Original Article

Improving Myocardial Infarction Detection from Echo Images using Contrastive Guided Adversarial Denoising Diffusion Probabilistic Model

Belavendhiran Arockia Valanrani[1], Selvakumar Devi Suganya[2]

[1,2]*Department of Computer Science, Vellalar College for Women, Erode, Tamilnadu, India.*

[1]*Corresponding Author : Valansajeev17phd@Gmail.Com*

Abstract - Early detection and diagnosis of Myocardial Infarction (MI) are crucial for preventing cardiac damage or death. Deep Learning (DL) methods effectively diagnose MI, but data scarcity is a primary challenge. Generative Adversarial Networks (GAN) models provide sufficient images by generating quality echo imagses. However, the diversity of the images generated by GAN is limited due to its predominant usage in generating and translating images between different sources. Low diversity leads to degrade the performance of deep learners in MI diagnosis. To solve this, this paper proposes the Contrastive Guided Adversarial Denoising Diffusion Probabilistic Model with GAN (CGADDPM-GAN) model to generate high-quality echocardiography images with high diversity for efficient MI detection. The combination of CGADDPM and GAN assisted in learning the Reverse Denoising Task (RDT) to represent the important anatomical features in produced image samples. In CGADDPM, the Diffusion Probabilistic Model (DPM) is used to generate samples that coincide with data within a limited range for network parameterization. Contrastive Learning Loss (CLL) is integrated with DPM to improve learned representations' quality through a learnable nonlinear transformation. Representation Learning (RL) is introduced with Contrastive Learning (CL) to enhance performance through normalized embeddings and parameter adjustments, resulting in smaller batch sizes. The synthesized images from CGADDPM-GAN are fed into an Encoder-Decoder Convolutional Neural Network (E-D CNN) for segmentation. The features from segmented images are fine-tuned through feature engineering. The fine-tuned features are then utilized in CNN for training and predicting MI. The complete framework is the Deep network model for MI detection (MIDepnet), which provides synthesized echocardiography images with a large diversity and high accuracy in MI detection.

Keywords - Myocardial Infarction, Echocardiography, Generative Adversarial Networks, Contrastive Learning, Representation Learning.

1. Introduction

Myocardial Infarction (MI) is a critical medical condition where blood flow to the heart is abruptly blocked, typically due to a blood clot, leading to potential damage to the heart muscle [1]. A shortage of blood flow to the circulatory system may result in serious cardiac muscle damage and death, as well as chest discomfort, body aches, dizziness, vertigo, perspiration, anxiousness, and difficulty breathing [2]. Rapid detection and treatment are important to reduce the mortality rate from MI. MI is diagnosed using several imaging models like Chest X-Ray (CXR), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Echocardiography (Echo), cardiac magnetic resonance imaging (CMR) and so on [3]. Echo is the most used approach owing to its implicit utilization, inexpensive, rapid efficiency, and reliability [4]. Echocardiography is a non-invasive technique that records heart function during motion using M-Mode imaging, providing clinicians with additional information about blood supply [5]. It captures physiological heart motion into three-dimensional frames using parameters like width, height, and time. The captured echo information can be a singular image during a specific cardiac interval or a video series spanning multiple cycles [6]. However, it derives some limitations like long procedure times, multiple estimation values, complex evaluations, operator subjectivity and broad feedback ranges, even under specific conditions. [7]. The strong desire for clinical expert training in echocardiography exacerbates these restrictions. For this reason, DLs are increasingly utilized to enhance echocardiography clinical examinations for MI prediction [8]. DL systems enhance the echocardiography diagnostic capabilities by quantifying cardio-motion, improving image quality, efficiently classifying cardio-views and computing cardiac motion parameters [9]. Several DL frameworks have been developed to predict various categories of MI using echocardiography data. The EDCNN model was suggested [10] for prompt MI recognition using low-quality

echocardiography. This method segmented whole Left Ventricle (LV) images using EDCNN. Then, feature engineering was applied to the segmented images to identify the potential MI signature from Apical 4-Chamber (A4C) and Apical 2-Chamber (A2C) views. Then, the fine-tuned images were fed into SVM for MI detection. However, the models struggle to synthesize the echo images with ground truth labels, necessitating extensive data and error-prone manual labeling. To solve this, MIDepnet is proposed in this paper using CGADDPM-GAN to solve the above-mentioned issues for efficient MI detection.

This method uses CGADDPM to generate and translate domains from echo images to provide finest image instances with wide variation. The CGADDPM assisted in learning the RDT with decreased sampling tasks, ensuring the retention of important morphological features in produced image samples. In CGADDPM, DPM is a parameterized Markov chain developed with variability prediction to produce the instances that data coincides after a limited segment. The transitions in this chain reverse the diffusion process, progressively increasing the data distortion when the signal is eradicated. Adjustments to dependent Gaussians in the sampling chain can improve network parameterization when propagation is minimal Gaussian noise. CLL and DPM are applied to synthesize data and enhance learnt depictions' accuracy through nonlinear adjustment. RL with CL improves performance through normalized embeddings and parameter adjustments, resulting in smaller batch sizes and shorter training time. The data obtained from the CGADDPM is integrated with the GAN model to synthesize new echocardiography images with large diversity and eliminate the image inadequacy problem. The obtained images are fed into E-D CNN for segmentation.

Then, feature engineering is applied to optimize the segmented LV wall to determine the MI signature [12]. Finally, the CNN model is employed for the MI prediction and classification. The following parts are structured: Section II examines related works, Section III discusses the recommended strategy, Section IV demonstrates its effectiveness, and Section V provides a summary and future improvements.

2. Literature Survey

A computerized MI diagnosis system was constructed [11] using the echocardiography and Polar Residual Network (PResNet). However, the accuracy and F1-score were less due to its inability to address low generalization ability and variations in infraction severity. A Convolutional Neural Sub-network (CNSN) was devised [12] by computing normalized cross-correlations from 2DE images between feature maps for motion estimation. But, accuracy and precision values were lower as the parameters were not optimized properly. A 3D CNN system was presented [13] to define the cardiac activities

and the motion from the echo images for MI detection. But, this model results in high computational time and memory space. A Trilateral Attention Network (TaNet) was created [14] using a Spatial Transformer Network for echo extraction, region cardiac segmentation and cardiac parameters evaluation in echocardiography images for MI prediction. However, increasing the data increases the misclassification error and lowers the accuracy rate. A 3D Cardiac Segmentation Network (3D-CSN) was constructed [15] using multi-labelled echo images to segment the median-line location for MI detection. However, this model required extensive synthetic data for training, and slight overfitting issues were experienced, resulting in lower accuracy and recall efficiency. A 3D-CNN model was developed [16] to automatically detect Regional Wall Motion Abnormalities (RWMAs) in MI patients using Xception and LSTM-U-Net for MI detection. However, this model obtains relatively lower accuracy as it determines highly sensitive borders near the edge area and lowers the significant information. An ensemble learning method was devised [17] to detect MI in echocardiographs using U-Net for LV image segmentation and a Support Vector Machine (SVM) for MI detection.

