Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n34.2025.09
pISSN: 1390-650X / eISSN: 1390-860X

Exploring deep generative models for improved data generation in hypertrophic cardiomyopathy

Exploración de modelos generativos profundos para una mejor generación de datos en la miocardiopatía hipertrófica

Swarajya Madhuri Rayavarapu[1,*], Gottapu Sasibhushana Rao[1]

Received: 06-08-2024, Received after review: 05-06-2025, Accepted: 10-06-2025, Published: 01-07-2025

Abstract

Data generation strategies are essential for addressing the challenge of limited training data in deep learning-based medical image analysis, particularly for hypertrophic cardiomyopathy (HCM) using magnetic resonance imaging (MRI). Unlike traditional augmentation techniques, deep generative models can synthesize novel and diverse MRI images, enriching the training datasets. This study evaluates several generative models, including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Deep Convolutional GANs (DCGANs), Auxiliary Classifier GANs (ACGANs), InfoGANs, and Diffusion Models, using the Structural Similarity Index Measure (SSIM) and Cross-Correlation Coefficient (CC) to assess image quality and structural fidelity. While VAEs demonstrated limitations such as noticeable noise and blurriness, GAN-based models, particularly DCGANs and ACGANs, generated higher-quality and anatomically accurate images. Diffusion models achieved the highest image fidelity among all the methods evaluated, but required longer generation times. These findings underscore the trade-off between image quality and computational efficiency and highlight the potential of deep generative models to augment MRI datasets, thereby improvingdeep learning applications for HCM diagnosis.

Keywords: Data Generation, Diffusion models, Generative Adversarial networks, Variational autoencoders.

Resumen

Las estrategias de generación de datos son fundamentales para superar el desafío de los datos de entrenamiento limitados en el análisis de imágenes médicas basado en aprendizaje profundo, en particular para la miocardiopatía hipertrófica (HCM) mediante resonancia magnética (MRI). A diferencia de los métodos de aumento tradicionales, los modelos generativos profundos pueden sintetizar imágenes de MRI novedosas y diversas. Este estudio evalúa múltiples modelos generativos: autocodificadores variacionales (VAE), redes generativas adversarias (GAN), GAN convolucionales profundas (DCGAN), GAN con clasificador auxiliar (ACGAN), InfoGAN y modelos de difusión, utilizando el índice de similitud estructural (SSIM) y el coeficiente de correlación cruzada (CC) para evaluar la calidad de imagen y la fidelidad estructural. Si bien los VAE mostraron limitaciones como el ruido y la borrosidad, los modelos basados en GAN, especialmente DCGAN y ACGAN, produjeron imágenes de mayor calidad y precisión anatómica. Los modelos de difusión lograron la mayor fidelidad de imagen, aunque a expensas de tiempos de generación más prolongados. Estos resultados destacan la compensación entre la calidad de imagen y la eficiencia computacional, y demuestran el potencial de los modelos generativos para ampliar los conjuntos de datos de MRI, mejorando así las aplicaciones de aprendizaje profundo para el diagnóstico de HCM.

Palabras clave: generación de datos, modelos de difusión, redes generativas adversarias, codificadores automáticos variacionales

[1,*]Department of Electronics and Communication Engineering, Andhra University, India.
Corresponding author ✉: madhurirayavarapu.rs@andhrauniversity.edu.in.

Suggested citation: S. Madhuri Rayavarapu and G. Sasibhushana Rao, "Exploring deep generative models for improved data generation in hypertrophic cardiomyopathy," *Ingenius, Revista de Ciencia y Tecnología*, N.° 34, pp. 116-125, 2025, DOI: https://doi.org/10.17163/ings.n34.2025.09.

1. Introduction

Artificial intelligence (AI) has become an increasingly powerful tool for analyzing medical data, including images, text, and signals, to support disease interpretation, identification, classification, and diagnosis [1]. The development of AI-based medical software relies heavily on large volumes of diverse data, such as electrocardiograms, X-rays, magnetic resonance imaging (MRI), computed tomography (CT), echocardiography, and dermatological images [2–8]. However, the collection and annotation of these datasets remain costly and time-consuming, often requiring expert clinicians to manually label the data [9].

Although numerous medical datasets are available, publicly accessible data for rare cardiac conditions such as hypertrophic cardiomyopathy (HCM) remain scarce. This limitation stems from factors including the low prevalence of the disease, privacy restrictions, and class imbalance within existing datasets. Consequently, there is an urgent need for advanced techniques capable of generating high-quality synthetic medical images to augment current datasets. Deep learning models, particularly those developed for medical image analysis, benefit greatly from large, balanced, and precisely labeled datasets, which improve diagnostic accuracy and generalizability [10]. Moreover, diverse training data are essential for mitigating overfitting and enhancing model robustness.

