I. J. Engineering and Manufacturing, **2025, 3, 56-68**
Published Online on June 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2025.03.05

Medical Image Synthesis Using Variational Autoencoder and Generative Adversarial Networks

Sinchana Ganesh
Department of Computer Applications, JSS Science and Technology University, India
Email: sinchanaganesh6@gmail.com

Madhushree B.
Department of Computer Applications, JSS Science and Technology University, India
Email: madhushree@jssstuniv.in
ORCID iD: https://orcid.org/0000-0002-6117-9468

Sowmya K. N.
Department of Information Science and Engineering, JSS Academy of Technical Education, India
Email: kn_sowmya@rediffmail.com
ORCID iD: https://orcid.org/0000-0001-7212-8677

H. R. Chennamma*
Department of Computer Applications, JSS Science and Technology University, India
Email: hrchennamma@jssstuniv.in
ORCID iD: https://orcid.org/0000-0002-3858-7834
*Corresponding Author

Received: 10 October, 2024; Revised: 16 December, 2024; Accepted: 13 March, 2025; Published: 08 June, 2025

Abstract: Nowadays, image synthesis has become essential in the medical field for lever- aging deep learning technique to improve decision- making. Our proposed research work combines Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) to synthesize medical im- ages, enhancing diagnostics, medical training, and image analysis. The model presented combines a Discriminator, and a Variational Autoencoder to capitalize on the strengths of both VAEs and GANs. The Decoder is tasked with generating synthetic medical images, the Discriminator evaluates their distinguishing factor, and the VAE learns a probabilistic mapping from input to latent space, ensuring a structured representation of underlying medical features. The training process involves a decoder creating realistic medical images, a discriminator distinguishing real from synthetic ones, and a VAE capturing meaningful data variations in the latent space. Verified on the dataset sourced from the Kaggle. The model refines its parameters iteratively using a training loop, resulting in enhanced quality and variety of generated medical images. The proposed VAE- GAN model demonstrates its efficacy by generating diverse and realistic medical images. The structured latent space contributes to interpretability, making the images suitable for purposes like data augmentation, anomaly detection, and machine learning model training.

Index Terms: Medical image synthesis, VAE, GAN, Hybrid model, Data augmentation, Healthcare applications

1. Introduction

In the constantly advancing realm of healthcare, the synthesis of lifelike medical images holds immense potential for enhancing diagnostic tools, medical training, and the efficiency of image analysis algorithms. This study introduces a pioneering approach that amalgamates the potency of Variational Autoencoders and Generative Adversarial Networks. The synergy of these two powerful models results in a novel framework for producing synthetic medical images that are both varied and lifelike, while maintaining a structured latent space. At the core of the proposed architecture comprises crucial components which is, a Variational Autoencoder (VAE) to encode medical images into a structured latent space and a Generative Adversarial Network (GAN) to enhance the quality and realism of the generated outputs. The VAE encoder extracts meaningful patterns from input images, while the decoder reconstructs them based on the learned latent representations. The GAN discriminator further refines these images by distinguishing between real and synthetic ones, improving their fidelity and clinical applicability. When applied to a dataset sourced from Kaggle, the model showcases its prowess in generating diverse and realistic medical images. The structured latent space developed by the VAE provides more than just a foundation for image generation, it also enhances interpretability. This makes the generated images particularly valuable for data expansion, identifying anomalies, and supporting machine learning training processes.

In the broader context of healthcare and medical imaging, this research highlights numerous practical applications, from enriching existing datasets to preserving patient privacy while sharing data. The potential of this model underscores its importance in addressing challenges in the ever-evolving field of medical technology.

Here are some possible uses:

1.1 Data Augmentation for Medical Image Analysis

The generated images may assist in augmenting existing medical image datasets. This augmented dataset can then be employed for training machine learning models, enhancing the model's generalization ability and robustness.

1.2 Privacy-Preserving Data Sharing

Generated medical images may help to share datasets while preserving patient privacy. Since the generated images are not created from real patient data, there's less risk of exposing sensitive information (illustrated in Figure.1). However, real-world validation is still necessary to confirm the accuracy and clinical usability of these synthetic images.

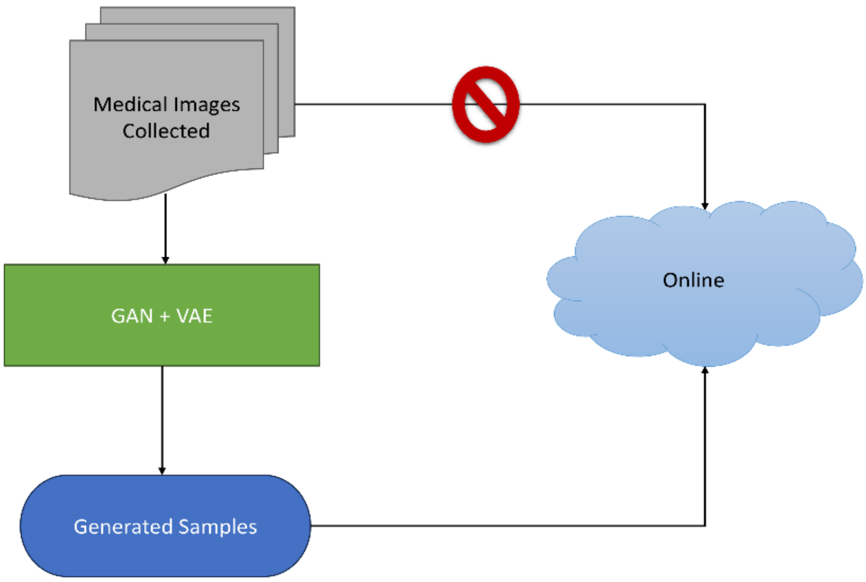


Fig. 1. Privacy Maintenance

2. Related Works

Convolutional Neural Networks (CNNs) are tailored deep learning models formulated to process structured like a gride data structures, as in images. Their effectiveness in image processing tasks is due to their capability to identify spatial patterns and hierarchical features. CNNs have gained widespread use due to the significant speed improvements provided by Graphics Processing Units (GPUs), which are hardware accelerators. The parallelism offered by GPUs enable CNNs to efficiently process large image datasets [1, 2].