However, this model was trained with a limited dataset, which reduced the accuracy and F1 score. An IoT-based technique was suggested [18] with a Fuzzy C-Means (FCM) algorithm for segmentation and Pre-Trained Recurrent CNN (PRCNN) for echocardiography categorization and MI prediction. However, the model parameters were not fine-tuned appropriately, resulting in uncertainty issues. An early diagnosis of MI was demonstrated [19] using 1D-CNN to calculate and analyze myocardial segments with LV wall motion and displacement modification.

However, accuracy and recall rate were lower as the model's network parameters were not fine-tuned for real-time diagnosis. Due to privacy concerns and the limited availability of labeled medical imaging datasets, various pipelines have been developed to generate synthetic 3D echocardiographic images with corresponding ground truth labels. These methods aim to reduce the need for extensive data collection and manual annotation. While GANs are typically used to create high-resolution anatomical models from echocardiograms, their image variety is limited by single-source data. To overcome this, the paper introduces a novel DPM integrated into the GAN framework, enhancing image quality and diversity for improved MI detection accuracy. This approach helps build more reliable DL models for clinical use, improving MI diagnosis.

3. Proposed Methodology

This part illustrates the framework of the MIDepnet model designed for efficient MI identification. Figure 1 illustrates the block diagram of the proposed study. Table 1 depicts the notation list.
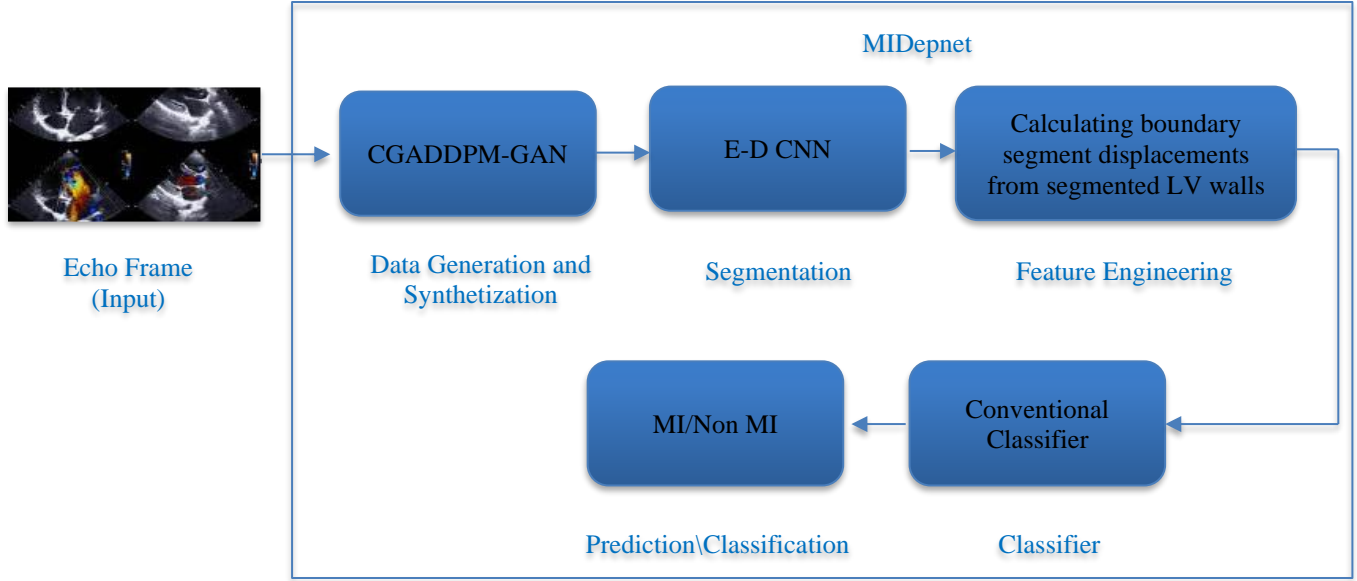
Fig. 1 Pipeline of the MIDepnet model

Table 1. Notation List

Notations	Description
$\mathcal{A}$	Feature Space
$\mathcal{B}$	Label Space
ϵ	Neighbourhood of data
θ	Learnable Parameter
$\mathcal{L}_{Adv}$	Adversarial Loss
$1\{.\}$	Indicator Loss
a , b	Data and label
$d_{\mathcal{A}}$	Generated Features
δ, Δ	Adversarial set
f	Forward Data Dispersion
r	Reverse Data Dispersion
Φ	gradient of the data density
c	Gaussian changes
$\mathbb{U}$	Uniformity Covariance Matrix
β_c	Injected noise
Q	Estimated Residual
G	Constant Parameter
μ_c	Mean
$\hat{\mu}_c$	Posterior Mean
σ	Variance
γ	Forward process variances
$\mathbb{V}$	KL Divergence
$\mathcal{G}$	Generator
$\mathcal{D}$	Discriminator
l_{cont}	Contrastive Loss
τ	Softmax Temperature
ϱ	Hyperparameter stabilizing the l_{cont}
$e (.)$	Neural Network Base Encoder
$j(.)$	Projection Head
H	Data Synthetic Depiction
W	Variation reduction

ψ	Weight control parameter
$l_{u,v}$	Loss function for instances pair (u, v)

3.1. Problem Definition and Preliminaries

Assume the feature space $\mathcal{A}$ with corresponding label space $\mathcal{B}$ and the true joint data distribution as $d = d_{\mathcal{A}*\mathcal{B}}$. Consider a labelled training data $d_{train} := \{(a_x, b_y)\}_{x=1}^{n}$ which aims to learn a robust classifier $F_{\theta}: \mathcal{A} \to \mathcal{B}$, parameterized by a learnable θ to achieve lower adversarial loss as in Equation 1,

$$\min_{\theta} \mathcal{L}_{Adv}(\theta) := \mathcal{E}_{(a,b)\sim d} \left(\max_{\delta \in \Delta} l(a + \delta, b , \theta) \right) \quad (1)$$

