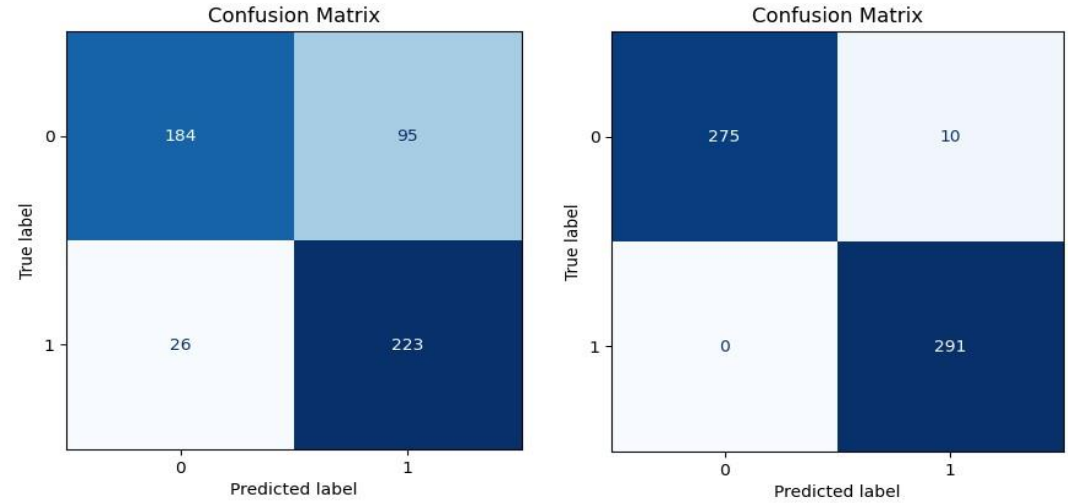
	precision	recall	f1-score	support
0	0.94	0.91	0.92	264
1	0.88	0.83	0.85	266
2	0.95	0.97	0.96	313
3	0.92	0.98	0.95	300
accuracy			0.92	1143
macro avg	0.92	0.92	0.92	1143
weighted avg	0.92	0.92	0.92	1143

	precision	recall	f1-score	support
0	0.98	0.98	0.98	332
1	0.92	0.97	0.94	326
2	0.98	0.93	0.95	310
3	1.00	1.00	1.00	308
accuracy			0.97	1276
macro avg	0.97	0.97	0.97	1276
weighted avg	0.97	0.97	0.97	1276

(a) Without Hybrid GAN (b) With Hybrid GAN

Figure 11. Classification of Brain Tumor (a) Without GAN (b) With Hybrid GAN

Figure 12 presents confusion matrix results for skin cancer classification. Without Hybrid GAN augmentation, the model performance is limited (accuracy 0.77). With Hybrid GAN support, classification accuracy increases to 0.98, along with notable gains in precision, recall, and F1-score, confirming its effectiveness in complex cancer classification tasks.



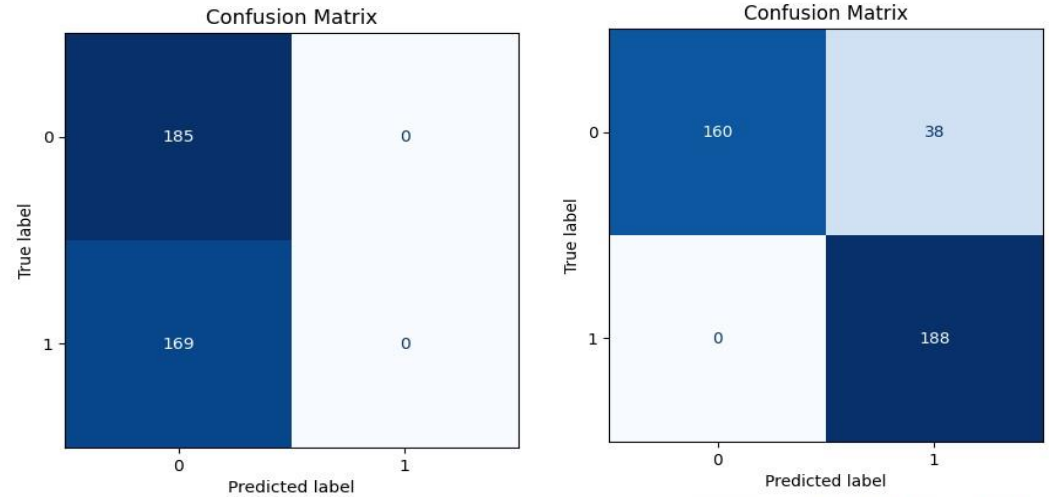
	precision	recall	f1-score	support
0	0.88	0.66	0.75	279
1	0.70	0.90	0.79	249
accuracy			0.77	528
macro avg	0.79	0.78	0.77	528
weighted avg	0.79	0.77	0.77	528