Generative Adversarial Networks (GANs) entails two components: a generator and a discriminator. The generator generates synthetic data, and the discriminator evaluates the data. This adversarial training between the two leads to the generation of highly realistic data. GANs have proven effective across domains, including image generation, style transformation, and data augmentation [3, 4].

Variational Autoencoders (VAEs) represent a distinct class of generative models that establishes a probabilistic mapping across input data and a latent space. VAEs comprise an encoder that projects input data into a latent space distribution and a decoder that reconstructs data from latent samples. VAEs are notable for their potential to generate diverse outputs while maintaining a structured latent space representation [5, 6].

By combining GANs and VAEs, the strengths of both models can be harnessed. GANs are highly proficient at producing realistic data, whereas VAEs excel in creating a well-organized latent space. The integration of these two approaches aims to generate high-quality, varied, and structured synthetic data. Typically, this involves embedding the VAE within the GAN by using the discriminator for adversarial training, creating a hybrid model [7].

In GANs, the loss function consists of the following:

- Discriminator Loss: Evaluates the discriminator's capability to differentiate between real and generated data, offering feedback to boost the efficiency of both the decoder and discriminator.

For VAEs, the loss function generally includes two parts:

- Reconstruction Loss: It calculates the difference among the original input and the reconstructed output, pushing the model to accurately reconstruct input data.
- KL Divergence Loss: Regularizes the distribution in the latent space to encourage a well-structured and continuous representation [8].

Similar to our proposed work, Rguibi et al [9] have explored VAE-GAN hybrids for medical image generation, but our research advances these methods further by achieving near-perfect similarity between real and generated images, as indicated by an Fréchet inception distance (FID) score approximately zero. Further- more, our model demonstrates superior robustness by generating images from a larger, more diverse dataset, ensuring its relevance extends to practical applications in real-world situations. The paper [9] utilized a VAE-GAN model and achieved an FID score of 0.1334 on a dataset of 1,341 medical images. While this result is noteworthy, it is limited by the small dataset size.

3. Proposed Work

The proposed VAE-GAN model demonstrates its efficacy by generating diverse and realistic medical images. Our model incorporates an innovative feature known as dynamic latent space adjustment. This mechanism allows for precise control over latent representations, enabling the emphasis or de-emphasis of specific anatomical characteristics such as bone density, lung opacity, or tissue textures. This capability facilitates the creation of tailored synthetic images that simulate rare medical conditions or address unique diagnostic scenarios, marking a transformative step in advanced medical training. Unlike standalone VAEs, which struggle with generating highly detailed images, or GANs, which lack structured latent spaces, our hybrid model leverages the strengths of both. The discriminator's learning rate is adjusted dynamically in response to the decoder's performance, which prevents issues like mode collapse and promotes a steady enhancement in image quality during training. One of the standout aspects of our hybrid framework is its adaptability across different imaging domains. With minimal architectural modifications, the model can be repurposed to generate synthetic medical images such as CT scans or MRI visuals. This cross-domain compatibility is achieved by leveraging pre-trained weights, making the system highly versatile and suitable for multi-modal medical imaging applications. The synthetic images generated by our approach have immense practical value in real-world healthcare. They can significantly enhance machine learning model performance by augmenting datasets for rare medical conditions, thereby improving generalization in clinical diagnostic tasks. Furthermore, the organized structure of the latent space enables controlled modifications in specific anatomical features, offering a foundation for personalized healthcare solutions. In contrast to prior works like those referenced in [9], which primarily focus on adversarial training, our hybrid VAE-GAN model integrates a well-structured latent space to enable more interpretable and controllable image generation. This allows our framework to synthesize visuals that are not only realistic but also closely correspond to specific medical conditions, making it especially suitable for applications like anomaly detection and diagnostic validation.

3.1 Dataset Used

The dataset utilized in this study is sourced from Kaggle [10] and focuses specifically on X-ray im- ages. This dataset is comprehensive, encompassing approximately 10,000 X-ray images, each annotated with relevant labels. These labels are crucial as they provide necessary insights into the features or conditions depicted in the images, which equips the model to effectively learn from the data. The diversity within the dataset ensures a broad representation of different medical conditions and variations in X- ray imaging, which is instrumental for training a robust model capable of generating synthetic X-ray images. By leveraging such a diverse dataset, the model can better capture the nuanced patterns and anomalies present in real-world X-ray data, enhancing its ability to produce realistic and clinically relevant synthetic images, while the dataset is diverse, it does not fully cover rare diseases, which is why real-world validation on hospital-acquired datasets is necessary. In Figure. 2, a portion from the dataset is illustrated, showcasing the type of X-ray images used for training the proposed model. This visual representation helps to contextualize the dataset's content along with the nature of the images that contribute to the model's learning process.

Before training, the X-ray images underwent preprocessing steps to ensure uniformity and enhance model performance. These steps included resizing all images to a standard dimension, normalizing pixel intensities for consistent input distribution, and applying histogram equalization to improve contrast. Augmentation techniques such as flipping and rotation were also employed to increase dataset variability, further aiding the model in generalizing to diverse data distributions.

While the model performs well on this dataset, external validation on clinical datasets from hospitals or medical institutions is necessary to ensure real-world applicability.

The use of a smaller dataset in this study was primarily driven by computational limitations associated with training deep learning models on large-scale medical imaging datasets. Training high-resolution medical image synthesis models, such as VAE-GAN, requires substantial GPU memory, processing power, and extended training times. Given these constraints, a smaller dataset was selected to ensure feasibility while maintaining meaningful experimentation.

While a larger dataset would provide greater variability and better generalization, computational constraints necessitated the use of a smaller dataset. However, through augmentation, and regularization Techniques used dropout and weight decay to prevent overfitting, the model was able to generate diverse and clinically meaningful synthetic images despite dataset limitations.

Since the dataset is relatively small, the model may learn patterns specific to the training set rather than generalizing well to unseen medical images.

Overall, the richness and scale of the dataset are integral to the development of an effective model for generating synthetic medical images, reinforcing the validity and applicability of the proposed approach.