Where, $l(a, b , \theta) = 1\{b \neq f_{\theta}(a)\}$ is the loss operation, $1\{.\}$ is the indicator function and $\Delta = \{\delta: \|\delta\|_{\infty} \leq \epsilon\}$ is the adversarial items defined by $l_{\infty} - norm$. Eq. (1) minimizes the worst-case loss within a $\epsilon -$ neighborhood of input characteristics. In the mathematical terms of adversarial training, the robust classifier F_{θ} is trained on the training set $d_{train} := \{(a_x, b_y)\}_{x=1}^{n}$ by addressing the sample mean approximation in Equation 2,

$$\min_{\theta} \hat{\mathcal{L}}_{Adv}(\theta) := \frac{1}{n} \sum_{x=1}^{n} \max_{\delta \in \Delta} l(a_x + \delta_x, b_x, \theta) \quad (2)$$

3.2. Contrastive Guided Adversarial Denoising Diffusion Probabilistic Model
3.2.1. Adversarial Training

Synthetic data generation enhances learning algorithm performance by increasing training set dimensions and generating sufficient data. Mainstream generation procedures include feature generation and conditional generation based on desired features. Assume the dispersion of the generated features as $d_{\mathcal{A}}$ and the generated synthetic data as $d_{syn} :=$

$\{(\hat{a}_x, \hat{b}_y)\}_{x=1}^{\hat{N}}$. The features values $\hat{a}_x$ are determined by synthetic distribution $\hat{d}_A$ and $\hat{b}_x$ are the pseudo-classes assigned by the classifier trained on the training data d_{train}.

By integrating the synthetic and original data, the training set $d_L \coloneqq d_{train} \cup d_{syn}$ with $N + \hat{N}$ samples are provided in Equation 3,

$$\min_\theta \left\{ \psi \left(\frac{1}{n} \sum_{x=1}^N \max_{\delta \in \Delta} \ell \left(a_x + \delta_x + b_x, \theta \right) \right) + (1 - \right.$$
$$\left. \psi) \left(\frac{1}{\hat{n}} \sum_{x=1}^N \max_{\gamma_x \epsilon \Delta} \ell \left(\hat{a}_x + \delta_x + \hat{b}_x, \theta \right) \right) \right\} \qquad (3)$$

Where, $\psi \in (0,1)$ is the variable managing the weights of synthetic data.

3.2.2. Denoising Diffusion Probabilistic Model
The notion of DDPM is a forward Markov chain with Gaussian fluctuations $f(a_c|a_{c-1})$ to drive the noise to the actual data dispersion $f(a_0)$ Particularly, the forward Gaussian transition (c) is defined as follows,

$$f(a_c|a_{c-1}) \coloneqq \mathcal{Z}(\sqrt{\beta_c}\, a_{c-1}, (1 - \beta_c)\mathbb{U}) \qquad (4)$$

In Equation 4, $\beta_c, c = 1, 2, \dots\dots C$ is a subsiding series to manage the derived noises divergence, and $\mathbb{U}$ is the uniformity covariance matrix. The connective probability for the above Markov chain can be depicted as $f(a_0:C) = f(a_0) \Pi_{C=1}^c f(a_c|a_{c-1})$. In DDPM, it is assumed that $f_\theta(a_{0:c}) = r_\theta(a_c)\Pi_{c=1}^C r_\theta(a_{c-1}|a_c)$ for the reverse task, where $r_\theta(a_{c-1}| a_c)$ is a parameterized using a neural network. The training target is to reduce the Kullback-Leibler (KL) divergence between the forward and reverse tasks, $d_{KL}(f(a_{0:C}), f_\theta(a_{0:C}))$ as in Eq. (5),

$$\min_\theta \mathbb{V}_{c,a_0,\epsilon} \left[\left\| \epsilon - \epsilon_0 \left(\sqrt{\hat{\beta}_c}\, a_0 + \sqrt{1 - \hat{\beta}_c}\epsilon, c \right) \right\|^2 \right] \quad (5)$$

In Equation 5, $a_0 \sim f(a_0), \hat{\beta}_c = \prod_{R=1}^c \beta_r$ for $c = 1,2,\dots,C, \epsilon \sim \mathcal{Z}(0, \mathbb{U})$ and $\epsilon_\theta(a,c)$ represents the neural network parameterized by θ to be trained simultaneously. The original process of a time-reversed Markov chain is generated after learning the time-reversed task parameterized by θ, as illustrated in Equation 6.

$$a_{c-1} = \frac{1}{\sqrt{\beta_c}} \left(\beta_c - \frac{1 - \beta_c}{\sqrt{1 - \hat{\beta}_c}} \epsilon_\theta(a_c, c) \right) + \sigma_c \Phi_c, \quad c = C,$$
$$C - 1, \dots, 1 \qquad (6)$$

Where, $\Phi_c \sim \mathcal{Z}(0,1)$ if $c > 1$ and $\sigma_c = 0$ if $c = 1$. The Denoising Diffusion Implicit Model (DDIM) accelerates the DDPM task by generalizing it to a non-markovian procedure, resulting in a sampling orientation much smaller than C. DDIM is designed to prove the forward transformation $F(a_{c-1}|a_c, a_0)$ such that $f(a_c, a_0) = \mathcal{Z}(\sqrt{\beta_c}\, a_0), (1 - \beta_c)\mathbb{U}$ for all $c = 1, \dots, C$. The DDIM shares similarities with DDPM, which adapts the pre-trained DDPM model, facilitating the sampling task without incurring additional costs. The main sampling generating phase in DDIM is depicted in Equation 7,

$$a_{c-1} = \sqrt{\beta_{c-1}} \begin{pmatrix} \frac{a_c - \sqrt{1 - \beta_c}\epsilon_\theta\,(a_c,c)}{\sqrt{\beta_c}} \\ \text{detected } a_0 \end{pmatrix} + \begin{pmatrix} \frac{a_c - \sqrt{1 - \beta_{c-1}}\epsilon_\theta\,(a_c,c)}{\sqrt{\beta_c}} \\ \text{pointing to } a_c \end{pmatrix} (7)$$

In Equation 7, a_{c-1} is the process of generation becomes predictable and consistent and determined by a_c and a_0. The variance β_c is determined by forward decisions, model design, and Gaussian distribution parameterization in the RDT.

Moreover, the training process may be enhanced by fine-tuning the random terms of W using a stochastic gradient descent. The preceding enhancements might help to eliminate the variation reduction by rewriting W in Equation 8 as,

$$\mathbb{V}_f \left[\frac{d_{kl}\left(f(a_C|a_0)||r(a_c)\right)}{W_C} + \right.$$
$$\frac{\sum_{c>1} d_{kl}(f(a_{c-1}|a_c, a_0)||r_\theta(a_{c-1}|a_c, a_0))}{W_{c-1}} -$$
$$\left. \frac{\log r_\theta(a_0|a_1)}{W_0} \right] \coloneqq W \qquad (8)$$

DPPM constructs a new relationship between diffusion algorithms and noise reduction score balancing, resulting in a simpler and weighted variational limit target. Forward Task and W_c: The forward process deviations β_c are adjusted to variables, ignoring the re-parameterization.