	precision	recall	f1-score	support
0	1.00	0.96	0.98	285
1	0.97	1.00	0.98	291
accuracy			0.98	576
macro avg	0.98	0.98	0.98	576
weighted avg	0.98	0.98	0.98	576

(a) Without Hybrid GAN (b) With Hybrid GAN

Figure 12. Classification of Skin (a) Without GAN (b) With Hybrid GAN

Figure 13 demonstrates improved classification outcomes with Hybrid GAN integration. The baseline model shows weak performance with an accuracy of 0.52, whereas the Hybrid GAN-augmented model achieves balanced predictions across all classes and a substantially higher accuracy of 0.90, emphasizing the framework's robustness across datasets.



	precision	recall	f1-score	support
0	0.52	1.00	0.69	185
1	0.00	0.00	0.00	169
accuracy			0.52	354
macro avg	0.26	0.50	0.34	354
weighted avg	0.27	0.52	0.36	354

	precision	recall	f1-score	support
0	1.00	0.81	0.89	198
1	0.83	1.00	0.91	188
accuracy			0.90	386
macro avg	0.92	0.90	0.90	386
weighted avg	0.92	0.90	0.90	386

(a) Without Hybrid GAN (b) With Hybrid GAN

Figure 13. Classification of Leukemia (a) Without Hybrid GAN (b) With Hybrid GAN

4. Discussion

Table 5 summarizes generator and discriminator loss values across five medical imaging datasets and different GAN architectures. These losses are reported to reflect training stability rather than direct image quality. The Hybrid GAN demonstrates the most stable convergence behavior, achieving lower loss values,

particularly for Diabetic Retinopathy and Leukemia datasets, indicating reliable and efficient adversarial training.

Table 5. GANs Losses Analysis

Medical Modality	DCGAN (G/D)	C-GAN (G/D)	CycleGAN (G/D)	SRGAN (G/D)	Proposed MixGANMed (G/D)
Diabetic Retinopathy	4.81 / 4.46	4.61 / 4.42	2.40 / 2.02	4.63 / 4.09	0.52 / 0.43
X-ray Pneumonia	0.59 / 0.29	0.52 / 0.22	0.69 / 0.38	0.70 / 0.40	1.21 / 1.21
Brain Tumor MRI	0.78 / 0.47	0.76 / 0.45	1.20 / 0.65	1.42 / 0.82	1.54 / 1.75
Skin Cancer	6.24 / 6.25	6.22 / 6.21	4.21 / 4.12	6.25 / 6.12	1.01 / 1.22
Blood Cell Leukemia Cell	5.56 / 5.36	5.23 / 5.23	3.33 / 3.13	5.23 / 5.10	0.29 / 0.22

*G: Generator Loss, D: Discriminator Loss

Table 6 reports image quality evaluation metrics. PSNR and SSIM are used exclusively for assessing the super-resolution stage, where paired low- and high-resolution images are available. For overall realism and distribution-level similarity of generated images, Fréchet Inception Distance (FID) is adopted as the primary evaluation metric, as it measures the closeness between real and synthetic image distributions.