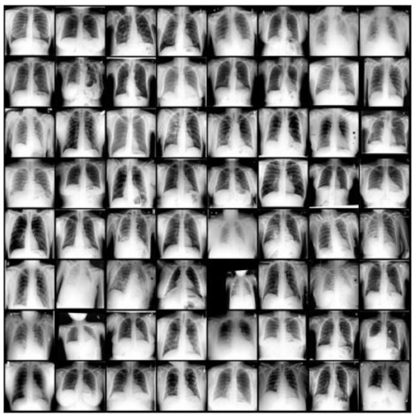


Fig. 2. Chest X-Ray Image Dataset.

3.2 Architecture of the model used (VAE+GAN)

In the proposed architecture, the discriminator consists of six convolutional layers with batch normalization and Leaky ReLU activation. This design enables effective feature extraction while stabilizing adversarial training, ensuring the synthetic images closely resemble real medical images. The VAE’s encoder employs a 9-layer of convolutional and fully connected layers to effectively capture hierarchical medical image features. The deeper structure allows the model to extract detailed anatomical patterns while preserving latent space continuity. The decoder employs a seven-layer architecture composed of fully connected and transpose convolutional layers. This structure is optimized to reconstruct high-fidelity medical images from the latent space while ensuring smooth spatial transitions and preserving fine anatomical details. These structures have been meticulously crafted to produce realistic medical images and ensure an effective feature representation within a well- organized latent space. Batch normalization is employed post every convolutional layer to stabilize the training process, it also helps standardize activations, preventing exploding or vanishing gradients, which are common in deep networks. This smooths training and accelerates convergence. Leaky ReLU activations are utilized, providing a small gradient when the input is negative, which helps prevent dead neurons and improves the model’s capability to learn intricate data distributions, this also ensures continuous learning throughout training, reducing vanishing gradient issues. The architecture of our VAE-GAN hybrid model is explicitly defined with detailed layer counts for each component, which distinguishes it from the architecture presented in the latest research [9]. Careful consideration is given to feature representation, ensuring that the latent space captures meaningful and interpretable features. The combination of convolutional, fully connected, and transpose convolutional layers within each of the encoder and decoder lets the model to learn intricate patterns and structures present in medical images.

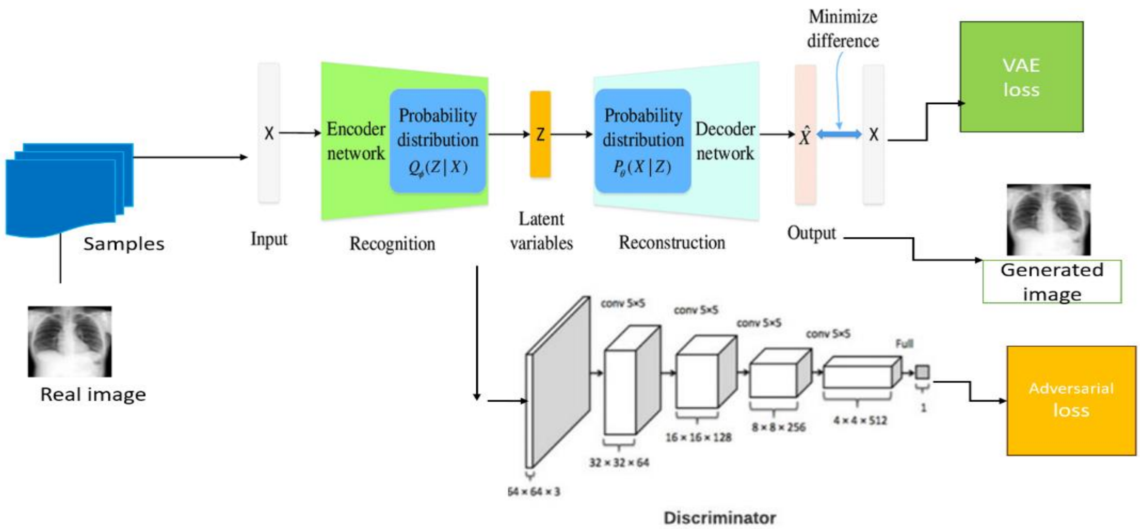


Fig. 3. GAN+VAE Architecture

3.2.1 VAE

- Encoder:

The encoder's function is to transform source data (such as images) into a condensed, lower dimensional latent space. It extracts the critical features of the input, representing them as latent vectors. In Variational Autoencoders (VAEs), the encoder learns to produce a distribution (mean and variance) rather than a single point in the latent space. This distribution is used to sample latent vectors, ensuring a continuous and structured latent space.

$$q(z|x) = N\ (z;\ \mu(x), \sigma^2(x)I) \tag{1}$$

This equation (1) represents the probabilistic encoder in the VAE. It encodes the input (x) into a probability distribution over the latent space (z) , where $\mu(x)$ represents mean and $\sigma 2(x)$ is variance.

- Decoder:

The decoder's role is to take latent vectors (either sampled from the latent space or provided by the encoder) and reconstruct them back into the original data form, in our model, VAE-GANs, the decoder acts as a generator and works in tandem with the discriminator. The decoder aims to produce outputs that can deceive the discriminator, thereby enhancing the overall performance of the generative model.

$$p(x|z) = N\ (x;\ \mu(z), \sigma^2(z)I) \tag{2}$$

The decoder generates the reconstructed output x given a sample from the latent space (z) . Similar to the encoder, $\mu(z)$ is the mean and $\sigma 2(z)$is the variance.