Because the estimated residual r has no learnable parameters, W_c remains static throughout training and may be deleted. The forward task is represented in Equation 9,

$$W_c = d_{kl}\left(f(a_C|\, a_0)||\mathcal{Z}(0,1) \approx 10^{-5}\right) \qquad (9)$$

Reverse Task and $W_{1:c-1}$: The selection in $r_\theta\,(a_{c-1}|a_c) = \mathcal{Z}\,(a_{c-1}; \mu_\theta(a_c, c), \sum_\theta(a_c, c))$ for $1 < c \le C$. Initially, $\sum_\theta(a_c, c) = \sigma_c^2 \mathbb{U}$ is adjusted to train the latency related constants. Mathematically, both $\sigma_c^2 = \gamma_c$ and $\sigma_c^2 = \hat{\gamma}_c = \frac{1 - \hat{\beta}_{c-1}}{1 - \hat{\beta}_c} \gamma_c$ are acquired with similar performances.

The initial selection is relevant for $a_0 \sim \mathcal{Z}(0, \mathbb{U})$ and the second is also intended to be relevant for a_0 when it is deterministically adjusted to the main point. Then, a detailed parameter configuration organized by the preceding evaluation as W_c to define the mean $\mu_\theta(a_c, c)$ is given in Equation 10,

$$W_{c-1} = \mathbb{V}_f \left[\frac{1}{2\sigma_c^2} \|\hat{\mu}_c(a_c, a_0) - \mu_\theta(a_c, a_0)\|^2 \right] + G \quad (10)$$

Where G is the constant that is independent of θ. The straight parameterization μ_θ detects $\hat{\mu}_c$ which is the forward process posterior mean. Also, the re-parameterization in Equation 4 is further expanded by Equation 10 as Equation 11

$$a_c(a_0, \epsilon) = \sqrt{\widehat{\beta_c}}\, a_0 + \sqrt{1 - \widehat{\beta_c}}\, \epsilon \text{ for } \epsilon \sim \psi(0, \mathbb{U}) \quad (11)$$

The forward process posterior is applied in Equation 12

$$\hat{\mu}_c(a_c, a_0) := \frac{\sqrt{\hat{\beta}_{c-1}}\,\gamma_c}{1 - \hat{\beta}_c} a_0 + \frac{\sqrt{\hat{\beta}_c}\,(1 - \hat{\beta}_{c-1})}{1 - \hat{\beta}_c} a_c \text{ and } \hat{\gamma}_c :=$$
$$\frac{1 - \hat{\beta}_{c-1}}{1 - \hat{\beta}_c} \gamma_c \quad (12)$$

Applying the Equation 11 in Equation 10, the Equation 13 is obtained,

$$W_{c-1} - G = \mathbb{V}_{a_0, \epsilon} \left[\frac{1}{2\sigma_c^2} \left\| || \hat{\mu}_c(a_c\,(a_0, \epsilon), \frac{1}{\sqrt{\hat{\beta}_c}} \left(a_c(a_0, \epsilon) - \right. \right. \right.$$
$$\left. \left. \left. \sqrt{1 - \hat{\beta}_c}\epsilon \right), \sigma_\theta(a_c(a_0, \epsilon)), c \right\|^2 \right] \quad (13)$$

The Equation 14 depicts that the σ_θ which should identify $\frac{1}{\sqrt{\hat{\beta}_c}} \left(a_c - \frac{\gamma_c}{\sqrt{1 - \hat{\beta}_c}} \epsilon \right)$ is provided at a_c. Since, a_c

will be fed as input to the model to select the parameterization in Equation 15,

$$W_{c-1} - G = \mathbb{V}_{a_0, \epsilon} \left[\frac{1}{2\sigma_c^2} \left\| \frac{1}{\sqrt{\hat{\beta}_c}} \left(a_c\,(a_0, \epsilon) - \frac{\gamma_c}{\sqrt{1 - \hat{\beta}_c}} \epsilon \right) - \right. \right.$$
$$\left. \left. \mu_\theta\left(a_c(a_0, \epsilon) \right), c \right\|^2 \right] \quad (14)$$

$$\mu_\theta(a_c, c) = \hat{\mu}_c \left(a_c, \frac{1}{\sqrt{\hat{\beta}_c}} \left(a_c - \sqrt{1 - \hat{\delta}_c}\, \epsilon_\theta\,(a_c) \right) \right) =$$
$$\frac{1}{\sqrt{\hat{\beta}_c}} \left(a_c - \frac{\gamma_c}{\sqrt{1 - \hat{\beta}_c}} \epsilon_\theta\,(a_c, c) \right) \quad (15)$$

Where, ϵ_θ will be a function estimator projected to predict ϵ using a_c. The variational analysis is used to enhance denoising score coordination by estimating the finite-time residual instances with Langevin dynamics μ_0 and upgrading the parameterization of the reverse process mean operation approximation $\hat{\mu}_c$ or ϵ. $\epsilon-$ prediction parameterization reduces the propagation variational limit to the desired level, similar to reducing noises for score matching. However, it is essentially another parameterization

of $R_\theta(a_{c-1}|a_c)$, therefore, its usefulness is assessed by expecting ϵ to predict the $\hat{\mu}_c$. The GAN's efficiency is dependent on calculating the denoising distribution R_θ. By presenting the original image Y, the reverse diffusion procedure is triggered. The generator $\mathcal{G}$ synthesizes $a'_{c-1} \sim r_\theta(a_{c-1}|a_c, Y)$ to estimate $r_\theta(a_{c-1}|a_c, Y)$ by generating a'_{c-1}. The discriminator $\mathcal{D}\,(a'_{c-1}, a_c, C))$ is modified among the instances for either the actual ratio dispersion $f(a)$ or the identified $R_\theta(a)$. Images are generated and translated from one input to another using the gathered datasets.

3.2.3. Contrastive Learning Loss and Representation

The CGADDPM is proposed in this model to enhance the sample efficiency of the generative model. The contrastive loss (l_{cont}) of this model is completely guided by the DDPM. In every time step ts of the generation procedure, the given present value $a_c^{(x)}$ is added to the gradient of the $l_{cont}\,(a_c^{(x)}, a_{c-1}^{(x)}; \tau)$ with accordance to $a_c^{(x)}$ as the actual diffusion generative tasks, $a_{c-1}^{(x)}$ is the positive couple of $a_c^{(x)}$, τ and ϱ is the softmax temperature and hyperparameter stabilizing the l_{cont} within the diffusion procedure. The $l_{cont}(.)$ is very resilient in which multiple terms of l_{cont} will be coupled with various selection processes of positive and negative pairs. The CL effectively generates synthetic data, requiring multiple data augmentation for effective distinct predictions, benefiting from weaker data augmentation than unstructured distinctive learning. In CGADDPM, a learnable nonlinear transformation connecting depiction and CLL greatly improves visualization, facilitating larger batch sizes and longer training durations for mutual information forecasting. Figure 2 depicts the framework of CGADDPM-GAN.