Developing effective deep learning models for medical imaging poses several challenges. First, acquiring sufficient data is difficult due to patient confidentiality constraints and limited data sharing among institutions. Second, manual annotation of medical images is labour-intensive, time-consuming, and susceptible to variability across different imaging modalities. Third, datasets often exhibit class imbalance, as pathological cases typically constitute a minority compared to normal or healthy instances [11]. Collectively, these challenges complicate the development of reliable automated diagnostic systems.

To address these issues, researchers have adopted data augmentation techniques to increase dataset size and variability. Traditional augmentation methods include geometric transformations, kernel operations, and colour space adjustments [12–14]. While useful, these methods yield only limited variation because they manipulate existing images rather than creating entirely new ones. In contrast, deep learning-based augmentation approaches, such as Variational Autoencoders (VAEs), Neural Style Transfer, Diffusion Models, and Generative Adversarial Networks (GANs), have demonstrated superior capability in generating diverse and realistic synthetic images [15–17]. Unlike conventional methods, these generative models learn the underlying high-dimensional data distribution, enabling the creation of novel samples that closely resemble real-world data. As a result, deep generative models are particularly effective for mitigating data scarcity and class imbalance, especially for rare diseases such as HCM, thereby enhancing the training and generalization performance of diagnostic models.

Section 2 reviews the deep generative models used for synthetic image generation and describes the methodology, section 3 presents the results and discussion, and section 4 summarizes the conclusions.

2. Materials and methods

2.1. Deep generative models

This section provides a concise description of the different deep generative models employed for generating medical images. It focuses on three main types of models: Variational Autoencoders, Generative Adversarial Networks, and Diffusion Models.

2.1.1. Variational autoencoder

Max Welling and Diederik P. Kingma introduced the concept of the variational autoencoder (VAE) in 2013 [18]. A VAE describes an observation in latent space in a probabilistic manner. Instead of producing a single value for each latent state attribute, the encoder generates a probability distribution for each attribute. Applications of VAEs include data compression and the generation of synthetic data. Figure 1 illustrates the VAE architecture, with its main components detailed below.

- **Input**: The input to a VAE depends on the specific application and the domain of interest. For image-based VAEs, the input typically consists of entire images or image patches.
- **Encoder**:The encoder transforms the input data into the parameters of the latent space, which define the associated probability distribution. It typically consists of multiple neural network layers, such as convolutional or fully connected layers, that work together to reduce the dimensionality of the input. The output of the encoder is a set of mean and variance vectors that approximate the characteristics of a multivariate Gaussian distribution in the latent space.
- **Latent space**: Each point in the latent space represents a latent code, which is a reduced-dimensional description of the input data. During training, the encoder learns to generate latent codes that accurately capture the essential features of the input.
- **Decoder**: Before mapping the transformed data back to the input space, the decoder extracts a

relevant subset from the latent space. Like the encoder, it consists of several neural network layers that progressively sample the latent code and generate the output. The goal of the decoder is to reproduce the original input data as accurately as possible.

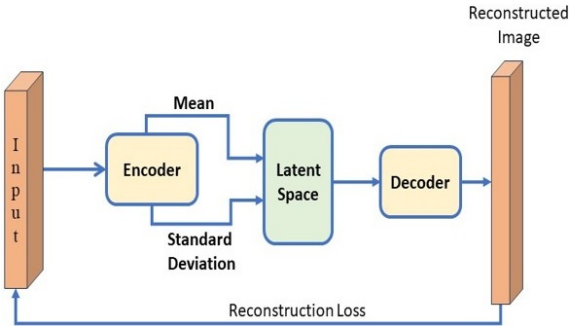

Figure 1. Variational autoencoder

VAEs rely on the mathematical representation of the latent space learned by the encoder network to approximate the data distribution, while the decoder network uses this representation to generate samples similar to the training data.