3.2.2 Generative Adversarial Networks (GANs)

- Discriminator (D):

The primary role of the discriminator is to tell apart inputs that are real (from the training set) and those that are fake (generated by the decoder). It provides a probability score between 0 and 1, reflecting the likelihood that the input is authentic. Throughout the training process, the discriminator evaluates the decoder's output, thereby directing the decoder's efforts. The decoder aims to produce samples indistinguishable from real data, while the discriminator works to enhance its accuracy in detecting synthetic data.

$$D(x) \rightarrow [0,1] \tag{3}$$

The discriminator assesses an input *x* from the decoder and returns a probability indicating whether the input is real $(D(x) \approx 1)$ or indicating that it is fake $(D(x) \approx 0)$.

$$L_{VAE} = L_{rec} + L_{KL} \tag{4}$$

$$L_{rec} = -E[logp(x|z)] \tag{5}$$

This term evaluates the extent to which the generated output (x) matches the input data. It represents the negative log-likelihood of the data conditioned on the latent variable.

$$L_{KL} = D_{KL}(q(z|x)||p(z)) \tag{6}$$

This term ensures that the learned latent space distribution $(q(z|x))$ is akin to a prior distribution $(p(z))$. It regularizes the latent space.

$$L_{GAN} = E[logD(x)] + E[log(1 - D(G(z)))] \tag{7}$$

This GAN loss consists of the discriminator's log probability of real data $(\log D(x))$ and the log probability of generated data being classified as fake $(\log(1 - D(G(z))))$.

3.3 Training Process

To synthesizing diverse and realistic medical images, the training process of the hybrid VAE+GAN model is designed with careful coordination between the components that is, Discriminator, and the Variational Autoencoder (VAE). Adaptive Moment Estimation (Adam) is utilized as the optimization algorithm. Adam adjusts learning rates for individual network parameters during training, providing several advantages:

- Faster convergence compared to traditional stochastic gradient descent (SGD).
- Efficient management of sparse gradients, tailoring it to be optimally fitting for large and intricate datasets.
- Less sensitivity to the initial parameter settings.

The VAE+GAN model involves the following training components:

- The discriminator evaluates generated images to classify them as real or fake.
- The decoder produces images from latent space representations, aiming to create realistic outputs.
- The encoder real images into latent representations.

The model was trained on a laptop equipped with an NVIDIA GeForce RTX (4GB VRAM) and 16GB RAM. Each epoch completed in less than 1 minute, demonstrating computational efficiency for training on mid-range hardware. However, training on significantly larger datasets may require multi-GPU scaling to maintain efficiency. While the training time was efficient on a single-GPU setup, scaling to larger datasets (e.g., 100,000+ images) may require multi-GPU parallelization.

Iterative process training involves updating the following components in an alternating fashion.

Update VAE:

- Encode real images into latent representations using the VAE encoder.
- Decode the latent representations back into images using the VAE decoder.
- Compute the reconstruction loss, which consists of Mean Squared Error (MSE) Loss which measures pixel-level differences between the original and reconstructed images, ensuring fidelity and clarity. And Kullback-Leibler (KL) Divergence Loss used to regularize the latent space to maintain a well-structured and continuous representation.
- Update the VAE encoder and decoder using the Adam optimizer to minimize this loss and enhance reconstruction quality.

Update Discriminator:

- Feed a batch of real images from the dataset to the discriminator.
- Feed a batch of generated images (from the decoder) to the discriminator.
- Compute the binary cross-entropy loss, which quantifies how well the discriminator distinguishes between real and synthetic images.
- Update the discriminator's parameters using the Adam optimizer to minimize this loss.

Update Decoder (Adversarial Learning):

- The decoder generates synthetic images and attempts to fool the discriminator into classifying them as real.
- Compute the adversarial loss, which helps improve the realism of the generated images.
- Update the decoder's parameters using the Adam optimizer to enhance the adversarial performance.

Iterate through these steps for multiple epochs until the model converges to a satisfactory performance level.

Through this iterative process, the VAE+GAN model learns to generate realistic and diverse images while preserving structural and semantic information from the training data.

Combining Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) can create powerful hybrid models for medical image synthesis. This approach leverages the strengths of both architectures.

- VAE Captures the underlying data distribution and generates diverse, realistic variations within that distribution.
- GAN Enforces high-fidelity image reconstruction by pitting a decoder against a discriminator.

The training begins with a real medical image as input, which is processed by the encoder network of the VAE. This encoder extracts the image's latent representation, compressing it into a code within the latent space.

The latent code becomes the key for manipulating the generated image. It is possible to adjust the code to control specific aspects of the synthesized image, like brightness, contrast, or even subtle anatomical variations. This manipulation allows for diverse image generation while staying within the realistic constraints learned from the dataset used for training.

The manipulated latent code is then fed into the VAE's decoder network. The decoder reconstructs a medical image based on the modified latent representation. However, in a VAE+GAN hybrid model, this reconstructed image isn't the final output. Instead, it serves as input to the GAN's discriminator network.

The VAE's decoder competes with GAN's discriminator. The decoder aims to produce images that fool the discriminator into believing they're real X-rays. In contrast, the discriminator works to differentiate between real and generated images.

This adversarial training loop pushes the decoder to produce increasingly realistic and high-fidelity X-ray images. The final output of the VAE+GAN hybrid model is the refined X- ray image generated by the GAN. Minimizing the VAE loss to ensure accurate reconstruction and a smooth latent space.

Minimizing the GAN loss to that the discriminator is refined to create images that appear increasingly realistic and which closely resemble real medical images. The discriminator's feedback pushes the decoder to create more realistic images, while the VAE's regularization prevents it from overfitting and collapsing to a single mode.

Our training process alternates update between the Discriminator, and VAE, ensuring that each component is fine-tuned to produce realistic images while maintaining a structured latent space. This structured training loop is critical for balancing the adversarial loss of the GAN and the reconstruction loss of the VAE. By iteratively updating these components, we avoid issues such as mode collapse and ensure that the generated images are diverse and realistic.

In our model's emphasis on the latent space structure ensures that the generated images retain meaningful medical features, which is crucial for applications like anomaly detection. The previous studies, such as [9], while focusing on realism, lacks this structured approach to latent space and therefore may not offer the same level of control over image diversity and feature representation.

In the existing system [9] which also combines VAE and GAN objectives, but the training process lacks the detailed description of alternating updates and loss balancing. In our work, the balance between the VAE's reconstruction loss and the GAN's adversarial feedback is carefully managed to optimize both image realism and latent space representation.