In this context, RL utilizes contrastive cross-entropy loss that involves normalized embeddings and precise variables. The normalized embeddings are vectors scaled to unit length, which is essential for maintaining consistent distances in high-dimensional space. This allows the model to distinguish between similar and dissimilar samples effectively. Normalization can be processed by L2 Normalization and mean normalization. Normalized embeddings improve the model's capacity to distinguish similar and dissimilar pairings using cosine similarity comparisons. This technique speeds training convergence by modifying gradient scaling and maintaining consistent parameter updates while limiting the influence of outliers, resulting in improved generalization and prediction accuracy. To create two integrated views, denoted as $\hat{a}_u$ and $\hat{a}_v$, the stochastic data augmentation model transforms data samples into positive pairs. This approach sequentially implements three types of augmentations within CLL: arbitrary cropping followed by resizing to the original dimensions, random color deformation, and random Gaussian blur.
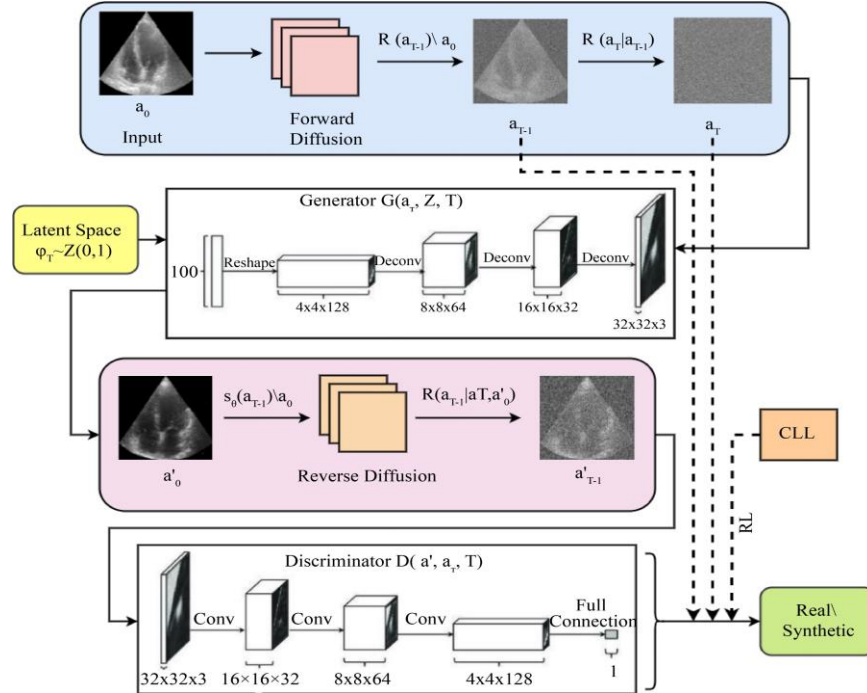
Fig. 2 Structure of CGADDPM-GAN for data generation and synthetization

Effectively combining arbitrary cropping and color deformation is critical for successful data development. The neural network base encoder $e(.)$ retrieves the representation vectors from augmented data samples. A $e(.)$ and a projection head $j(.)$ is performed to increase covenant using CLL in CGADDPM. After training is completed, the $j(.)$ is eliminated and perform encoder $e(.)$ and depiction H for data synthetization. The framework allows for the selection of different network structures without any limitations. The ResNet model is commonly adopted to determine the $H_u = e(\hat{a}_u) = ResNet\ (a_u)$ in which $H_u \in \mathbb{R}^D$ represents the result while performing the mean pooling layer. A minor neural network $j(.)$ that maps the visualization to the space in which the CLL is performed.

A Multi-Layer Perceptron (MLP) with a single hidden layer to determine the $R_u = j(H_u) = w^{(2)}\sigma(w^{(1)}H_u)$ where μ is a ReLU non-linearity. The contrastive loss on R_u is illustrated for the distinctive prediction operation rather han H_u which is more valuable. In a given set of $\{\hat{a}_g\}$, affirmative instances pairs $\hat{a}_u$ with $\hat{a}_v$ is insisted for the prediction of contrastive process, which aims to detect $\hat{a}_v$ in $\{a_g\}_{g \neq u}$ for a particular $\hat{a}_u$. In random mini-batch with samples n, inconsistent detection task is determined on pairs of augmented data, resulting in 2n data points. As per the positive pairs, the next $2(n-1)$ are attended to augment the samples within a minibatch as negative samples. Assume similarity $simi(x,y) = x^\top y/\|x\|\|y\|$ which is specified as a dot product among the l_2 normalized for x,y using the cosine resemblances. The loss function for the positive pair of instances (u,v) is formulated as shown in Equation 16,

$$l_{u,v} = -\log\frac{\exp(simi\ (a_u,a_v')/\tau)}{\sum_{g=1}^{2n} 1_{[g \neq u]}.\exp(simi\ (a_u,a_v')/\tau)} \quad (16)$$

Where, a_u, a_v' indicates the denoised representation obtained from the CGADDPM to capture the essential input features after reducing noises for instances (u,v). $1_{[g \neq u]} \in \{0,1\}$ is an indicator operation measured to 1 iff $g \neq u$. The fitness loss is computed for all affirmative pairs combined with (u,v) and (v,u) in a smaller batch.

The overall structure of the contrastive loss remains intact, focusing on maximizing the similarity between positive pairs while minimizing it for negative pairs. Incorporating the principles of DDPM into the contrastive learning framework enhances the sample efficiency by leveraging the generative capabilities of DDPM for better representation learning.

3.3. GAN
The data determined from the CGADDPM is integrated with GAN to improve segmentation, feature extraction, and classification. In addition, it generates large echo-cardio image samples and, balances the dataset for training with a large diversity and eliminates the echo image insufficiency problem.