The encoder network maps an input sample x to a latent representation z according to Equation (1), while the decoder network uses Equation (2) to reconstruct the input space from the latent representation z. The functions f and g denote the encoder and decoder networks, respectively.

$$z = f(x) \tag{1}$$

$$x' = g(z) \tag{2}$$

The loss function of the VAE is defined in Equation (3). In this equation, the first term represents the reconstruction loss, while the second term corresponds to the Kullback–Leibler (KL) divergence.

$$Loss = E\left[L\left(x, x'\right)\right] - D_{KL}\left(q\left(z \mid x\right) || \, p\left(z\right)\right) \tag{3}$$

$$Where \; L\left(x, x'\right) = logp\left(x \mid x'\right) \tag{4}$$

2.1.2. Generative adversarial networks

Generative Adversarial Networks (GANs) are an unsupervised learning method that leverages the well-established zero-sum game theory framework for two players. This concept was introduced by Goodfellow in 2014 [19]. In a GAN, the generator creates new samples based on real data, while the discriminator estimates the underlying data distribution by distinguishing between real and generated samples, see figure 2.

- **Generator**: The generator component of a GAN creates synthetic data by transforming random noise into samples that resemble real data.

- **Discriminator**: The discriminator component of a GAN acts as a classifier that distinguishes between real data and artificially generated data produced by the generator.

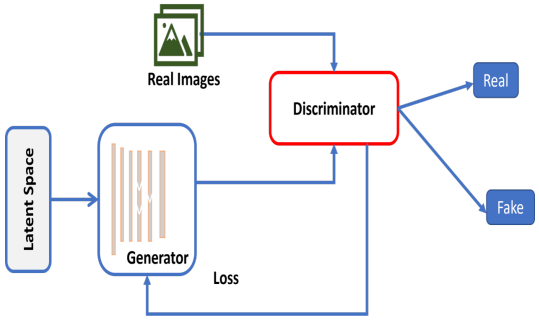

Figure 2. Architecture of a generative adversarial network (GAN)

GANs operate based on a mathematical framework in which the discriminator network provides feedback on the realism of generated samples, while the generator network maps latent space representations to the original data space. This adversarial process enables GANs to learn a generative model of the data and produce diverse, realistic synthetic samples. Compared to other generative methods, GANs offer notable advantages, including the ability to handle complex data distributions and generate high-resolution images [20].

The loss function of the GAN is defined in Equation (5).

$$\begin{aligned} min_G max_D V(D,G) = E_x log(D(x)) + \\ E_z log(1 - D(G(z))) \end{aligned} \tag{5}$$

In Equation (5), Zepresents the latent space or noise input, which is supplied to the generator. The discriminator is denoted by D, and the generator is denoted by G. The discriminator receives both the generated images and the real data samples. The representation of real data is expressed as $D(x)$, while the representation of generated data is expressed as $D(G(z))$. Both the discriminator and generator networks are trained simultaneously: the discriminator aims to minimize the classification score of generated samples, whereas the generator seeks to maximize it.

Several variants of GANs have been introduced in recent years. Some of these variants are described in the following sections.

2.1.3. Deep convolutional GANs

Deep Convolutional Generative Adversarial Networks (DCGANs) are a novel variant of convolutional neural networks (CNNs) designed with specific architectural constraints, as introduced in [21]. To meet these requirements, DCGANs incorporate three key architectural modifications. First, they replace fully connected hidden layers and pooling layers with convolutional layers, using fractional-strided convolutions in the generator and strided convolutions in the discriminator to enhance network performance. Second, they apply ReLU activations to all layers of the generator except the output layer, while employing LeakyReLU activations throughout the discriminator. Third, batch normalization is applied consistently in both the generator and the discriminator.

2.1.4. Auxiliary classifier GAN

Odena et al. [22] introduced the Auxiliary Classifier Generative Adversarial Network (ACGAN), which incorporates an additional classifier to enhance the model's performance. The ACGAN discriminator includes a classifier that categorizes samples into discrete classes, thereby improving training stability. In an ACGAN, the generator G uses both noise Z and a category label C sampled from a distribution to generate a synthetic sample, denoted as $x_{fake} = G(c, z)$. The discriminator D distinguishes between real and fake samples while also considering both the authenticity and the class labels.

The objective functions of the ACGAN are defined in Equations (6) and (7).

$$L_s = E\left[log(P(S = real \mid Xreal))\right] + E\left[log(P(S = fake \mid Xfake))\right] \tag{6}$$

$$L_c = E\left[log(P(C = c \mid Xreal))\right] + E\left[log(P(C = c \mid Xfake))\right] \tag{7}$$

The terms L_s and L_c denote the probabilities of correctly identifying the source and the class, respectively. X denotes the input image, C is the class label, and S is the source.