3.4 Quantitative and Qualitative Metrics Evaluation

The evaluation criteria employed for evaluating the image quality are Fréchet Inception Distance (FID), it measures realism and feature similarity, it compares feature distributions between real and generated images using a deep neural network. FID can prove that the synthetic images maintain real-world anatomical accuracy, crucial for medical applications. This means synthetic data can be safely used to expand training datasets for AI-driven medical diagnosis. Inception Score (IS), it gauges variety and fidelity of the generated images it also indicates clear, well-defined images that are also varied rather than repetitive and helps assess whether the model generates a broad range of images, essential for dataset augmentation. Mean Squared Error (MSE) assesses the average squared difference to tell apart from the real and generated images, it ensures that the synthetic images faithfully replicate real medical images. Peak Signal-to-Noise Ratio (PSNR) evaluates the proportion of signal relative to noise within an image. Since MSE and PSNR alone fail to capture high-level structural and perceptual similarities essential for clinical analysis FID and IS are also included to ensure both perceptual quality and diagnostic relevance. The principal aim was to qualitatively assess the clinical relevance and accuracy of the generated images, going beyond numerical metrics.

This assessment aimed to ensure that the generated images exhibit features consistent with real medical conditions. In addition to this the expert review served as a complement to quantitative metrics, adding a qualitative layer through the nuanced judgment of medical professionals. The aim was to validate the clinical authenticity of the generated images. The research paper recognized the integral role of visual inspection and expert review in contributing to a comprehensive evaluation strategy. By addressing the subjective nature, clinical utility, and sensitivity to anomalies in visual inspection, and emphasizing domain expertise, complementary with quantitative metrics, and consensus building in expert review, the paper ensured a holistic understanding of the generated medical images.

4. Experiments and Results Discussion

The following segment offers a comprehensive examination of the experiments conducted to deter- mine the efficiency of the hybrid VAE+GAN model in producing medical images. The experiments primarily focus on judging the standard quality, variety, and realism of the synthesized images through thorough quantitative and qualitative assessments. Using a comprehensive dataset consisting of 10,000 chest X-ray images from Kaggle, the model was trained and tested under various configurations and settings to ensure robust performance. The results presented here highlight the efficacy of the hybrid approach, showcasing significant improvements over traditional methods. These experiments not only validate the theoretical advantages of combining VAEs and GANs while also offering practical insights into their application in the medical imaging domain.

- Hyperparameters: Learning Rate (lr): Defines how quickly the model updates its parameters during training. A moderately low value of 0.0002 is commonly chosen for this.
- Epochs: Refers to how many times the model processes the entire dataset during training. In this experiment, training was conducted for 1000 epochs.
- Batch Size: Is size of data samples processed during a single training iteration. Although increasing the batch size can speed up the training process, it requires a greater amount of memory resources. A typical choice is 64.
- Latent Size: Defines the dimensionality of the latent space. A larger latent size allows for greater capacity to encode intricate and detailed features.
- Optimizer: Adam optimizer with specified betas is commonly used in GANs and VAEs.
- Advantages of Hybrid GAN is that the VAE component enhances diversity in generated images, ad- dressing the common GAN issue of producing limited variations. VAE ensures that the latent space has a structured representation, aiding interpretability and control over generated features. GANs excel in generating realistic images, providing a hybrid model with a balance between realism and diversity.
- Latent Space and VAE's Contribution: The VAEs map input data to a structured latent space, providing a more organized representation of features. The structured latent space enables a deeper understanding of the variations and features encoded in the data. VAE's contribution to the latent space helps generate high-resolution medical images with meaningful features.

4.1 Results Discussion

The successful generation of medical images using the hybrid GAN+VAE architecture represents a significant stride towards addressing challenges in medical data synthesis. The results derived from the model underscore its efficacy in producing realistic and diverse medical images, showcasing its potential applications in various medical domains.

The model was trained using 10,000 images from the Kaggle dataset. During the training process, the model generates images at each epoch. The overall count of images the are generated can be controlled by specifying the total epochs and the chosen batch size. For this research, the model was set to generate 64 synthetic images at each epoch in 8x8 grid like in Fig. 2.

The process in which the images are generated from random noise to the synthetic image is depicted in Fig. 4.

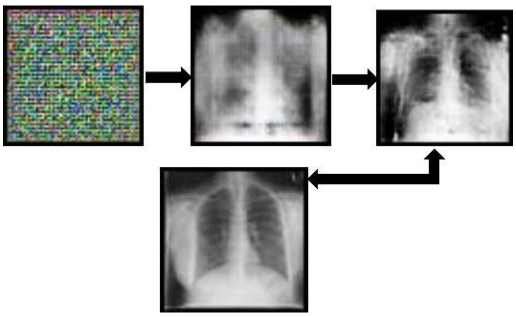
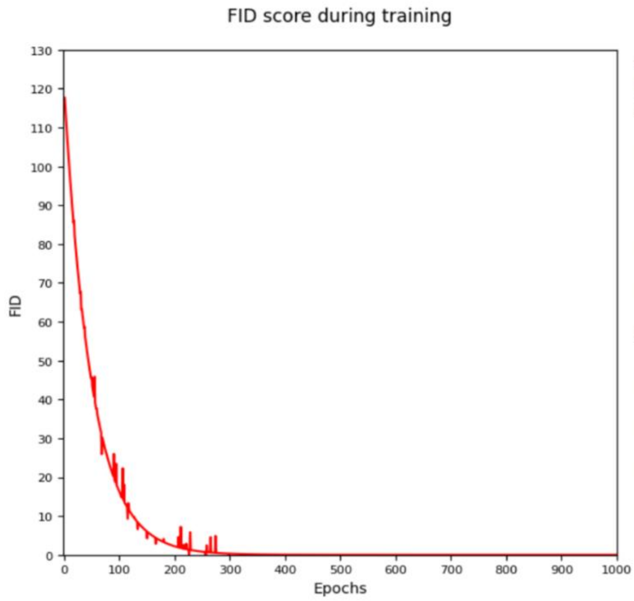


Fig. 4. VAE+GAN Generated Image.

The synthesized images were evaluated using several metrics, including Fréchet Inception Distance (FID), Inception Score (IS), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR), which confirmed their diversity and realism.