GAN utilizes a generator and discriminator to minimize the difference between real and synthetically generated images. When combining CGADDPM with GAN, it learns the probability distribution function of a training dataset is approximated by eliminating the input image with Gaussian noise and denoising the initial noisy input image, resulting in high-fidelity synthetic samples.
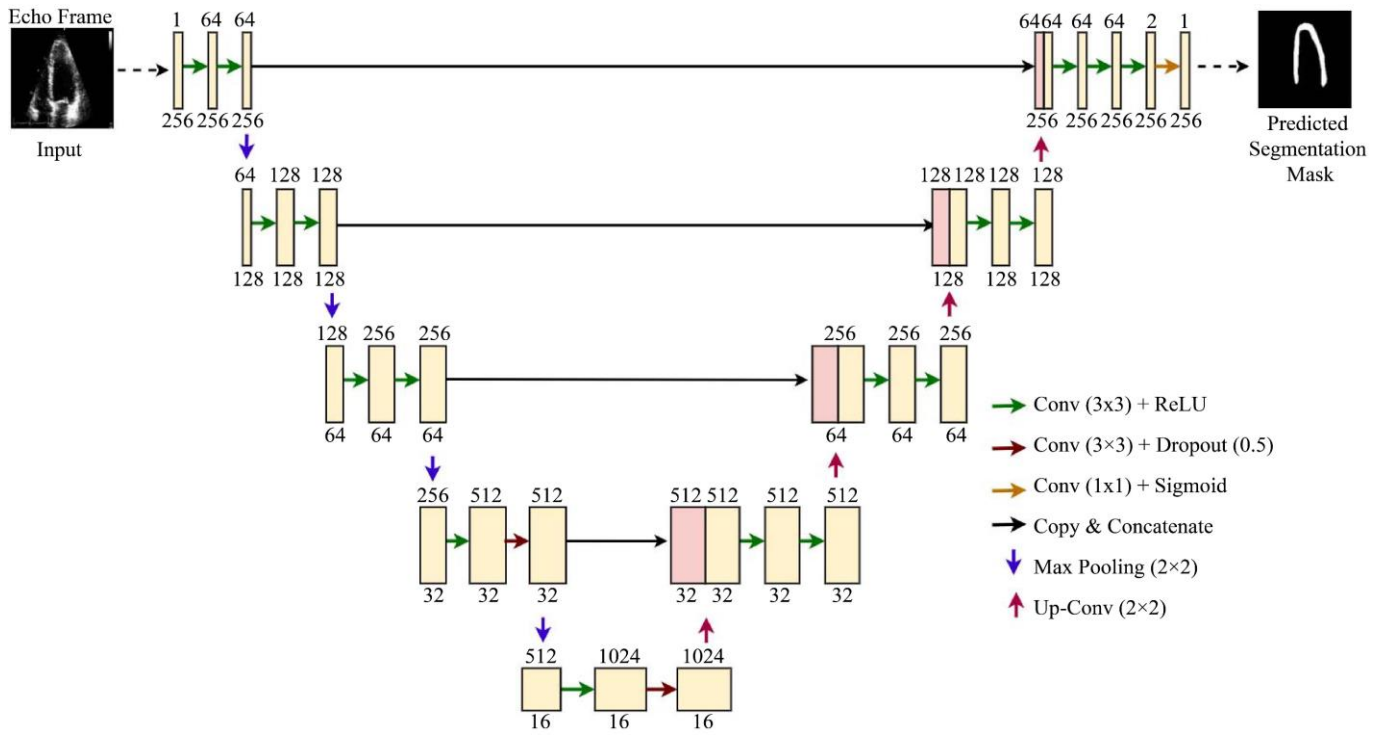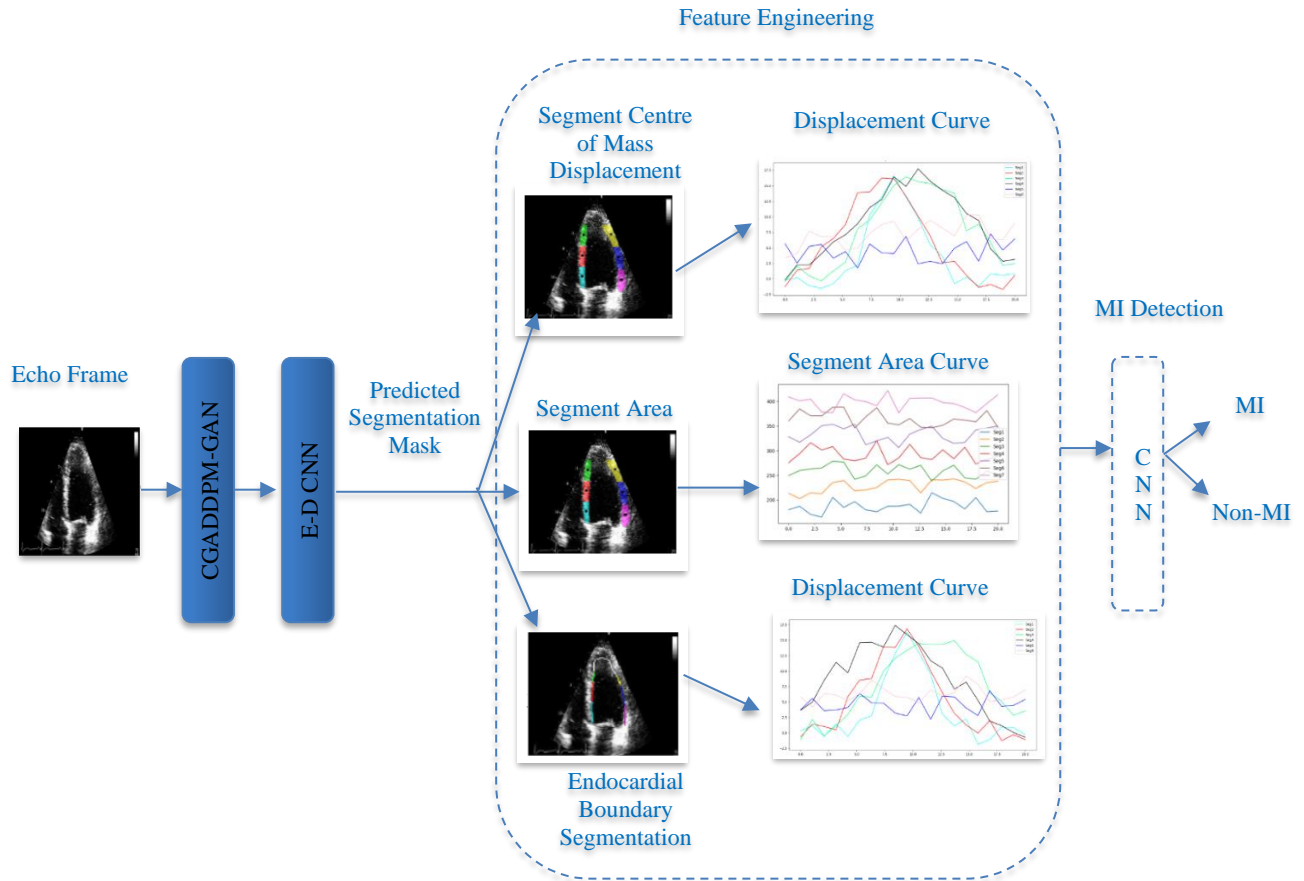
Fig. 3   Structure of E-D CNN

