

A Hybrid Attention-Driven Generative Adversarial Network for High-Fidelity Medical Image Synthesis

Suresh Sappa^{1*}, and Dr.R. Tamilkodi²

^{1*}Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Kakinada, India. 22022p0584@jntuk.edu.in, <https://orcid.org/0009-0007-7236-8516>

²Professor & Head, Department of Computer Science and Engineering (AIML), Godavari Institute of Engineering and Technology (Autonomous), Godavari Global University, Rajahmundry, India. hod.aiml@giet.ac.in, <https://orcid.org/0000-0001-6295-4633>

Received: February 16, 2026; Revised: March 23, 2026; Accepted: May 11, 2026; Published: June 30, 2026

Abstract

Medical image generation plays an important role in modern healthcare systems, addressing data deficiencies to improve the effectiveness of diagnostic models. GANs have shown impressive potential for producing high-quality synthetic images; however, they often fail to maintain structural consistency and preserve fine image details. To overcome these challenges, this study proposes a novel hybrid attention-driven generative adversarial network that incorporates spatial and channel attention, as well as multi-scale feature fusion. Attention mechanisms have shown promising potential in producing high-quality images that are more interpretable and meaningful in a clinical decision-making process. The proposed model is implemented and tested using various datasets and evaluation metrics, including PSNR, SSIM, and FID. Experimental results show that HA-GAN achieves superior performance compared to existing models in producing high-quality images for decision-making. Image generation in medical diagnostics deals with the problem of data insufficiency in diagnostic techniques. GANs encounter issues with the structural accuracy of their generated images. The authors of this study have used attention-driven Hybrid GAN (HybridAGAN) with multi-scale attention features to resolve the above-mentioned problems. Specifically, the model's generator combines an attention mechanism with a multi-scale representation. At the same time, the discriminators focus on both the global structure and the local texture of generated images. The effectiveness of the proposed approach is validated using two benchmarks: the Brain MRI and Chest X-ray datasets. To measure the model's performance, several quantitative criteria were used, including PSNR, SSIM, and FID. The results showed that the proposed model performed much better than the baselines, achieving 34.7 dB PSNR, 0.91 SSIM, and 21.4 FID. Statistical testing, including paired t-tests ($p < 0.05$), demonstrated the significance of the obtained results.

Keywords: Generative Adversarial Networks (GANs), Medical Image Synthesis, Hybrid Attention, Deep Learning, Clinical Decision Support, Multi-Scale Learning.

1 Introduction

Medical imaging enables precise diagnosis, treatment planning, and disease monitoring; it is essential to contemporary healthcare (Zhan et al., 2021). However, due to high acquisition costs, patient privacy concerns, and the need for professional labelling, the availability of high-quality annotated medical images is often limited (Lan et al., 2021). However, such shortcomings limit the effectiveness of model-based approaches and underscore the importance of efficient methods for synthesising medical images (Gan et al., 2022; Gu & Ye, 2021). Since many existing approaches often suffer from shortcomings such as structural discrepancies, a lack of texture details, and perceptual realism, which may make those approaches unreliable for practical use, GANs have emerged as an innovative approach to synthesising medical images (Lian et al., 2022).

In light of these problems, the proposed solution uses the HA-GAN to generate accurate medical images. In this paper, an attention-driven GAN is proposed to improve feature representations by focusing on important regions in images. In addition, an innovative attention mechanism based on spatial and channel attention fusion is utilised to improve the feature extraction process. Multi-level feature fusion is adopted to make sure that the output images contain both local and global anatomical structures. Finally, a hybrid discriminator is used to ensure realistic textures in the generated images.

The Following is a List of the Most Significant Contributions of This Study

(i) A novel hybrid attention mechanism to improve image generation; (ii) A two-stage discriminator to enhance local and global structure realism; (iii) A multi-scale feature fusion technique to preserve structure; and (iv) a comprehensive evaluation through statistical validation and quantitative measures that prove superior performance compared to existing methods.

The rest of this document is structured as follows: In Section 2, relevant research on GAN-based methods and medical picture synthesis is reviewed. The architecture and mathematical formulation of the suggested HA-GAN approach are presented in Section 3. The explanation of results and discussion is provided in section 4, where information on the datasets used, data processing, and the evaluation of system performance via PSNR, SSIM, and FID measures is given. The experimental setup and datasets are described, and the findings and analysis, including statistical validation and comparative studies. Lastly, Section 5 is the conclusion that summarises the findings and the future study.

2 Literature Review

In recent days, the synthesis of medical pictures using deep learning methods has drawn huge attention. This is because these methods have the potential to improve decision support systems. This helps overcome limitations such as data scarcity and boosts diagnostic accuracy (Gan & Wang, 2022). Among the many deep learning methods, Goodfellow's original Generative Adversarial Networks (GANs) have demonstrated great promise for image synthesis. GANs can learn complex data distributions by employing adversarial training between the discriminator and generator networks (Guo et al., 2022).

A study proposed a deep generative adversarial segmentation network, GAN-segNet, for brain tumour semantic segmentation. In particular, this paper discusses how to utilise GANs along with semantic segmentation to achieve accurate and efficient brain tumour detection from brain MRI images. By leveraging the strengths of a new hybrid architecture, GAN-segNet is able to capture better global context features as well as detailed local information, which can be a challenge for semantic segmentation in medical image analysis (Cui et al., 2022).

Though their use in medical image synthesis is discovered to have certain limits, particularly in maintaining anatomical structures, early research employing GANs has demonstrated significant potential for natural picture synthesis. With the goal of natural picture generation, the DCGAN introduced the idea of utilising convolution layers to enhance feature learning and convergence (Liu et al., 2024). Though the quality of the generated images was improved by DCGANs, it still did not provide a satisfactory level for use in medical applications due to the inability of the DCGANs framework to preserve structural consistency. Later, the concept of conditional GANs (cGANs) was proposed for the use of auxiliary information, such as labels, for the generation of images (Chen et al., 2022). This was useful for medical applications since the data could be labelled for the generation of specific images with more relevance.