These metrics verify the efficiency of the hybrid model in producing diverse and realistic medical images well-suited for applications like data augmentation, anomaly detection, and machine learning model training.




```
Mean (Real): [0.00397116 0.04147372 0.16294074 0.00079249 0.65659344]
Mean (Generated): [0.00397116 0.04147372 0.16294074 0.00079249 0.65659344]
Covariance (Real): [[ 5.09205292e-05  8.84905237e-05 -1.38304539e-04 -7.25148637e-08
   3.32085668e-05]
 [ 8.84905237e-05  1.56141215e-03  1.47400755e-04  8.30545792e-06
   1.19599179e-03]
 [-1.38304539e-04  1.47400755e-04  5.13642017e-03 -1.22863292e-05
   2.23255220e-03]
 [-7.25148637e-08  8.30545792e-06 -1.22863292e-05  2.85111188e-06
   2.77711344e-05]
 [ 3.32085668e-05  1.19599179e-03  2.23255220e-03  2.77711344e-05
   7.70598972e-03]]
Covariance (Generated): [[ 5.09205292e-05  8.84905237e-05 -1.38304539e-04 -7.25148637e-08
   3.32085668e-05]
 [ 8.84905237e-05  1.56141215e-03  1.47400755e-04  8.30545792e-06
   1.19599179e-03]
 [-1.38304539e-04  1.47400755e-04  5.13642017e-03 -1.22863292e-05
   2.23255220e-03]
 [-7.25148637e-08  8.30545792e-06 -1.22863292e-05  2.85111188e-06
   2.77711344e-05]
 [ 3.32085668e-05  1.19599179e-03  2.23255220e-03  2.77711344e-05
   7.70598972e-03]]
```

Fig. 5. FID score during training.

4.1.1 Inception Score (IS) 1.29 +/- 0.1

The score suggests that the generated images exhibit high diversity, covering a wide range of possible variations. However, the standard deviation of 0.1 indicates that the generated images might appear remarkably similar with each other, potentially lacking some variability.

4.1.2 Mean Squared Error (MSE) 0.558891236782074

The value suggests a moderate level of the disparity between the generated images and their real counterparts. A lower MSE generally indicates better reconstruction quality.

4.1.3 Peak Signal-to-Noise Ratio (PSNR) 50.66 dB

This value is considered relatively good, indicating a high signal-to-noise ratio and suggesting the generated images bear a strong resemblance to the real images in terms of pixel-level accuracy.

4.1.4 Fréchet Inception Distance (FID) score ≈ 0

The FID score from Figure.5 shows exceptionally low FID score achieved by our model, and demonstrates its capability to generate synthetic medical images with unprecedented accuracy and real- ism, surpassing the results of previous studies [9]. This significant improvement in image quality, combined with our model's ability to handle a larger and more diverse dataset, positions our approach as a cutting-edge solution for medical image augmentation, anomaly detection, and privacy-preserving data generation. In Figure. 6, the key loss functions—discriminator loss, and VAE loss are visualized.

Discriminator loss: The discriminator's loss decreases, suggesting that it is improving at identifying real versus generated images.

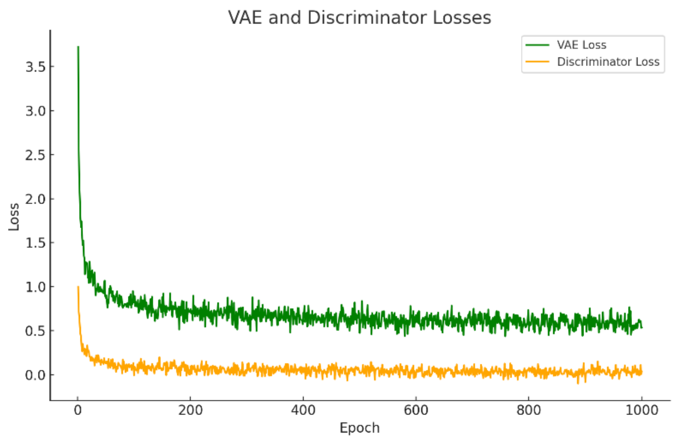


Fig. 6. Loss Function.

VAE loss: The VAE loss gradually decreases over time, implying that the VAE maintains a good balance between the reconstruction loss and the KL divergence.

In this model, the VAE's loss function incorporates both reconstruction loss and KL divergence to ensure the latent space to be smooth and structured, thereby enabling effective feature extraction from medical images. Meanwhile, the GAN's loss function is driven by the discriminator's performance in differentiating between real and synthetic images, providing adversarial feedback to refine the decoder's outputs.

In the previous studies [9] mentions a combined loss function for VAE and GAN but does not specify how the adversarial and reconstruction losses are balanced during training. Our approach offers a more explicit handling of loss functions, ensuring that the model can generate excellent standard images with both authenticity and interpretability, while minimizing the problem of overfitting or mode collapse.

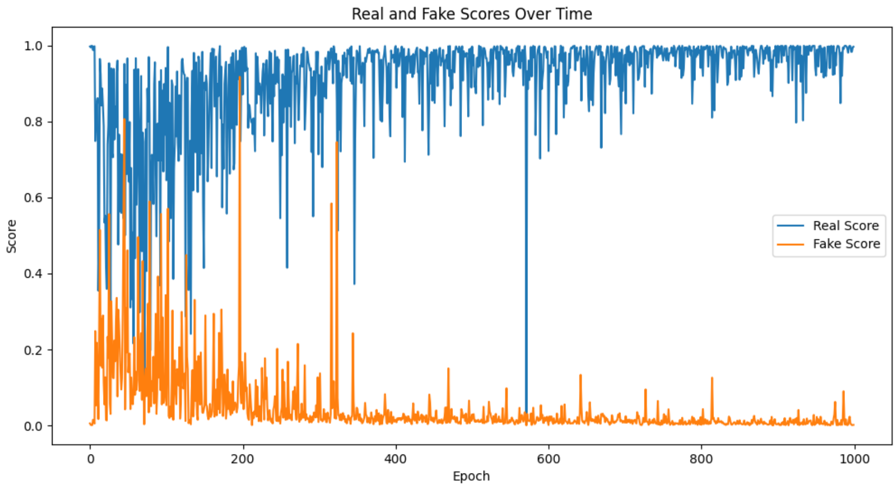


Fig. 7. Real and Fake scores

In Figure.7, the two lines on the graph represent the real and fake scores, which are metrics employed to assess the standard of the generated images, with the real score providing a measurement of authenticity of how similar the generated images are to the actual medical images in the training data. The fake score evaluates how well the discriminator to accurately tell apart from real and artificially generated images. In a typical GAN training process, the real score tends to rise steadily as the decoder gets more effective at producing realistic images. The fake score tends to drop progressively as the discriminator gets better at spotting fake images.