Fig. 4 Schematic Representation of Proposed MI Detection Model

3.4. EDCNN

In order to segment the full LV wall of each frame of each echo frame, the generated images from the CGADDPM-GAN are input into the EDCNN model. The EDCNN used in this model, which has four layers for downsampling and four levels for upsampling, was based on the U-Net architecture [20]. The number of feature maps and picture dimensions corresponding to each layer is labeled above and below the relevant layers. The EDCNN model's structure is shown in Figure 3. Initially, the ground-truth masks for LV wall segmentation are annotated using the pseudo-labelling approach. Experts then confirm the generated masks for every echo frame to ensure the right label frames are obtained. Following the creation of the ground-truth, the retrieved LV wall masks are segmented using EDCNN, which additionally establishes the intersections and displacements between the LV wall segments and the Endocardial Boundary (EB).

3.5. Feature Engineering

The potential MI signals are found by analyzing the segmented LV wall pictures. To define the EB in this work, the inner boundary of the segmented LV wall will be removed. After that, the boundary is split into six parts, and the echo determines each segment's displacement. The six segments of each echo are used to derive features such as color-coded LV wall with EB segments, segment (center) with EB displacement curves, segment area curves, and maximum displacement snapshot of the EB. These features help cardiologists predict MI more accurately and objectively [10].

3.6. MI Detection

The fine-tuned LV wall segments are fed into the CNN model for the MI prediction [21]. CNN assists clinicians in detecting MI at its early stages by analyzing subtle changes in echo images that may not be immediately apparent to the human eye. Early diagnosis allows for prompt intervention and management, potentially preventing further damage to the heart. The below presented algorithm depicts the proposed MIDepnet framework. Figure 4 depicts the schematic representation of the proposed model.

Algorithm: Proposed MIDepnet Model

1 **Input:** Dataset with the collection of 2D echocardiography recordings
2 **Output:** MI prediction and classification
3 Sending the input data a_0
4 Initialize $a_c = \left\{ a_c^{(x)} \right\}_{x=1}^{M} \sim \varphi_c \sim \mathcal{Z}(0,1)$
5 For $C = c$
6 While $c \neq 1$ do
7 For $x = 1: M$ do
8 Selecting $a_p^{(x)}$ as the positive pair of $a_c^{(x)}$
9 Transfer $a_c \sim f(a_c | a_{c+1}, a_0)$ using $R_\theta(a_c | a_{c+1})$
10 End for
11 Send a_0 employing the $R_\theta(a_0 | a_1)$
12 Compute $\quad \Delta a_c^{(x)} = \varrho. \nabla_{a_c^{(x)}} l_{cont}\left(a_c^{(x)}, a_p^{(x)}; \tau \right) + \epsilon_0(a_c^{(x)}, c)$
13 Acquire a_c using $R(a_c)$
14 For $C = c$
15 Obtain a_c by $R_\theta(a_c | a_{c-1})$
16 End for
17 End While
18 End for
19 Return $a_0 = \left\{ a_0^{(x)} \right\}_{x=1}^{M}$
20 For all sampled $\left\{ a_g \right\}_{g=1}^{n}$ do; $g \in \{1, ..., n\}$
21 Determine $u \in \{1,2,...,n\}$ and $v \in \{1,2,...,n\}$ do
22 Construct $simi_{u.v} = R_u R_v / (\|R_u\| \|R_v\|)$ \\ pairwise similarity
23 Specify $l_{u,v}$ (as in Eq. (15)); $l(CL) = \frac{1}{2n} \sum_{g=1}^{n} [l(2g - 1, 2g) + l(2g, 2g - 1)]$
24 Upgrade the network F and P to reduce the l_{cont} operation
25 End for
26 Reinitiate $f(.)$
27 Obtain the synthesized data
28 Input the synthesized data to EDCNN to detect LV segmentation mask
29 Perform feature engineering to acquire endocardial boundary
30 Employ the CNN model to classify MI and Non-MI

4. Results and Discussion

4.1. Dataset Description

HMC-QU Dataset [22]: This dataset has a partition of left ventricular walls. It contains over 10,000 echocardiography samples from 2018 and 2019, containing over 800 instances of acute ST-elevation. A subset of 109 A4C view echocardiography recordings has ground-truth differentiation masks at each pixel during a heartbeat covering the whole left ventricle. In order to experiment, 2346 frames were taken into account. Realistic Synthetic 2D Ultrasound Dataset [23]: It is a publicly available dataset with a collection of simulated echocardiography images comprising 105 sequences or 6165 frames with A2C, A3C and A4C views. It also includes 6060 pairs of synthetic ultrasound images with myocardial displacement fields. The dataset comprises an electromechanical heart model and template cine loop recordings incorporating 2D apical chamber views from five distinct motion patterns. For the experiment, 2289 frames have been considered.

4.2. Experimental Setup and Performance Evaluation

This section initially compares the quality of generated images by CGADDPM-GAN with other GAN models like DDM-GAN [24] and Enhanced cycle-GAN [25] using the Mean Square Error (MSE) metric that is represented in Equation 17. Next, the MI detection efficiency of the MIDepnet model is compared with existing models like

CNSN [12], 3D-CNN [13], 3D-CSN [15], 1D-CNN [19] and EDCNN-SVM [10] using the metrics mentioned from Equation 18 to Equation 21. Using the gathered dataset shown in section 4.1, both the current and suggested models were implemented in Python 3.11 with an Intel® CoreTM.i5-4210 CPU @ 3GHz, 4GB RAM, and 1TB HDD running Windows 10 64-bit. Table 2 provides the parameter details of existing and proposed models for performance evaluation. The gathered datasets are distributed into 70% (training) and 30% (testing). The following performance indicators are used to assess the model's triumph in estimating MI from the gathered datasets.

MSE: It represents the deviation between real a and generated a' images of GAN models

$$MSE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(a_{ij} - a'_{ij})^2}{NM} \quad (17)$$

Accuracy: It is calculated using Equation 17 as the ratio of correctly predicted instances (MI patients) to total occurrences (MI and non-MI patients), as in Equation 18,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

A "True Positive" (TP) situation occurs when the model correctly predicts an MI patient. The state is denoted as True Negative (TN) when the model correctly recognizes Non-MI patients. False Positive (FP) refers to situations where the model mistakenly identified non-MI participants as MI. The quantity of incorrectly forecasted MI as non-MI participants is known as False Negative (FN).