Table 6. GANs Image Quality Analysis

Dataset	Model	Avg FID	Avg IS	Avg PSNR (dB)	Avg SSIM
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	DCGAN	445.96	1.36	14.57	0.50
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	CycleGAN	413.28	1.11	13.77	0.50
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	SRGAN	417.18	1.90	13.00	0.48
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	Proposed GAN	30.402	4.642	36.742	0.93

Table 7 compares computational complexity in terms of training time and parameter count. The Hybrid GAN exhibits a lightweight and efficient design, requiring minimal training time and fewer parameters compared to other GAN variants, particularly SRGAN, which demands significantly higher computational resources.

Table 7. GANs Complexity Analysis

Dataset	Model	Avg Training Time (hrs)	Trainable Parameters
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	DCGAN	5.0	1.2M
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	CycleGAN	7.0	1.8M
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	SRGAN	8.0	2.1M
DR, X-ray, Brain MRI, Skin Cancer and Leukemia	Proposed GAN	0.96	0.67M

Table 8 provides a comparison between the proposed Hybrid GAN and existing state-of-the-art methods across multiple medical imaging modalities. While prior approaches achieve reasonable performance, the proposed model consistently delivers superior image quality metrics while maintaining low training cost, demonstrating both effectiveness and efficiency.

Table 8. Comparative Analysis of Existing Systems

Method	Image Domain	Dataset	Epochs	PSNR	SSIM	FID	IS	Time
DRForecastGAN [1]	Color	DR	200	16.9	0.65	34.4	–	–
GAN-VSP [2]	Color	DR	10k	35.10	0.90	–	0.12	–
IFGAN [3]	Gray	MRI	100	31.11	0.93	–	–	15.9h
Pix2Pix GAN [5]	Gray	MRI, PET	1000	27.11	0.89	–	–	–
Proposed Hybrid GAN	Color + Gray	Multi-dataset	1000	36.74	0.93	30.40	4.64	0.75h

Table 9 presents classification performance with and without Hybrid GAN augmentation. Across all five datasets, the Hybrid GAN markedly enhances classification accuracy and robustness, achieving near-optimal results in several cases. These findings confirm the Hybrid GAN's strong contribution to improving both synthetic image generation and downstream medical image classification tasks.

Table 9. Performance Comparison with and Without Hybrid GAN

Dataset	Without Hybrid GAN				With Hybrid GAN			
	ACC	P	R	F1	ACC	P	R	F1
Brain Tumor	0.91	0.91	0.92	0.91	0.97	0.97	0.97	0.97
Diabetic Retinopathy	0.37	0.07	0.19	0.10	0.89	0.91	0.90	0.90
Skin Cancer	0.76	0.78	0.78	0.76	0.98	0.98	0.98	0.98
Pneumonia	0.36	0.17	0.49	0.25	0.98	0.98	0.98	0.98
Skin Cancer	0.76	0.78	0.78	0.76	0.98	0.98	0.98	0.98

5. Conclusions

A novel hybrid generative adversarial framework is introduced to produce high-quality medical images across multiple diagnostic domains. The proposed approach integrates the strengths of well-established GAN variants, including DCGAN, cGAN, and SRGAN, effectively addressing key limitations of traditional methods such as training instability, slow convergence, and inadequate structural detail preservation. The model is extensively validated on five benchmark medical datasets—Diabetic Retinopathy, Pneumonia, Brain Tumors, Skin Cancer, and Leukemia—where it consistently surpasses conventional GAN models in terms of training stability, generator and discriminator loss behavior, and overall image realism. Both qualitative and quantitative evaluations confirm that the synthesized images closely preserve critical anatomical characteristics while minimizing visual artifacts, ensuring strong clinical relevance for diagnostic applications. In addition, the framework demonstrates notable computational efficiency, requiring only 0.75 hours of training with 0.67 million parameters, achieving a favorable balance between resource usage and synthesis quality. Classification performance is further enhanced by integrating a Squeeze-and-Excitation–based CNN, which leads to substantial improvements in accuracy, precision, recall, and F1-score across all evaluated datasets, supporting reliable automated medical diagnosis.