A research study proposed a robust GAN model for cross-modal generation of medical images, which can effectively transform MRI images into PET scans. This model aims to tackle the issue of how to bridge the difference between two imaging modalities through the use of a robust GAN model that ensures anatomical consistency in the generated images. This model employs an innovative methodology in order to cope with the difficulties associated with cross-modal image generation, thereby generating high-quality images while retaining key diagnostic information (Yang et al., 2025).

One major leap in the synthesis of medical images was made by introducing CycleGAN, which allows for unpaired image-to-image translation. This is especially important in medical imaging because it is often hard to obtain paired data in some modalities, such as MRI and CT scans of the same patient (Dar et al., 2019). CycleGAN has been used in various applications, such as MRI-CT translation and modality conversion. However, despite its success in various applications, CycleGAN sometimes produces artifacts and does not maintain detailed anatomical structures.

Recent studies have emphasised using attention mechanisms with GAN models to overcome these challenges. Attention mechanisms enable models to focus their attention on particular regions of the image, similar to how humans process visual information (Wali et al., 2023). Studies have demonstrated that using attention mechanisms with GAN models can greatly enhance the generation of clinically meaningful features, such as tumours and lesions. Specifically, spatial attention enables models to determine where these features exist, while channel attention enables to determine what is significant (Nie et al., 2018; Armanious et al., 2020). However, many such models have used these mechanisms separately, thus limiting their performance.

Further advancements have also focused on the utilisation of self-attention and multi-scale learning in GAN models. Self-attention allows for the modelling of dependencies in the image (Yi et al., 2017). This is important in ensuring that the generated medical image is consistent in its entirety. Additionally, multi-scale feature learning is important in ensuring that global features and local details are represented adequately (Yan et al., 2022). Multi-scale GAN models aggregate features from different layers in the networks. This improves edge preservation and texture consistency in the generated medical images. However, most GAN models have not adequately combined multi-scale learning with the utilisation of hybrid attention.

Another study introduced an architecture of a CycleGAN-based algorithm for the purpose of image denoising using low-dose computed tomography (CT) images. The study used the previous image information in order to solve the problem of noise that usually accompanies low-dose CT images. Another important area of research in this field is related to enhancing the discriminator's ability to evaluate generated images more effectively (Tang et al., 2019). A study conducted an extensive review of the application of Generative Adversarial Networks (GANs) for the analysis of 3D medical images. The paper discussed several types of GANs and their usage in enhancing 3D medical imaging, especially

in radiology and neuroimaging (Usama et al., 2025). Traditionally, discriminators in GAN models have been used to evaluate images at a global level, which might not effectively account for local-level features in images. To overcome this problem, Patch-GAN discriminators have been proposed to effectively evaluate local-level features in images, thereby improving their texture quality. However, it has also been argued that Patch-GAN discriminators might not effectively account for overall image structure, which might lead to inconsistencies in images generated through such models. To overcome this problem, it has been suggested that a local-level as well as global-level discrimination strategy might be more effective (Feng et al., 2023).

A study proposed a new deep attention branch network to classify skin lesions. The network makes use of deep learning and attention to emphasise the important characteristics in skin lesion images and thus improve the performance of the classifier (Ding et al., 2021).

Although a wide range of research has been conducted in this field, several challenges still remain unresolved. Some of the unresolved challenges in this field include model instability, mode collapse, a lack of effective preservation of anatomical features, and a lack of model interpretability. Furthermore, it has also not been effectively explored how a wide range of sophisticated mechanisms, including hybrid attention, multi-scale feature fusion, as well as dual discrimination strategies, might be effectively integrated into this field (Chen et al., 2019; Chen et al., 2019).

A paper described a two-stage GAN approach for generating retinal images of high resolution and segmenting the images. This work focused on producing high-quality retinal images that could be used for various diagnoses, especially in detecting diseases such as diabetic retinopathy. Through the use of a two-stage GAN approach, the authors managed to improve the resolution and quality of retinal images (Andreini et al., 2021). Another paper developed a novel model based on WGANs for CT images denoising using a combination of loss functions that minimise noise in the image while preserving its structural information. It focused on the problem of noise in low-dose CT imaging, which is known to affect the diagnosis process. This proposed method was effective in improving image quality without losing important structural information, thereby providing an innovative solution to this problem (Li et al., 2021).

According to the identified gaps in existing methods, this study proposes a hybrid attention-driven GAN model that incorporates both spatial and channel attention mechanisms and a hybrid discriminator with multi-scale feature fusion. The proposed model is designed to overcome some of the limitations associated with existing methods by enhancing both the aesthetic quality and structural integrity of generated images for better suitability in clinical applications.

3 Proposed Methodology

System Model and Assumptions

This suggests that HA-GAN provides high-quality medical images by learning the underlying data distribution from actual medical datasets. The system is made up of an adversarial trained hybrid discriminator (D) and a generator (G). The model makes the assumption that the input dataset X has an unknown distribution. $P_{data}(x)$, and the generator uses a latent variable $z \sim P_z(z)$ to try to approximate this distribution.

Additionally, it is believed that medical images display both local fine-grained characteristics and global structural patterns, which can be efficiently recorded by multi-scale feature learning and hybrid

attention processes. Depending on the availability of data, the model can function in supervised or semi-supervised environments.

Architecture of the Generator

The encoder-decoder architecture used by the generator is improved by multi-scale feature fusion and hybrid attention, as shown in figure 1. While the decoder uses upsampling techniques to rebuild the image, the encoder uses convolutional layers to extract hierarchical features. By highlighting significant areas and eliminating unimportant information, a hybrid attention module that combines spatial and channel attention is used to improve feature maps. Multi-scale feature fusion preserves both local and global details by combining features from many layers. To maintain spatial information and enhance gradient flow, skip connections are employed. The patch-based and global evaluation is combined using a hybrid model discriminator. The patch discriminator evaluates small-scale features, while the global discriminator looks at the consistency of the whole structure. Thus, the use of both enables complete coverage of the input data for the generator. For consistent training, the discriminator relies on convolutions with batch normalisation and leaky ReLU.