When training the discriminator D, the primary objective is to maximize the total loss $(L_s + L_c)$. In contrast, the generator G is trained to maximize the difference $L_s - L_c$ between the losses.

2.1.5. Information Maximizing GAN (Info GAN)

InfoGAN was proposed by Chen et al. [23] as an information-theoretic approach to improve interpretability in GANs by learning meaningful latent variables. The term "info" indicates the mutual information shared between the generated distribution $G(z, c)$ and the latent code c. To reduce the correlation between x and c, InfoGAN maximizes the mutual information $I(c; G(z, c))$. A regularization term incorporating this information objective is added to the standard GAN loss function. To estimate a tractable lower bound of the mutual information $P(c \mid x)$, an auxiliary distribution $Q(c \mid x)$ is introduced to approximate $P(c \mid x)$.

The objective function of InfoGAN is defined in Equation (8):

$$\frac{min}{G, Q}\frac{max}{D}L(D, G, Q) = V(D, G) - \lambda I(G, Q) \tag{8}$$

Where λ is the regularization constant, typically set to one.

2.2. Diffusion models

Sohl-Dickstein et al. were the first to introduce diffusion models [24]. Building on this idea, Ho et al. [25] proposed Denoising Diffusion Probabilistic Models (DDPMs), marking the first demonstration that diffusion models can achieve performance comparable to other generative models for image synthesis tasks.

Diffusion models are advanced machine learning algorithms that generate high-quality data by gradually adding noise to a dataset and then learning how to reverse this process [26]. As a subset of deep learning-based generative models, their primary objective is to produce synthetic data that is both realistic and plausible given a set of input conditions. Due to their many advantages over other generative methods, such as the ability to generate highly diverse data, handle high-dimensional datasets, and learn complex distributions effectively, diffusion-based generative models have recently gained significant popularity across various scientific disciplines [27].

A diffusion model is a probabilistic generative framework that involves two multi-step processes: forward diffusion and reverse diffusion. In the forward diffusion process, noise is gradually added to the input data until the original information is completely obscured. In contrast, the reverse diffusion process employs a trainable neural network to progressively remove the noise and reconstruct the original data distribution. Synthetic outputs are generated by applying this trained neural network to noisy samples, see figure 3.

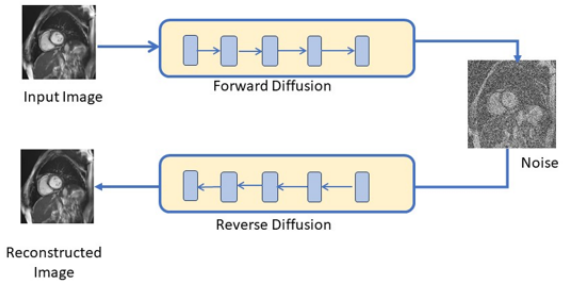

Figure 3. Architecture of a diffusion model

To initiate the forward diffusion process, samples are drawn from a simple distribution, typically a Gaussian. This initial sample then undergoes a sequence of small, reversible transformations. Through a Markov chain, each step incrementally increases the complexity of the sample, which can be interpreted as the progressive addition of structured noise.

In a forward diffusion process, small amounts of Gaussian noise are progressively added to a data point x drawn from the true data distribution $q(x)$, producing a series of increasingly noisy samples denoted by $x_1, x_2, x_3, \ldots, x_T$. This process is mathematically defined by Equations (9) and (10).

$$q(x_t \mid x_{t-1}) = N(x_i, \sqrt{1-\beta_t}x_{t-1,}\beta_t I) \tag{9}$$

$$q(x_{1:T} \mid x_0) = \prod_{t}^{T} = q(x_t \mid x_{t-1}) \tag{10}$$

The reverse diffusion process is defined by Equations (11) and (12).

$$p\theta(x_{0:T}) = p(x_T) \cdot \prod_{l=1}^{\bar{I}} p\theta(x_{t-1} \mid x_t) \tag{11}$$

$$p\theta(x_{t-1} \mid x_t) = W(x_{t-1}; u_\theta(x_t, t), \sum \theta(x_t, t)) \tag{12}$$

2.3. Methodology

Images of hypertrophic cardiomyopathy (HCM) are generated using various deep generative models, with Magnetic Resonance Imaging (MRI) serving as the primary modality for detecting and evaluating the disease. MRI provides detailed and high-resolution images of the heart's structure and function without the use of ionizing radiation, making it particularly well-suited for cardiac assessment. Key features of HCM, such as ventricular wall thickening and myocardial fibrosis, can be effectively evaluated through MRI scans. This imaging technique also enables assessment of disease severity and progression, supporting clinical decision-making. Furthermore, the MRI data serve as the basis for developing and validating automated techniques for HCM detection and analysis.