Future work may extend this framework to additional imaging modalities such as CT and PET, as well as rare disease datasets. Incorporating explainable AI techniques could further improve model transparency and trustworthiness, while federated learning strategies would enable secure multi-institutional training. These advancements are expected to broaden real-world clinical applicability and strengthen the framework's impact on medical image synthesis and diagnostic automation.

Funding: This research received no external funding.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. You might choose to exclude this statement if the study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. U. Akbar, W. Wang, and A. Eklund, "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, vol. 6, no. 1, p. 15022, 2023, doi: 10.1088/2632-2153/ad9a3a.
2. K. Afnaan, T. Singh, and P. Duraisamy, "Hybrid deep learning framework for bidirectional medical image synthesis," in Proc. 15th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT), 2024, pp. 1–6, doi: 10.1109/ICCCNT61001.2024.10725975.
3. M. Ali, M. Ali, M. Hussain, and D. Koundal, "Generative adversarial networks (GANs) for medical image processing: Recent advancements," Arch. Comput. Methods Eng., pp. 1–14, 2024, doi: 10.1007/s11831-024-10174-8.
4. Y. Chen et al., "ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation," Biomed. Signal Process. Control, vol. 92, p. 106100, 2024, doi: 10.1016/j.bspc.2024.106100.
5. Y. S. Devi and S. P. Kumar, "Diabetic retinopathy (DR) image synthesis using DCGAN and classification of DR using transfer learning approaches," Int. J. Image Graph., vol. 24, no. 5, p. 2340009, 2023, doi: 10.1142/S0219467823400090.
6. E. Dugas, J. Jared, Jorge, and W. Cukierski, "Diabetic retinopathy detection," Kaggle, 2015. [Online]. Available: https://kaggle.com/competitions/diabetic-retinopathy-detection. [Accessed: Feb. 9, 2024].
7. A. Fanconi, "Skin cancer: Malignant vs. benign," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign. [Accessed: Apr. 21, 2025].
8. A. S. Fard, D. C. Reutens, S. C. Ramsay, S. J. Goodman, S. Ghosh, and V. Vegh, "Image synthesis of interictal SPECT from MRI and PET using machine learning," Front. Neurol., vol. 15, p. 1383773, 2024, doi: 10.3389/fneur.2024.1383773.
9. P. Friedrich, Y. Frisch, and P. C. Cattin, "Deep generative models for 3D medical image synthesis," in Generative Machine Learning Models in Medical Image Computing, Springer, 2024, pp. 255–278. [Online]. Available: http://arxiv.org/abs/2410.17664.
10. M. Hamghalam and A. L. Simpson, "Medical image synthesis via conditional GANs: Application to segmenting brain tumours," Comput. Biol. Med., vol. 170, p. 107982, 2024, doi: 10.1016/j.compbiomed.2024.107982.
11. Y. Heng, M. Yinghua, F. G. Khan, A. Khan, and Z. Hui, "HLSNC-GAN: Medical image synthesis using hinge loss and switchable normalization in CycleGAN," IEEE Access, vol. 12, pp. 55448–55464, 2024, doi: 10.1109/ACCESS.2024.3390245.
12. S. Islam et al., "Generative adversarial networks (GANs) in medical imaging: Advancements, applications, and challenges," IEEE Access, vol. 12, pp. 35728–35753, 2024, doi: 10.1109/ACCESS.2024.3370848.
13. A. Jha and H. Iima, "CT to MRI image translation using CycleGAN: A deep learning approach for cross-modality medical imaging," in Proc. Int. Conf. Agents Artif. Intell., vol. 3, pp. 951–957, 2024, doi: 10.5220/0012422900003636.
14. D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122–1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.
15. M. Kumar, A. S. Chivukula, and G. Barua, "Deep learning-based encryption scheme for medical images using DCGAN and virtual planet domain," Sci. Rep., vol. 15, no. 1, p. 1211, 2025, doi: 10.1038/s41598-024-84186-6.
16. S. Madhav, T. M. Nandhika, and M. K. K. Devi, "Super resolution of medical images using SRGAN," in Proc. 2nd Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE), 2024, pp. 1–6, doi: 10.1109/ic-ETITE58242.2024.10493588.
17. A. A. Mamo, B. G. Gebresilassie, A. Mukherjee, V. Hassija, and V. Chamola, "Advancing medical imaging through generative adversarial networks: A comprehensive review and future prospects," Cogn. Comput., vol. 16, no. 5, pp. 2131–2153, 2024, doi: 10.1007/s12559-024-10291-3.
18. P. Mooney, "Chest X-ray images (Pneumonia)," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. [Accessed: Feb. 9, 2024].
19. P. Nandal, S. Pahal, A. Khanna, and P. R. Pinheiro, "Super-resolution of medical images using real ESRGAN," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3497002.
20. N. Paul, "Brain MRI images for brain tumor detection," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection. [Accessed: Apr. 21, 2025].
21. R. Raad et al., "Conditional generative learning for medical image imputation," Sci. Rep., vol. 14, no. 1, p. 171, 2024, doi: 10.1038/s41598-023-50566-7.
22. M. C. S. Akhil, B. S. S. Sharma, A. Kodipalli, and T. Rao, "Medical image synthesis using DCGAN for chest X-ray images," in Proc. Int. Conf. Knowl. Eng. Commun. Syst. (ICKECS), vol. 1, no. 1, pp. 1–8, 2024, doi: 10.1109/ICKECS61492.2024.10617031.