In a typical GAN training process, the real score should increase through time as the decoder gets more effective in producing realistic images. The fake score should decrease over time as the discriminator gets better at spotting fake images.

The specific details of the graph will vary based on the dataset used, the GAN architecture, and the training parameters. However, the general trend should be for the real score to increase and the fake score to decrease. The real and fake scores are not always directly comparable. The scale of the axes may be different, and the scores may be calculated in different ways. The training process may experience instability, and the real and fake scores may fluctuate over time. Even if the real score is high, it does not mean that the generated images are perfect. There may still be subtle differences between the real and generated images.

- High-Quality Image Generation: The model generated synthetic medical images. These images exhibited high fidelity and realism, as evidenced by a high PSNR value of 50.66 dB
- Diversity in Images: The Inception Score (IS) of 1.29 +/- 0.1 suggests that the generated images encompass a broad spectrum of potential variations, ensuring that the model does not produce repetitive or redundant images.
- Accurate Image Reconstruction: The Mean Squared Error (MSE) of 0.558891236782074 suggests a noticeable yet moderate distinction between the real and generated images, validating the model's capacity to precisely reconstruct original images from input data from the latent space.
- Effective Training Process: The loss functions for the GAN's discriminator(D), and VAE (as shown in Figure. 6) steadily converged during training, signifying that the model is learning in a stable manner.
- Realism in Generated Images: The Fréchet Inception Distance (FID) score approaches zero, demonstrating that the generated images are highly similar to real images. This score is a strong metric of the model's capacity to produce standard and realistic images.

Overall, the results support the notion that the hybrid GAN+VAE model is proficient in generating medical images that bear close likeness to the real images, successfully capturing the essential features needed for realistic synthesis.

This could have a number of potential applications, such as data augmentation for training medical image analysis algorithms or generating synthetic images for anonymization purposes. But we must also keep in mind that some generated images may exhibit blurry textures or anatomically inconsistent structures, reducing their clinical usefulness.

Table 1. Comparison Table

Models	PSNR value	Data Modality
GAN [11]	34.42 dB, 33.01 dB	Different Medical images
SRGAN [12]	26.4505 dB	Different Medical images
SRCNN-GAN [13]	26.18 dB	Chest X-ray images
FA-GAN [14]	42.11 dB	Brain MRI images
VAE+GAN [9]	21.11 dB	Chest X-ray images
Proposed VAE+GAN	50.66 dB	Different Medical images

4.2 Limitations and Future Work

The Kaggle dataset used in this study covers a wide range of medical conditions, making it useful for training robust models. However, it may still lack sufficient cases of rare diseases, which could limit the model's ability to generalize to less common conditions. Testing on specialized hospital datasets will help ensure that the model performs well across both common and rare cases. Additionally, the model should be evaluated in real clinical workflows, such as radiology departments, where AI-generated images assist in decision-making.

AI-generated medical images must always be reviewed by medical professionals before being used for training or diagnosis. There is a risk of over-reliance on synthetic data, which could reduce clinician involvement in decision-making. Future research should focus on collaborating with hospitals, conducting clinical trials, and integrating the model into AI-powered diagnostic tools.

While our model demonstrates efficient training times (less than 1 minute per epoch) on the current dataset, training on significantly larger real-world datasets presents scalability challenges. High-resolution medical image synthesis using GANs and VAEs requires substantial GPU memory and computational resources. Future research should explore multi-GPU training strategies and distributed computing frameworks (e.g., PyTorch, TensorFlow) to improve performance when handling large datasets.

Our proposed model incorporates batch normalization, which stabilizes training but increases memory consumption. Future iterations could explore distributed training techniques, such as multi-GPU training in PyTorch or TensorFlow, to improve scalability when handling larger datasets.

Our model uses an adaptive feedback loop for dynamic discriminator learning rate adjustment, optimizing the training process compared to conventional GAN architectures. Given the 4GB VRAM limitation, the model effectively balances memory usage with training speed. However, larger batch sizes could be tested on higher VRAM GPUs (e.g., RTX 3090 or A100) to further optimize performance.

While our model demonstrates strong quantitative results (FID $\approx$ 0, PSNR 50.66 dB), real-world validation remains essential. Future work will involve collaborating with hospitals and research institutions to assess the clinical utility of generated images through expert evaluation and real patient datasets. Additionally, our model will be tested within AI-powered diagnostic systems to determine its effectiveness in medical workflows.

Furthermore, deploying AI-generated medical images in real-world applications requires compliance with regulatory standards such as HIPAA (for data privacy) and FDA/CE approvals (for medical imaging AI). Future studies will address these considerations to ensure ethical and legal compliance in clinical environments.

In conclusion, upcoming improvements will focus on broadening dataset diversity, enhancing scalability, validating real-world clinical performance, and ensuring regulatory compliance. These enhancements will be critical in closing the gap between AI-driven image synthesis and its practical deployment in medical environments.

5. Conclusion

In summary, this study denotes a pivotal stride in the domain of medical image synthesis by harnessing the synergistic power of Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). The integrated model, comprising a Discriminator, and VAE, showcases a novel approach to generating synthetic medical images with notable advancements in both quality and diversity. The meticulous training process, orchestrating the collaboration of the discriminator, and VAE, demonstrates the model's ability to produce high- fidelity medical images while maintaining a structured latent space. Leveraging a carefully curated dataset from Kaggle, the model excels in capturing and reproducing intricate medical features, showcasing its versatility across various healthcare applications. The results of this research highlight the successful synthesis of medical images using a hybrid VAE+GAN architecture.