The dataset for hypertrophic cardiomyopathy was obtained from the Cardiac ACDC Dataset.

2.3.1. Dataset

The cardiac MRI scans used in this study were sourced from the open-source Cardiac ACDC dataset [28]. The real clinical examinations that form the ACDC dataset were provided by the University Hospital of Dijon. To protect patient privacy, all data underwent thorough anonymization and processing in accordance with the criteria established by the local ethics committee of the French Hospital of Dijon. The dataset includes a sufficient number of examples to train machine learning algorithms and to reliably assess changes in key physiological parameters derived from cine-MRI, such as diastolic volume and ejection fraction. It encompasses a range of diverse cardiac pathologies and is divided into five categories with an equal proportion of cases in each. In total, the dataset comprises 150 examinations, each obtained from a different patient.

2.3.2. Experimental setup and training process

The Python programming language and the PyTorch framework were used to implement the various deep generative models. Model training was performed in an Anaconda Navigator and Jupiter Notebook environment, with Graphics Processing Unit (GPU) acceleration enabled on an Intel I7 laptop.

Each generative model, including VAE, GAN, Deep Convolutional GAN (DCGAN), InfoGAN, ACGAN, and Diffusion Models, was trained using the Adam optimizer with an initial learning rate of $0,0002$, $\beta_1 = 0,5$ y $\beta_2 = 0,999$. The batch size was set to 32, and models were trained for up to 100 epochs unless early stopping was triggered based on validation loss. For the diffusion models, training was extended to approximately 115 epochs. For GAN-based architectures, the appropriate adversarial loss functions were applied; InfoGAN additionally optimized a mutual information loss term, and ACGAN incorporated an auxiliary classification loss. VAEs were trained by minimizing a combination of reconstruction loss and Kullback–Leibler (KL) divergence. Training stability was enhanced through techniques such as gradient penalty and label smoothing where applicable. Model checkpoints were saved regularly to preserve the best-performing weights according to validation metrics.

Upon completion of training, approximately 1,000 images were generated. These images were then subjected to a qualitative evaluation to assess their similarity to the original dataset.

2.3.3. Evaluation metrics

The SSIM, cross-correlation, and mean squared error (MSE) metrics were used to evaluate the deep generative models, as they capture different aspects of image quality and similarity that are crucial in medical imaging contexts.

2.3.4. Structural similarity index (SSIM)

The Structural Similarity Index (SSI), also known as SSIM, is a widely used metric for quantifying the similarity between two images [29]. SSIM evaluates the similarity of structural information by considering luminance, contrast, and structural patterns within the images. It works by performing three comparisons between corresponding patches: luminance comparison, contrast comparison, and structural comparison. These results are then combined to produce an overall SSIM index, which ranges from -1 to 1, where a value of 1 indicates perfect structural similarity.

For two images, x and y, SSIM is computed using the equation (13):

$$(x, y) = \frac{(2 \cdot \mu_x \mu_y + c_1) \cdot (2 \cdot \sigma_{xy} + c_2)}{(c_1 2 + \mu_x 2 + \mu_y 2) \cdot (c_2 2 + \sigma_x 2 + \sigma_y 2)} \quad (13)$$

$\sigma_y 2$, $\sigma_x 2$ are the variances of images y and x, respectively.

$\sigma_{x,y}$ is the covariance between y and x.

μ_x, μ_y are the mean values of images x and y, respectively.

The constants c_1 and c_2 are determined based on the dynamic range of the pixel values. The SSIM value equals one if and only if both x and y are identical.

2.3.5. Cross correlation coefficient

In image processing, cross-correlation is a technique used to measure the similarity between two signals or images [30]. It involves sliding one image (or signal) over another and calculating a similarity measure at each position.

2.3.6. Mean square error

One of the most widely used and straightforward full-reference metrics is the mean squared error (MSE), which is calculated by squaring the intensity differences between corresponding pixels in the distorted and reference images [31].

3. Results and discussion

The comparison between real and generated images produced by different deep generative models was conducted using three metrics: SSIM, cross-correlation coefficient, and mean squared error (MSE). These metrics quantify the degree of similarity between the generated images and the original ones.