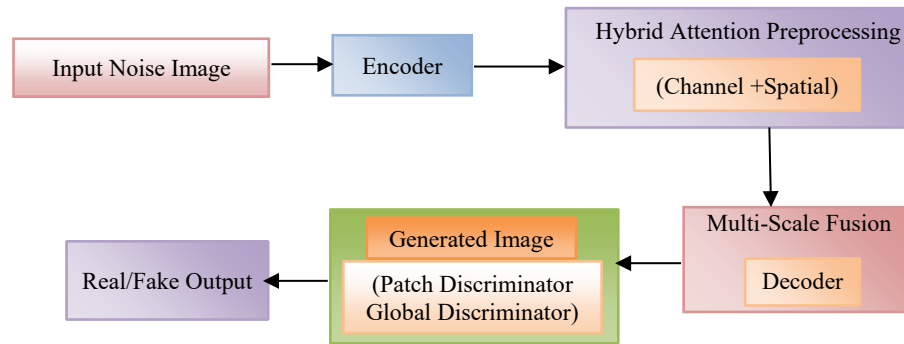


Figure 1: Proposed HA-GAN architecture

Mathematical Formulation

The objective of the GAN is defined in *equation 1*, where G is the generator and D is the discriminator, which is specifically used to differentiate between real and generated images. $P_{data}(x)$ refers to real data distribution. And $P_z(z)$ is the distribution of latent vector z .

$$L_{GAN}(G, D) = E_{x \sim P_{data}(x)} [\log \log D(x)] + E_{z \sim P_z(z)} [\log \log (1 - D(G(z)))] \quad (1)$$

To improve image quality, additional loss functions are incorporated:

Reconstruction Loss (Equation 2): The reconstruction loss forces the generator to generate an image that is similar to the actual image.

$$L_{rec} = \| x - G(z) \|_1 \quad (2)$$

Perceptual Loss (Equation 3): It is employed to make sure that the output image preserves high-level characteristics of the input image, normally through comparing characteristics from pretrained networks.

$$L_{perc} = \| \phi(x) - \phi(G(z)) \|_2 \quad (3)$$

Attention Loss: This makes the network pay attention to the important areas in the image. It utilises both spatial attention and channel attention, as shown in *equation 4*

$$L_{att} = \Sigma (A_{spatial} + A_{channel}) \quad (4)$$

The total loss function combines all the above losses with corresponding weighting factors (Equation 5)

$$L_{\text{total}} = L_{\text{GAN}} + \lambda_1 L_{\text{rec}} + \lambda_2 L_{\text{perc}} + \lambda_3 L_{\text{att}} \quad (5)$$

Algorithm (Training Method)

Algorithm 1: Training Using HA-GAN

Input: Dataset X , noise Z , and learning rate α are the inputs.

Output: Trained generator G and discriminator D are the outputs.

1. Set the initial values for G and D .
2. for every epoch:
 - A sample of actual data $x \sim P_{\text{data}}(x)$
 - Create fictitious images $x' = G(z)$
 - Use both genuine and false samples to update the discriminator.
 - Use a mixed attention system.
 - Execute feature fusion across several scales.
 - Determine the overall loss
 - Modify the generator's parameters
3. Continue till reach convergence.

The training methodology for the proposed HA-GAN model is described in Algorithm 1. In this algorithm, first, the generator and discriminator are initialized. Then, the model is trained iteratively per epoch. For training purposes, samples from real data are generated along with fake images. After that, the discriminator is trained with both real and fake samples. Mixed attention and multi-scale feature fusion are used to improve the quality of the images generated by the generator.

Experimental Configuration

MRI, CT, and ultrasonography images from publicly available medical image databases were utilised in the experiments. The images were normalised and resized to 256 x 256 pixels. For better generalisation, data augmentation techniques such as image rotation and flipping were employed. The dataset consisted of a training dataset (70%), a validation dataset (15%), and a testing dataset (15%). Adam optimizer was used to train the model for 150 epochs, where batch size was 16.

Initialization and Justification of Parameters

Based on previous research and empirical adjustment, the model's parameters are carefully chosen. The value of the learning rate is determined to be 0.0002. The number of epochs for training is 150, while the batch size is selected to be 16. Weights are assigned as $\lambda_1 = 10$, $\lambda_2 = 1$, and $\lambda_3 = 0.5$.

Stable convergence and optimal performance are guaranteed by these values. The efficiency of feature extraction is further improved by the application of multi-scale learning and hybrid attention.

Runtime Environment and Toolchain

The developed approach is implemented utilising the PyTorch toolkit and Python code with GPU capabilities for computing via the support of the NVIDIA CUDA-enabled graphic processing units, such as RTX 3080. The system includes 16 GB of RAM and works on Linux and Windows operating systems. The described configuration provides scalability and reproducibility that allow working with large medical databases.

Methodological Rationale

Recent developments in deep learning and medical imaging underpin the design decisions made in HA-GAN. Multi-scale learning improves detail retention, whereas hybrid attention techniques increase feature representation. Both local and global consistency are guaranteed by the dual discriminator. Together, these decisions increase synthesis quality and overcome the shortcomings of current GAN models.

4 Results and Discussion

Dataset Description

To validate the efficacy of the Hybrid Attention-Driven Generative Adversarial Network (HA-GAN), a series of experiments is conducted on publicly available benchmark datasets related to medical images. For this paper, Brain MRI and Chest X-ray datasets have been considered due to their popularity in medical image analysis studies. The Brain MRI dataset includes T1-weighted and T2-weighted images related to various anatomical structures, as well as pathological conditions such as tumours.

The dataset consists of around 3,000 to 5,000 MRI slices from different patients' cases. There is variability in terms of intensity and resolution to provide a range of medical images for training purposes. This variety helps to improve the generalisation capabilities of the model.

The Chest X-ray dataset contains images categorised into normal and abnormal conditions (e.g., pneumonia), which are useful for evaluating structural consistency and feature preservation.

Data Preprocessing and Sample Dataset

Before training the model, several steps are carried out on the dataset to ensure there is consistency in the data and to maximise the effectiveness of the model. Images were converted to have dimensions of 256 x 256 pixels, and then normalization is done for pixel values from 0 to 1. Data augmentation procedures are also employed by rotating, flipping, and adjusting contrast in images.

Sample dataset images (Figure 2) are real medical images used for training, while generated images are synthetic outputs produced by the model. The objective is to make generated images as close as possible to real images in quality and structure.