By utilizing 10,000 medical image dataset sourced from Kaggle which consists of chest X-ray images, the model achieved the task of generate high-quality synthetic images, addressing key challenges in medical imaging such as data scarcity and patient privacy. However, while the dataset provided valuable diversity, testing on hospital-acquired datasets remains essential to ensure generalization across rare medical conditions. The result and proposed work underscores the architectural clarity, training process sophistication, balanced loss functions, and broad application scope of our paper. In particular, the evaluation metrics employed and the emphasis on real-world medical applications distinctly enhance the practical value of our paper compared to the more theoretical and less detailed approach presented in the previous papers [9]. These attributes collectively signify substantial advancements in methodology and applicability.

One of the paper’s key strengths lies in the hybridization of VAE and GAN, where the decoder’s proficiency in generating realistic data is complemented by the encoder’s ability to provide a structured la- tent space. This unique amalgamation results in a model that not only excels in realism but also facilitates interpretability, crucial for applications for instance anomaly detection and machine learning model training.

The results of this research demonstrate the effectiveness of the proposed model, showcasing diverse and realistic synthetic medical images.

The structured latent space obtained through the VAE proves instrumental for boosting the interpretability of the generated images, setting the stage for widespread applications in data augmentation, diagnostic tool development, and medical image analysis. However, it’s vital to acknowledge the paper’s limitations, such as the need for a robust evaluation framework and potential challenges associated with data heterogeneity. Future research endeavors may focus on refining these aspects, pushing the boundaries of medical image synthesis and addressing emerging challenges in the ever-evolving landscape of healthcare technology.

Despite these strengths, the study acknowledges certain limitations, including the need for a more robust evaluation framework and the potential impact of data heterogeneity on model performance. While the model achieves efficient training times (less than 1 minute per epoch) on the current dataset, scalability remains a challenge when working with larger, high-resolution medical datasets. Future research should explore multi-GPU scaling and distributed training strategies to improve computational efficiency**.**

Overall, this paper contributes a promising methodology to address data scarcity and privacy concerns in medical image generation, opening avenues for further exploration and innovation in the intersection of deep learning and healthcare.

References

[1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017 May 24;60(6):84-90.
[2] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. InProceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 675-678).
[3] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Advances in neural information processing systems. 2014;27.
[4] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. 2015 Nov 19.
[5] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. 2013 Dec 20.
[6] Kingma DP, Mohamed S, Jimenez Rezende D, Welling M. Semi-supervised learning with deep generative models. Advances in neural information processing systems. 2014;27.
[7] Larsen AB, Sønderby SK, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. InInternational conference on machine learning 2016 Jun 11 (pp. 1558-1566). PMLR.
[8] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks. arXiv preprint arXiv:1701.04862. 2017 Jan 17.
[9] Rguibi Z, Hajami A, Zitouni D, Maleh Y, Elqaraoui A. Medical variational autoencoder and generative adversarial network for medical imaging. Indonesian Journal of Electrical Engineering and Computer Science. 2023 Oct;32(1):494-505.
[10] https://www.kaggle.com/datasets/sinchanaganesh/crx-dataset
[11] Ren Z, Stella XY, Whitney D. Controllable medical image generation via GAN. J. Percept. Imaging. 2022 Mar 1;5:000502-1.
[12] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. Photo-realistic single image super-resolution using a generative adversarial network. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4681-4690).
[13] Dittimi TV, Suen CY. Single image super-resolution for medical image applications. InInternational Conference on Pattern Recog- nition and Artificial Intelligence 2020 Oct 9 (pp. 660-666). Cham: Springer International Publishing.
[14] Jiang M, Zhi M, Wei L, Yang X, Zhang J, Li Y, Wang P, Huang J, Yang G. FA-GAN: Fused attentive generative adversarial networks for MRI image super-resolution. Computerized Medical Imaging and Graphics. 2021 Sep 1;92:101969.

Authors' Profiles

Sinchana Ganesh is a Master's student in Computer Applications at JSS Science and Technology University, Mysuru. Her research interests focus on Artificial Intelligence and Machine Learning. She is actively engaged in exploring AI-driven solutions for various computational challenges.

Madhushree B. received her MCA degree from JSS Science and Technology University, Mysuru in 2020 and pursuing her Ph.D in JSS Science and Technology University, Mysuru. Currently she is working as Assistant Professor in the department of computer applications, JSS Science and Technology University, Mysuru. Her research interests are image forensics and machine learning.

Sowmya K. N. earned her Ph.D. in the Faculty of Computer and Information Sciences from Visvesvaraya Technological University in 2021, specializing in video forensics. She is currently an Associate Professor at JSS Academy of Technical Education, Bengaluru, and a certified Cyber Crime Intervention Officer by ISAC, India. Her research focuses on video and image forensics, particularly in areas such as video forgery detection, weapon detection in surveillance, and facemask authentication. Additionally, she explores applications in computer vision and machine learning. Dr. Sowmya has made significant contributions to the field, including filing two Indian patents and publishing multiple research papers in peer-reviewed journals.

H. R. Chennamma received Ph.D. in Computer Science from the University of Mysore in the area of Digital Image Forensics in 2011. Subsequently, she served as a Post Doctoral fellow in the Department of Computer Science and Engineering, University of North Texas, USA in 2012. Currently, she is a Professor in the Department of Computer Applications, JSS Science and Technology University, Mysuru. Previously, she has been awarded a prestigious Fellowship under the scheme of Research and Development in Forensics from the Department of Home Affairs, Government of India in 2005. Chennamma served as a Senior Research Fellow (SRF) in Central Forensic Science Laboratory, Hyderabad. She is the recipient of two "Best Scientific Paper Awards". Her research areas of interest are Image and Video forensics, pattern recognition and Computer Vision.

How to cite this paper: Sinchana Ganesh, Madhushree B., Sowmya K. N., H. R. Chennamma, "Medical Image Synthesis Using Variational Autoencoder and Generative Adversarial Networks", International Journal of Engineering and Manufacturing (IJEM), Vol.15, No.3, pp. 56-68, 2025. DOI:10.5815/ijem.2025.03.05