23. M. K. Sherwani and S. Gopalakrishnan, "A systematic literature review: Deep learning techniques for synthetic medical image generation and their applications in radiotherapy," Front. Radiol., vol. 4, p. 1385742, 2024, doi: 10.3389/fradi.2024.1385742.
24. D. Shah, M. A. U. Khan, and M. Abrar, "Reliable breast cancer diagnosis with deep learning: DCGAN-driven mammogram synthesis and validity assessment," Appl. Comput. Intell. Soft Comput., vol. 2024, p. 1122109, 2024, doi: 10.1155/2024/1122109.
25. D. N. Sindhura, R. M. Pai, S. N. Bhat, and M. M. M. Pai, "A review of deep learning and generative adversarial networks applications in medical image analysis," Multimed. Syst., vol. 30, no. 3, p. 161, 2024, doi: 10.1007/s00530-024-01349-1.
26. V. Tanwar, "Leukemia cancer small dataset," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/visheshtanwar26/leukemia-cancer-small-dataset/data. [Accessed: Apr. 21, 2025].
27. S. Varshitha, N. Lavanya, M. Shirisha, S. Manmatti, K. P. A. Rani, and S. Gowrishankar, "Enhancing medical imaging resolution: Exploring SRGAN for high-quality medical image reconstruction," in Proc. IEEE Int. Conf. Women Innov. Technol. Entrepreneurship (ICWITE), 2024, pp. 1–8, doi: 10.1109/ICWITE59797.2024.10503215.
28. J. Wang et al., "Self-improving generative foundation model for synthetic medical image generation and clinical applications," Nat. Med., vol. 31, no. 2, pp. 609–617, 2025.
29. R. Wang, A. F. Heimann, M. Tannast, and G. Zheng, "CycleSGAN: A cycle-consistent and semantics-preserving generative adversarial network for unpaired MR-to-CT image synthesis," Comput. Med. Imaging Graph., vol. 117, p. 102431, 2024, doi: 10.1016/j.compmedimag.2024.102431.
30. R. Zakaria, H. Abedlamjid, D. Zitouni, and A. Elqaraoui, "Medical-DCGAN: Deep convolutional GAN for medical imaging," in Advances in Emerging Financial Technology and Digital Money, CRC Press, 2024, pp. 123–134, doi: 10.1201/9781032667478-10.