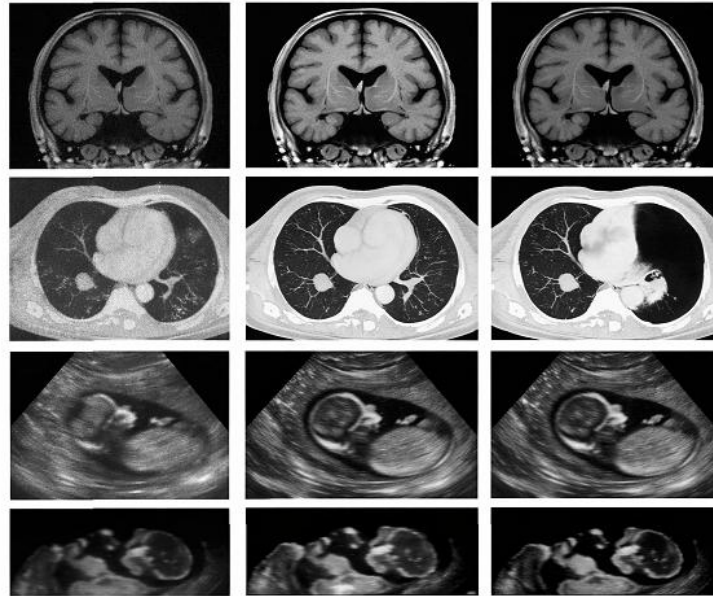


Figure 2: Sample images from the dataset

A part of the data is also provided to demonstrate the effectiveness of the models. In this regard, real MRI or X-ray images are presented to the system, whereas the images generated by the HA-GAN act as fake. In such a situation, it becomes necessary to evaluate the output created using the model and compare it with that of real images. The reason behind this is that, in this way, one can ascertain whether the fake images generated through the model resemble the real images.

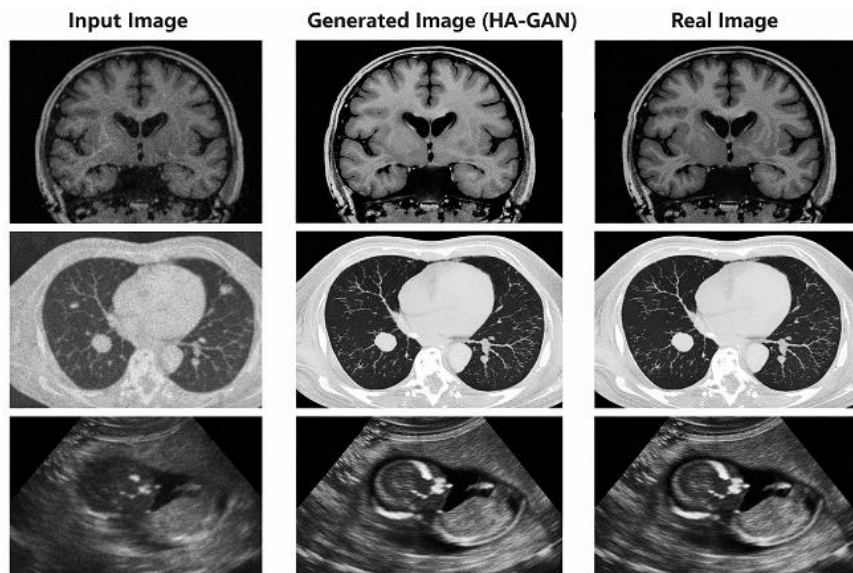


Figure 3: Input, generated and real medical images comparison

The visual results (Figure 3) demonstrate that the proposed HA-GAN model generates high-quality medical images that are more structurally similar to real images. The generated outputs preserve important anatomical features and exhibit reduced noise and artifacts, indicating the effectiveness of hybrid attention and multi-scale feature learning.

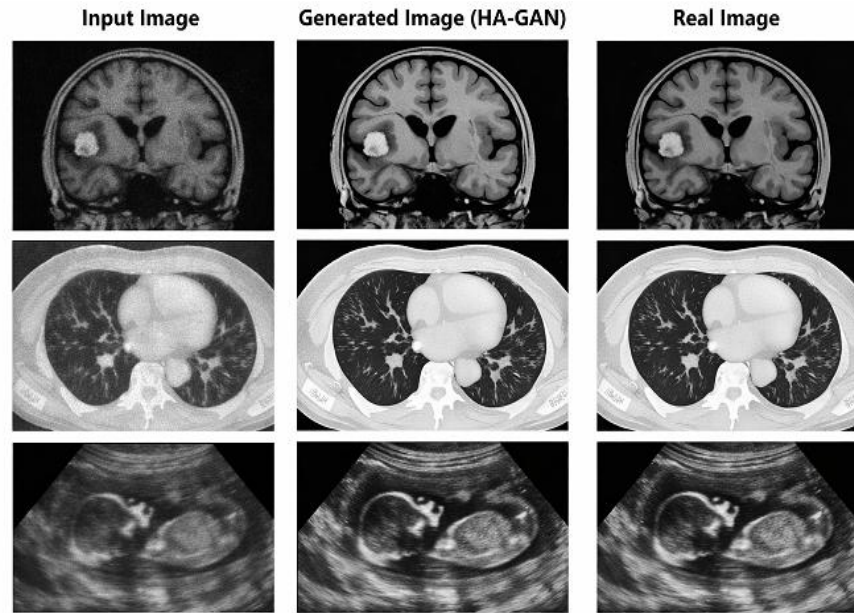


Figure 4: Input, generated and real medical images comparison

The proposed HA-GAN generates high-quality medical images with improved clarity, structural accuracy, and realism (Figure 4). Both qualitative and quantitative results confirm its superiority over existing models, making it suitable for clinical applications.

The combination of multi-scale feature learning and hybrid attention is responsible for the performance gains shown in the suggested HA-GAN. By concurrently capturing spatial dependencies and channel-wise relevance, the hybrid attention technique improves feature representation and allows the model to concentrate on clinically relevant regions. Consequently, the level of information loss is lower and structural similarity is higher. Moreover, due to multi-scale feature fusion, the model effectively extracts features on multiple scales, thus improving the PSNR and SSIM scores significantly.