The performance of the generated images for hypertrophic cardiomyopathy using different deep generative models is summarized in Table 1. The similarity index of the images generated by the VAE is lower than that of all the GAN-based and diffusion models. The generated images produced by the various deep generative models for hypertrophic cardiomyopathy are presented in Figure 4. The first image shows the original MRI scan, followed by the VAE-generated image, the GAN-generated image, the DCGAN- and InfoGAN-generated images, and finally the image generated using the diffusion model.

Table 1. Performance Analysis between Real and Generated Images using different deep generative models (VAE, GAN, DCGAN, ACGAN, and Diffusion Models)

Method	SSIM	Cross- Correlation Coefficient	MSE
VAE	0.903	0.8421	0.39
GAN	0.941	0.8821	0.1
DCGAN	0.958	0.9204	0.1
ACGAN	0.969	0.9371	0.09
Diffusion models	0.972	0.9403	0.06

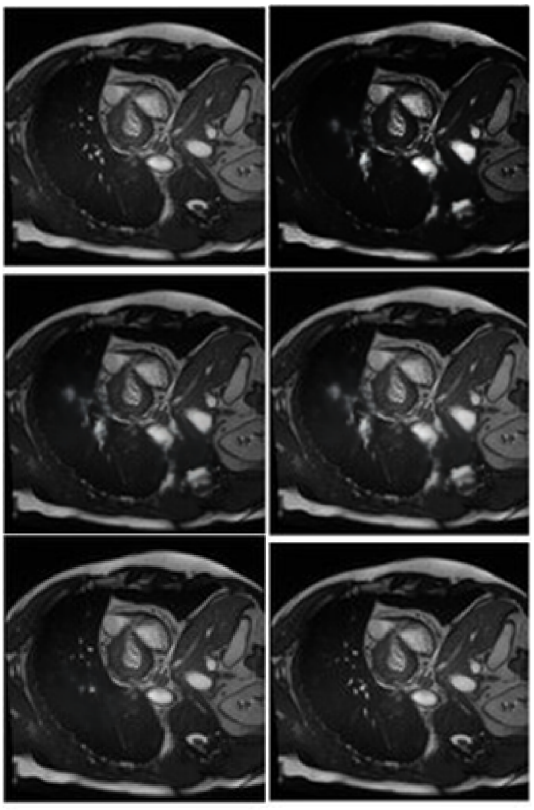

Figure 4. Original Image and Synthesized cardiac images generated using VAE, GAN, DCGAN, ACGAN, and Diffusion model, respectively

In addition to these similarity results, the training loss graphs for the various deep generative models are also presented. The loss function results for the VAE are shown in Figure 5, where it can be observed that the loss decreases as the number of epochs increases.

Figure 6 shows that the generator loss in the GAN increases initially, while the discriminator loss decreases as the number of epochs increases up to approximately 18 epochs. After that point, both losses converge around the 62nd epoch, which yields the best results. The training losses for the DCGAN and diffusion models are shown in Figures 7 and 8, respectively.

When compared to GANs, VAEs perform better during training due to their resilience against mode collapse and their ability to produce more diverse outputs. However, a major drawback is that the generated images are generally blurry and lack sharp detail.

Although diffusion models can produce highly realistic results and maintain stable training, the extensive diffusion process requires a lengthy sampling period, which can limit their practicality for generating images efficiently. This trade-off is evident in the performance metrics presented in Table 1.