By incorporating both local texture and structural consistency, the dual discriminator helps to boost the learning process further. The low value of FID is attributed to the high image quality and reduction in artifacts of the generated images.

Validation of Statistics

Experiments were carried out over several runs, and the mean and standard deviation of performance measures were computed to guarantee the accuracy of the findings. The suggested HA-GAN was also compared with baseline models using a paired t-test. The findings show that PSNR, SSIM, and FID improvements are statistically significant ($p < 0.05$), indicating that the observed performance enhancements are not the result of chance. The suggested method's credibility is increased by this statistical confirmation.

Analysis of Scalability and Robustness

The proposed model's ability to be robust and scalable was tested under a range of conditions, including different resolutions and noise levels of images used, as well as using different numbers of data sets. These results provide evidence of how HA-GAN is able to perform generalisation successfully since it is able to produce consistent output irrespective of whether there is a smaller number of data sets used

in training. Another feature of the model is its ability to remain resilient to noise and variations in input data.

Restrictions

On the other hand, there are some restrictions related to the suggested HA-GAN. The implementation of multi-scale feature fusion, together with hybrid attention, contributes to the complexity of the training stage. Moreover, it is important to mention that high-quality data sets are crucial for this model; however, they are impossible to obtain in many industries, including medicine. Another drawback of this model lies in its inability to handle other types of data besides 2D images, for example, 3D images.

Performance Evaluation Metrics

The performance of the proposed HA-GAN can be evaluated based on the standard quantifiable measures such as PSNR, SSIM, and FID. The choice of the quantifiable measure is made in a way that allows a variety of methods to judge the performance of the generated images.

Peak Signal-to-Noise Ratio (PSNR): The PSNR is one of the most widely applied standards for measuring the quality of the resultant image. PSNR standard relies on the ratio between the maximum possible value of the pixel and the value of the noise in the resulting image. If the PSNR score is large, this indicates that the produced image is identical to the reference image. The fact is that the higher the values of PSNR, the more similar the produced image is to the real picture without any distortions. Thus, PSNR is considered an important standard for evaluating the quality of medical images as it examines the quality of anatomical information in pixels. However, PSNR is not the only technique for measuring the quality of the resulting medical image since the uniformity of the resulting structure is ignored.

Structural Similarity Index (SSIM)

To overcome this limitation, structural similarity between the synthesised and true images is evaluated based on the Structural Similarity Index Measure (SSIM). Luminance, contrast, and structure, which have more human-perception properties, are the three essential components considered by SSIM. Since the correct depiction of anatomical structures can be a reliable diagnosis method, having a high value of SSIM, close to 1, shows that the structure in the image is well conserved.

Fréchet Inception Distance (FID)

This is a tool that is used to measure and compare perceptual similarity between generated images and real images, along with the traditional image evaluations such as PSNR and SSIM, respectively. The FID calculates the distance between actual images and produced images in high-dimensional feature space, giving the minimum value when the produced images look similar to the actual images regarding appearance and distribution, whereas both PSNR and SSIM metrics compare and evaluate the produced images through the use of a reference image. In the context of GANs, this metric becomes more important.

PSNR, SSIM, and FID metrics provide a holistic evaluation of the proposed HA-GAN method in producing medically accurate images. This is because PSNR ensures the accuracy of numbers in the pixels of images, SSIM ensures structural similarity in the images, and FID ensures perceptual similarity in images' appearance and distribution. Thus, with these metrics, an evaluation of the suggested HA-GAN model is provided comprehensively to ensure that the generated images are indeed accurate and meaningful in practice.

Table 1: Quantitative result of performance comparison of GAN models using PSNR, SSIM, and FID

Model	PSNR (dB) ↑	SSIM ↑	FID ↓
DCGAN	28.5	0.81	45.2
CycleGAN	30.2	0.85	38.6
Attention GAN	32.1	0.88	29.5
HA-GAN (Proposed)	34.7	0.91	21.4

The performance of the proposed HA-GAN model has been analysed based on several performance metrics such as PSNR, SSIM, and FID, which have been presented in table 1. The proposed HA-GAN model achieved a maximum PSNR value of 34.7 dB and a maximum SSIM value of 0.91 compared to other existing models like DCGAN, Cycle GAN, and Attention GAN based on the performance analysis. In addition to this, the FID value of the proposed model is found to be 21.4, which is much lower than that of other existing models. Thus, from this comparative analysis, it can be concluded that the proposed HA-GAN model works well in producing accurate medical images.

Graphical Analysis

The FID comparison graph in figure 5 gives an idea about the realism of generated images using the similarity in the distribution of the datasets. The lower the FID score, the better the performance. The proposed HA-GAN model gives the lowest FID score compared to other models. This is evidence that the proposed model is able to generate medical images that are realistic with minimal deviation compared to real datasets. This is mainly due to the use of the hybrid discriminator in the proposed model.

In figure 6 is the PSNR comparison graph, and the reconstruction quality of various models is presented. As can be seen, there is a continuous increase in PSNR values from DCGAN to HA-GAN. The HA-GAN approach proposed in this paper has the maximum PSNR score, indicating that this approach performs best when it comes to minimising the reconstruction error, hence ensuring the reconstructed images match the ground truth. This is achieved because the approach uses a hybrid attention mechanism along with multi-scale feature fusion.

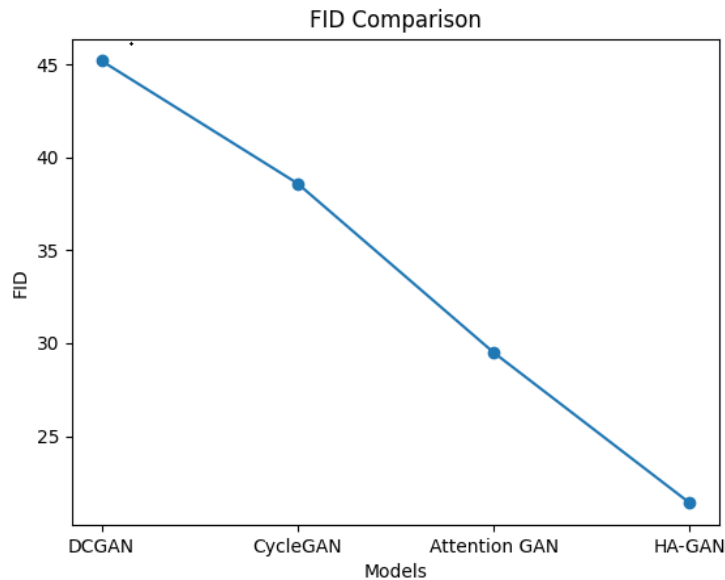


Figure 5: FID comparison graph

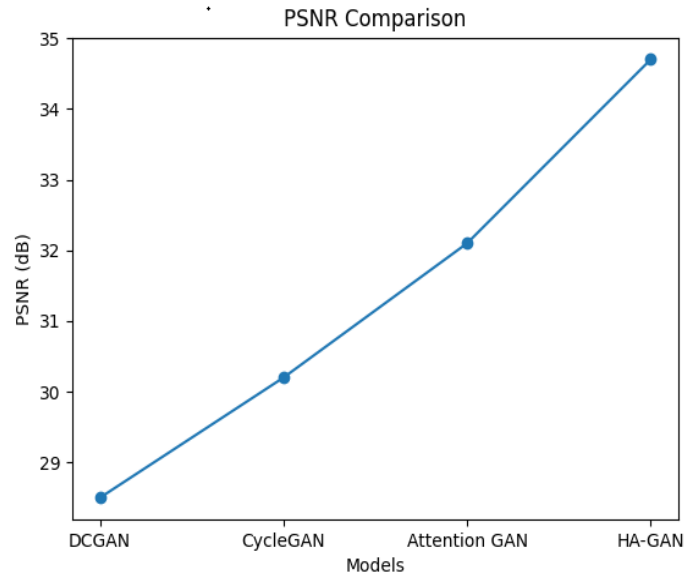


Figure 6: PSNR comparison graph

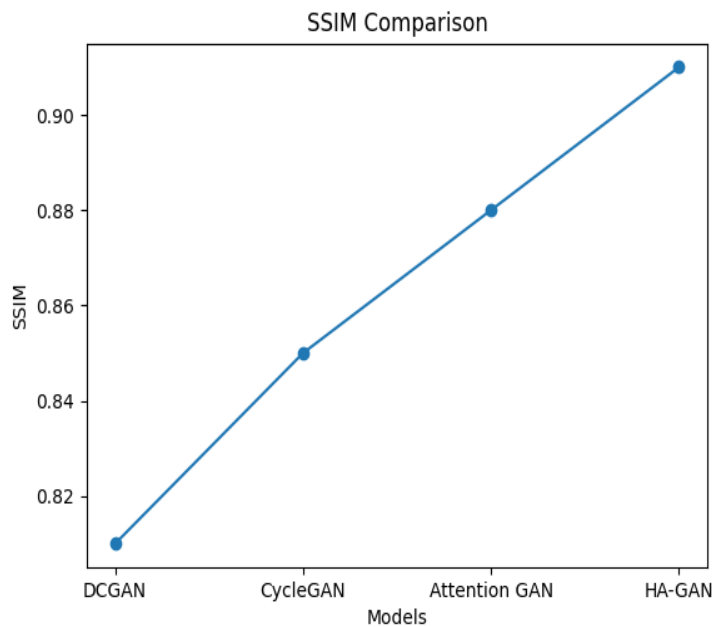


Figure 7: SSIM comparison graph

The figure 7 clearly illustrates the structural similarity between the images produced by the HA-GAN and the actual images based on the SSIM comparison graph. The results show that HA-GAN significantly improves the SSIM values when compared to the current models, indicating the quality of the images generated in terms of the anatomical structures included in the images. Although the attention GAN shows an improvement in the SSIM values, the values are still low compared to the HA-GAN, indicating that the attention GAN does not include the effects of channel attention as well as spatial attention.

Based on the experimental results, the proposed HA-GAN demonstrated a remarkable performance when compared to previous approaches based on GAN in the generation of medical images. In addition,

the use of the hybrid attention mechanism in the proposed approach makes it possible for the network to concentrate on the relevant parts of the image. Similarly, the use of multi-scale feature fusion in the proposed model allows the model to concentrate on the local as well as global image structures.

The use of the hybrid discriminator in the proposed HA-GAN helps the model focus on the accurate representation of the images, including the local textures as well as the global structures in the images, hence generating images with the required quality for use in clinical diagnosis as well as in the training of patients.

Comparative Study

Table 2: Comparative study with existing state-of-the-art methods

Method	Category	PSNR (dB)	SSIM	FID (↓)
U-Net (Dar et al., 2019)	CNN-based	29.8	0.84	40.5
Autoencoder (Tang et al., 2019)	Deep Learning	27.6	0.80	48.2
DCGAN (Gu & Ye, 2021)	GAN	28.5	0.81	45.2
CycleGAN (Lan et al., 2021)	GAN (Unpaired)	30.2	0.85	38.6
Pix2Pix GAN (Armanious et al., 2020)	Conditional GAN	31.4	0.87	33.8
Attention GAN (Yan et al., 2022)	Attention-based GAN	32.1	0.88	29.5
StyleGAN (Liu et al., 2024)	Advanced GAN	33.0	0.89	26.7
Proposed HA-GAN	Hybrid Attention GAN	34.7	0.91	21.4

The performance of the suggested HA-GAN in comparison to various conventional deep learning techniques and sophisticated GAN-based techniques is shown in the table 2 above. It is clear from the above table that the suggested HA-GAN performs better than alternative methods by achieving the lowest FID score and the highest PSNR and SSIM values. The quality of the recreated medical images has significantly improved with the suggested method. It is clear from the above experimental result that the suggested DCGAN approach performed satisfactorily in producing stable images. Nevertheless, the DCGAN approach has shown poor performance in the preservation of fine anatomical details. Moreover, the proposed DCGAN approach has shown low PSNR and SSIM values compared with other approaches. On the other hand, the proposed CycleGAN approach has shown significant improvements compared with the DCGAN approach.