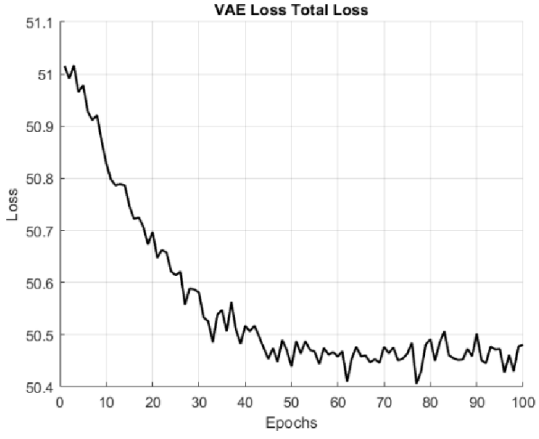

Figure 5. Training Loss of the VAE

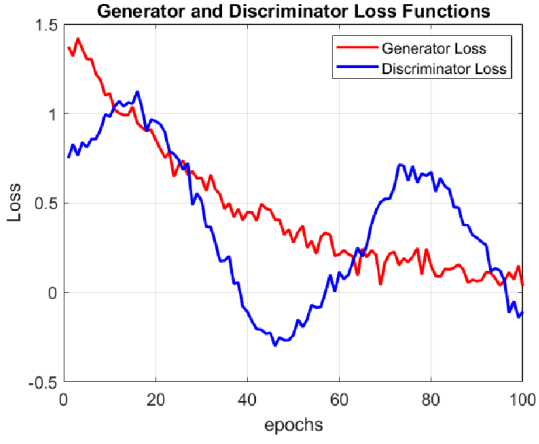

Figure 6. Training Loss of the GAN

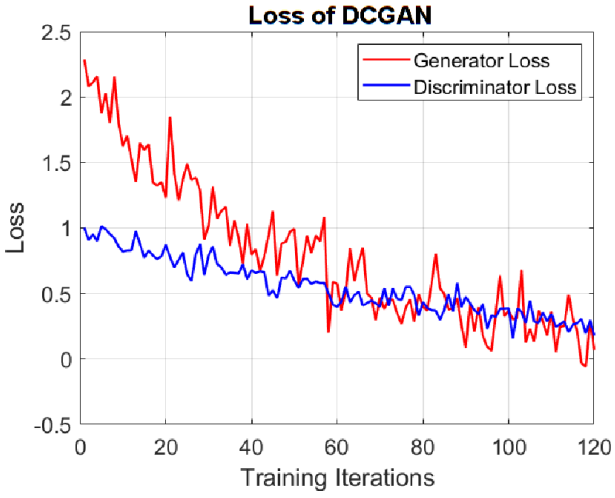

Figure 7. Training Loss of the DCGAN

Figure 8. Training Loss of the Diffusion model

4. Conclusions

Data generation strategies are critical for overcoming the challenge of limited training data in deep learning-based medical image analysis. Unlike conventional data augmentation techniques commonly applied in cardiac diagnostics, deep generative models can synthesize entirely new and diverse data samples. In this study, the performance of various data generation approaches was evaluated using the Structural Similarity Index Measure (SSIM) and the Cross-Correlation Coefficient (CC), both of which are standard metrics for assessing image quality and structural fidelity.

The variational autoencoder (VAE) approach achieved an SSIM of 0.9028 and a CC of 0.8421; however, the generated images exhibited noticeable noise and blurriness, revealing limitations in visual realism. In contrast, generative adversarial networks

(GANs) demonstrated improved performance, achieving an SSIM of 0.9428 and the same CC of 0.8421. Among the GAN variants, both deep convolutional GANs (DCGANs) and auxiliary classifier GANs (ACGANs) produced superior results, with SSIM values of 0.9576 and 0.9687, respectively, indicating a greater capability to generate high-quality and structurally accurate images.

Diffusion models outperformed both GANs and VAEs in terms of similarity metrics, achieving the highest SSIM scores; however, their practicality is constrained by substantially longer sampling times. This trade-off between image quality and computational efficiency should be carefully weighed when selecting an appropriate generative model for medical data augmentation in cardiac diagnostics, particularly for conditions such as hypertrophic cardiomyopathy.

Contributor roles

- **Swarajya Madhuri Rayavarapu:** Conceptualization, methodology, software.
- **Gottapu Sasibhushana Rao:** Investigation, supervision.

References

[1] C. González García, E. Núñez-Valdez, V. García-Díaz, C. Pelayo G-Bustelo, and J. M. Cueva-Lovelle, "A review of artificial intelligence in the internet of things," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, p. 9, 2019. [Online]. Available: http://dx.doi.org/10.9781/ijimai.2018.03.004

[2] Y. Shen, L. Chen, J. Liu, H. Chen, C. Wang, H. Ding, and Q. Zhang, "PADS-Net: GAN-based radiomics using multi-task network of denoising and segmentation for ultrasonic diagnosis of Parkinson disease," *Computerized Medical Imaging and Graphics*, vol. 120, p. 102490, Mar. 2025. [Online]. Available: https://doi.org/10.1016/j.compmedimag.2024.102490

[3] H. Zhang and Y. Qie, "Applying deep learning to medical imaging: A review," *Applied Sciences*, vol. 13, no. 18, p. 10521, Sep. 2023. [Online]. Available: https://doi.org10.3390/app131810521

[4] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26 731–26 769, Dec. 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-14305-w

[5] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, Sep. 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.107187

[6] S. K. Mathivanan, S. Srinivasan, M. S. Koti, V. S. Kushwah, R. B. Joseph, and M. A. Shah, "A secure hybrid deep learning framework for brain tumor detection and classification," *Journal of Big Data*, vol. 12, no. 1, Mar. 2025. [Online]. Available: https://doi.org/10.1186/s40537-025-01117-6