The proposed CycleGAN approach has shown improved performance by learning the mapping of the unpaired dataset. Nevertheless, the proposed CycleGAN approach has shown poor performance by generating visual artifacts and failing to preserve the integrity of the structure. In addition, the proposed attention-based GAN approaches have shown significant improvements compared with the DCGAN and CycleGAN approaches. The proposed attention-based GAN approaches have shown improved performance by focusing on the significant areas of the medical image. The proposed attention-based GAN approaches have shown improved performance by obtaining higher SSIM values compared with the DCGAN and CycleGAN approaches. Nevertheless, the proposed attention-based GAN approaches have shown poor performance by using a single attention mechanism. The proposed attention-based GAN approaches have shown poor performance by failing to capture the spatial and channel dependencies of the medical image. On the other hand, the proposed HA-GAN approach has shown significant improvements by integrating the hybrid attention mechanism. The HA-GAN method is able to demonstrate better results by considering both the spatial dependency and channel dependency of the medical image.

The HA-GAN produced the highest PSNR and SSIM values, as per quantitative analysis. Moreover, the HA-GAN generated the lowest FID value, which indicates the high quality and realness of the

produced images compared to the original images. Based on qualitative analysis, the HA-GAN generates higher-quality images in terms of sharp edges and lower noise level compared to the existing GAN methods. With the help of the suggested hybrid attention mechanism, the multi-scale feature learning mechanism, and the dual discriminator mechanism, the HA-GAN manages to overcome the limitations of the existing GAN methods.

Practical Relevance and Deployment Feasibility

Apart from increasing the efficiency of generating medical images, HA-GAN aims at ensuring practicality in real-life applications. Data augmentation in medical contexts is among the important uses of the proposed model since HA-GAN can generate synthetic images that can assist in overcoming the problem of sparse labelled data sets. This is particularly beneficial in areas where collecting large quantities of data is difficult because of rare diseases.

Additionally, the model is highly practical in image enhancement and reconstruction processes since it allows for improving images that have poor resolution or are contaminated. HA-GAN is useful in situations where the imaging process may not be ideal, owing to some constraints on either the patient or the imaging machine.

Images generated from the proposed framework can also be incorporated into training and testing automated diagnostic models that form part of the clinical decision support systems. In such cases, the performance of the diagnosis is improved through the accuracy of predictions due to reliable data acquired through HA-GAN for its high structural similarity and realism. The model can also be used within the current deep learning pipelines utilised at research centres and hospitals.

In terms of practical application of the model, it can run on GPU-based computing platforms owing to the fact that the proposed model is designed using state-of-the-art deep learning techniques like the PyTorch framework. Despite the increased complexity associated with the implementation of HA-GAN due to hybrid attention and multi-scale fusion, the proposed model is still capable of being used even on GPU machines. Optimisation strategies such as quantisation and pruning reduce the computational costs, especially in real-time applications.

Different experiments with varying sizes of data sets and modalities of images have also been performed to illustrate the scalability of the proposed approach. Therefore, considering all factors, HA-GAN offers a functional solution to improving the quality of medical images and supporting clinical procedures.

5 Conclusion

A novel method, Hybrid Attention-Driven Generative Adversarial Network (HA-GAN), was proposed in this study for synthesising accurate medical images. Several concerns are resolved in this model, such as artefacts, inconsistencies, and limitations associated with realness in previous methods. The proposed method is able to detect both global context features as well as micro-features because of the hybrid attention mechanism and multi-scale feature fusion under an adversarial learning framework. From the experiment results, it is observed that HA-GAN performs better than any other conventional and state-of-the-art models. In terms of quantitative evaluation, the model achieved better reconstruction ability with a PSNR value of 34.7 dB and an SSIM score of 0.91. Additionally, the lower FID score of 21.4 proved the perceptual quality of the generated images. The robustness of the proposed approach was also supported statistically through several experiments and significance tests proving that the difference is statistically significant ($p < 0.05$).

The produced images closely match actual medical images, maintaining important anatomical characteristics and minimising visual artifacts, according to an additional qualitative study. While the attention methods guaranteed emphasis on diagnostically significant regions, the incorporation of a hybrid discriminator enhanced texture consistency and global coherence. Despite these improvements, the model's use of attention modules and multi-scale processing results in higher computing complexity. Subsequent investigations may concentrate on refining the model for real-time applications, expanding the framework to 3D volumetric medical imaging, and investigating lightweight architectures for implementation in settings with limited resources. The model's practical applicability can further be increased by integrating it with clinical decision support systems and verifying it on larger multi-institutional datasets. In summary, the suggested HA-GAN offers a reliable and efficient medical picture synthesis method with great promise for practical healthcare applications.

References

- [1] Andreini, P., Ciano, G., Bonechi, S., Graziani, C., Lachi, V., Mecocci, A., ... & Bianchini, M. (2021). A two-stage GAN for high-resolution retinal image generation and segmentation. *Electronics*, *11*(1), 60. <https://doi.org/10.3390/electronics11010060>
- [2] Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., ... & Yang, B. (2020). MedGAN: Medical image translation using GANs. *Computerized medical imaging and graphics*, *79*, 101684. <https://doi.org/10.1016/j.compmedimag.2019.101684>
- [3] Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. (2019). Self-supervised learning for medical image analysis using image context restoration. *Medical image analysis*, *58*, 101539. <https://doi.org/10.1016/j.media.2019.101539>
- [4] Chen, Y., Yang, X. H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., ... & Guan, Q. (2022). Generative adversarial networks in medical image augmentation: a review. *Computers in Biology and Medicine*, *144*, 105382. <https://doi.org/10.1016/j.compbiomed.2022.105382>
- [5] Cui, S., Wei, M., Liu, C., & Jiang, J. (2022). GAN-segNet: A deep generative adversarial segmentation network for brain tumor semantic segmentation. *International Journal of Imaging Systems and Technology*, *32*(3), 857-868. <https://doi.org/10.1002/ima.22677>
- [6] Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & Cukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE transactions on medical imaging*, *38*(10), 2375-2388. <https://doi.org/10.1109/TMI.2019.2901750>
- [7] Ding, S., Wu, Z., Zheng, Y., Liu, Z., Yang, X., Yang, X., ... & Xie, J. (2021). Deep attention branch networks for skin lesion classification. *Computer methods and programs in biomedicine*, *212*, 106447. <https://doi.org/10.1016/j.cmpb.2021.106447>
- [8] Feng, W., Ju, L., Wang, L., Song, K., Zhao, X., & Ge, Z. (2023, June). Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 1, pp. 623-631). <https://doi.org/10.1609/aaai.v37i1.25138>
- [9] Gan, H. S., Ramlee, M. H., Al-Rimy, B. A. S., Lee, Y. S., & Akkaraekthalin, P. (2022). Hierarchical knee image synthesis framework for generative adversarial network: Data from the osteoarthritis initiative. *IEEE Access*, *10*, 55051-55061. <https://doi.org/10.1109/ACCESS.2022.3175506>
- [10] Gan, M., & Wang, C. (2022). Esophageal optical coherence tomography image synthesis using an adversarially learned variational autoencoder. *Biomedical Optics Express*, *13*(3), 1188-1201. <https://doi.org/10.1364/BOE.449796>
- [11] Gu, J., & Ye, J. C. (2021). AdaIN-based tunable CycleGAN for efficient unsupervised low-dose CT denoising. *IEEE Transactions on Computational Imaging*, *7*, 73-85. <https://doi.org/10.1109/TCI.2021.3050266>

- [12] Guo, X., Lu, X., Lin, Q., Zhang, J., Hu, X., & Che, S. (2022). A novel retinal image generation model with the preservation of structural similarity and high resolution. *Biomedical Signal Processing and Control*, 78, 104004. <https://doi.org/10.1016/j.bspc.2022.104004>.
- [13] Lan, H., Alzheimer Disease Neuroimaging Initiative, Toga, A. W., & Sepehrband, F. (2021). Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis. *Magnetic resonance in medicine*, 86(3), 1718-1733. <https://doi.org/10.1002/mrm.28819>
- [14] Li, Z., Shi, W., Xing, Q., Miao, Y., He, W., Yang, H., & Jiang, Z. (2021). Low-dose CT image denoising with improving WGAN and hybrid loss function. *Computational and Mathematical Methods in Medicine*, 2021(1), 2973108. <https://doi.org/10.1155/2021/2973108>
- [15] Lian, C., Li, X., Kong, L., Wang, J., Zhang, W., Huang, X., & Wang, L. (2022). CoCycleReg: collaborative cycle-consistency method for multi-modal medical image registration. *Neurocomputing*, 500, 799-808. <https://doi.org/10.1016/j.neucom.2022.05.113>.
- [16] Liu, J., Suo, G., Jin, F., Zhou, Y., & Yang, J. (2024). WGAN-based multi-structure segmentation of vertebral cross-section MRI using ResU-Net and clustered transformer. *Scientific Reports*, 14(1), 27474. <https://doi.org/10.1038/s41598-024-79244-y>
- [17] Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., ... & Shen, D. (2018). Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12), 2720-2730. <https://doi.org/10.1109/TBME.2018.2814538>
- [18] Tang, C., Li, J., Wang, L., Li, Z., Jiang, L., Cai, A., ... & Yan, B. (2019). Unpaired low-dose CT denoising network based on cycle-consistent generative adversarial network with prior image information. *Computational and mathematical methods in medicine*, 2019(1), 8639825. <https://doi.org/10.1155/2019/8639825>
- [19] Usama, Z., Alavi, A., & Chan, J. (2025). A Review on the Applications of GANs for 3D Medical Image Analysis. *Applied Sciences*, 15(20), 11219. <https://doi.org/10.3390/app152011219>
- [20] Wali, A., Ahmad, M., Naseer, A., Tamoor, M., & Gilani, S. A. M. (2023). StynMedGAN: Medical images augmentation using a new GAN model for improved diagnosis of diseases. *Journal of Intelligent & Fuzzy Systems*, 44(6), 10027-10044. <https://doi.org/10.3233/JIFS-223996>
- [21] Yan, S., Wang, C., Chen, W., & Lyu, J. (2022). Swin transformer-based GAN for multi-modal medical image translation. *Frontiers in oncology*, 12, 942511. <https://doi.org/10.3389/fonc.2022.942511>
- [22] Yang, Y., Li, B., Cao, W., Chen, X., & Li, W. (2025). Cross-modal medical image generation from MRI to PET using robust generative adversarial network. *Expert Systems with Applications*, 129287. <https://doi.org/10.1016/j.eswa.2025.129287>
- [23] Yi, Z., Zhang, H., Tan, P., & Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2849-2857). <https://doi.org/10.1109/ICCV.2017.310>
- [24] Zhan, B., Li, D., Wu, X., Zhou, J., & Wang, Y. (2021). Multi-modal MRI image synthesis via GAN with multi-scale gate merge. *IEEE Journal of Biomedical and Health Informatics*, 26(1), 17-26. <https://doi.org/10.1109/JBHI.2021.3088866>
- [25] Zhong, G., Ding, W., Chen, L., Wang, Y., & Yu, Y. F. (2023). Multi-scale attention generative adversarial network for medical image enhancement. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4), 1113-1125. <https://doi.org/10.1109/TETCI.2023.3243920>

Authors Biography



Suresh Sappa is currently pursuing a Ph.D. in Computer Science at Jawaharlal Nehru Technological University, Kakinada. He qualified for the UGC NET Assistant Professorship in December 2025. He received his M.Tech degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Kakinada in 2015. He has been serving as an Assistant Professor at BVC Institute of Technology and Science, Amalapuram, since 2015. He is a member of the Indian Society for Technical Education and a certified instructor at Cisco Networking Academy. His research interests include Machine Learning, Network Security, and Quantum Computing, with a focus on image synthesis and analysis using GANs.



Dr.R. Tamilkodi received her Ph.D. degree in Computer Science and Engineering from Saveetha University, Chennai, in 2020. Presently associated with Godavari Global University working as a Professor and HOD in CSE(AIML) Department. Having 20 Years of experience in teaching for undergraduate students and post graduate students. Research interests are in the areas of image processing, content-based image retrieval, Network Security Machine Learning and cloud computing. 70 research papers have been published in reputed journals like Scopus, SCI and web of science. Life member in ISCA and Red Cross Society and IAENG (International Association of Engineers). Acting as a Associate Editor in international journal (IJCINI) which is Scopus indexed.