[7] C. Chola, P. Mallikarjuna, A. Y. Muaad, J. V. Bibal Benifa, J. Hanumanthappa, and M. A. Al-antari, "A hybrid deep learning approach for COVID-19 diagnosis via CT and X-ray medical images," in *The 1st International Electronic Conference on Algorithms*, ser. IOCA 2021. MDPI, Sep. 2021, p. 13. [Online]. Available: https://doi.org/10.3390/IOCA2021-10909

[8] F. Y. Shih and H. Patel, "Deep learning classification on optical coherence tomography retina images," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 08, p. 2052002, Oct. 2019. [Online]. Available: https://doi.org/10.1142/S0218001420520023

[9] P. Gupta, S. Nandakumar, M. Gupta, and G. Panda, "Data programming enabled weak supervised labeling for ECG time series," *Biomedical Signal Processing and Control*, vol. 87, p. 105540, Jan. 2024. [Online]. Available: https://doi.org10.1016/j.bspc.2023.105540

[10] S. U. Amin, A. Hussain, B. Kim, and S. Seo, "Deep learning based active learning technique for data annotation and improve the overall performance of classification models," *Expert Systems with Applications*, vol. 228, p. 120391, Oct. 2023. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.120391

[11] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," *Artificial Intelligence in Medicine*, vol. 101, p. 101723, Nov. 2019. [Online]. Available: https://doi.org/10.1016/j.artmed.2019.101723

[12] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthcare Analytics*, vol. 5, p. 100340, Jun. 2024. [Online]. Available: https://doi.org/10.1016/j.health.2024.100340

[13] N. Nonaka and J. Seita, "Data augmentation for electrocardiogram classification with deep neural network," *arXiv*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2009.04398

[14] M. M. Rahman, M. W. Rivolta, F. Badilini, and R. Sassi, "A systematic survey of data augmentation of ECG signals for AI applications," *Sensors*, vol. 23, no. 11, p. 5237, May 2023. [Online]. Available: http://doi.org/10.3390/s23115237

[15] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Expert Systems with Applications*, vol. 161, p. 113696, Dec. 2020. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.113696

[16] J. Saldanha, S. Chakraborty, S. Patil, K. Kotecha, S. Kumar, and A. Nayyar, "Data augmentation using variational autoencoders for improvement of respiratory disease classification," *PLOS ONE*, vol. 17, no. 8, p. e0266467, Aug. 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0266467

[17] D. Bhattacharya, S. Banerjee, S. Bhattacharya, B. Uma Shankar, and S. Mitra, *GAN-Based Novel Approach for Data Augmentation with Improved Disease Classification.* Springer Singapore, Dec. 2019, pp. 229–239. [Online]. Available: https://doi.org/10.1007/978-981-15-1100-4_11

[18] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv*, 2013. [Online]. Available: https://doi.org/10.48550/arXiv.1312.6114

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv*, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1406.2661

[20] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "GANs for medical image synthesis: An empirical study," *Journal of Imaging*, vol. 9, no. 3, p. 69, Mar. 2023. [Online]. Available: https://doi.org10.3390/jimaging9030069

[21] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1511.06434

[22] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," *arXiv*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1610.09585

[23] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets," *arXiv*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1606.03657

[24] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," *arXiv*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1503.03585

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2006.11239

[26] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10 850–10 869, Sep. 2023. [Online]. Available: https://doi.org/10.1109/TPAMI.2023.3261988

[27] Z. Guo, J. Liu, Y. Wang, M. Chen, D. Wang, D. Xu, and J. Cheng, "Diffusion models in bioinformatics: A new wave of deep learning revolution in action," *arXiv*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.10907

[28] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018. [Online]. Available: http://doi.org/10.1109/TMI.2018.2837502

[29] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006. [Online]. Available: http://doi.org/10.1109/TIP.2006.881959

[30] G. Prieto, E. Guibelalde, M. Chevalier, and A. Turrero, "Use of the cross-correlation component of the multiscale structural similarity metric (R* metric) for the evaluation of medical images," *Medical Physics*, vol. 38, no. 8,

pp. 4512–4517, Jul. 2011. [Online]. Available: https://doi.org/10.1118/1.3605634

[31] A. Borji, "Pros and cons of GAN evaluation measures: New developments," *arXiv*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2103.09396