Precision: It determines the ratio of TP projections (properly predicted MI patients) among all optimistic

predictions generated by the model, as in Equation 19,

$$Precision = \frac{TP}{TP+FP} \quad (19)$$

Recall: It calculates the proportion of TP projections among all observed affirmative cases in the dataset as in Equation 20,

$$Recall = \frac{TP}{TP+FN} \quad (20)$$

F1-score: It devises the tradeoff ratio of precision and recall as in Equation 21,

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (21)$$

Figure 5 depicts the performance of proposed and existing models for efficient data generation. The proposed CGADDPM-GAN is compared with existing models like DDM-GAN [24] and Enhanced cycle-GAN [25]. It is determined that the proposed model is 2.9 % improved than DDM-GAN and 7.9 % improved than cycle-GAN for the HMC-QU Dataset in terms of MSE. For the synthetic dataset, 4.2% and 9.3% improved than DDM-GAN and Enhanced cycle-GAN, respectively. This is due to CGADDPM, which generates new samples from learned distributions, augmenting the training set, especially in situations with limited labeled data. CGADDPM and CLL leverage large volumes of unlabeled data and generate high-quality synthetic data with robust features, resulting in improved performance on MI prediction. The effectiveness of several models on the combined HMC-QU dataset for MI identification is shown in Figure 6.
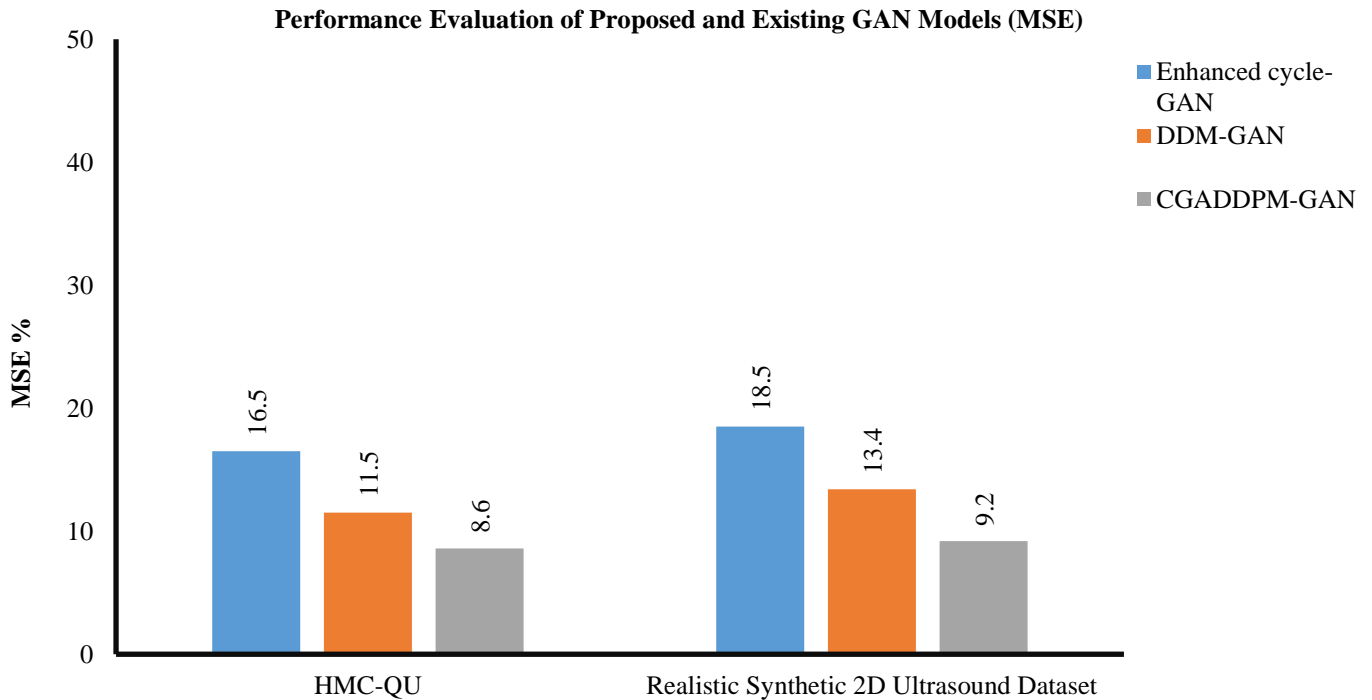

Performance Evaluation of Proposed and Existing GAN Models (MSE)

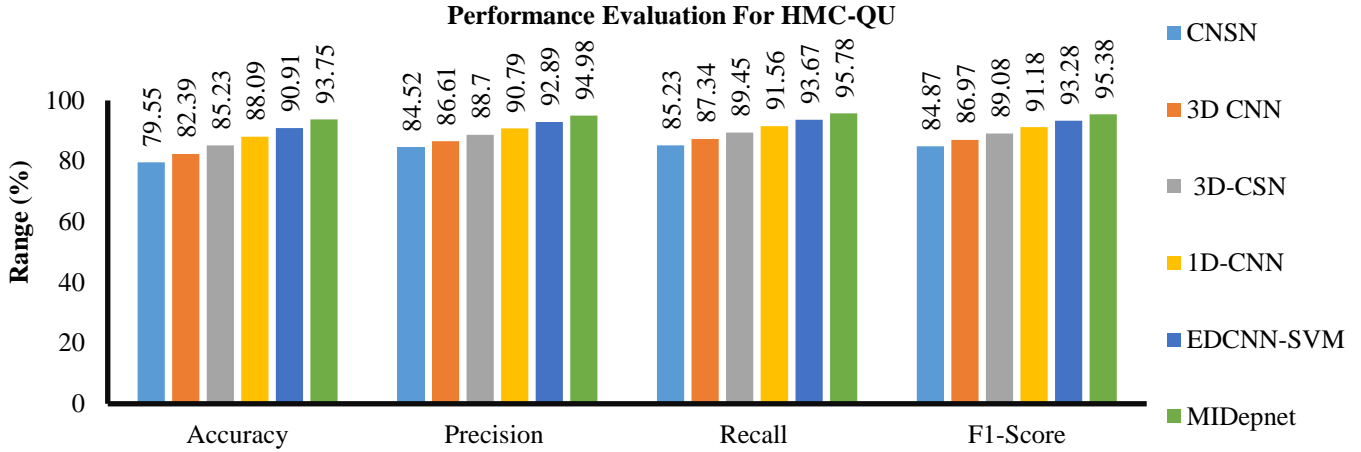**Fig. 5 Performance analysis of proposed CGADDPM-GAN and existing GAN models**

Fig. 6 Comparison analysis of the proposed and existing model for MI detection for HMC-QU Dataset
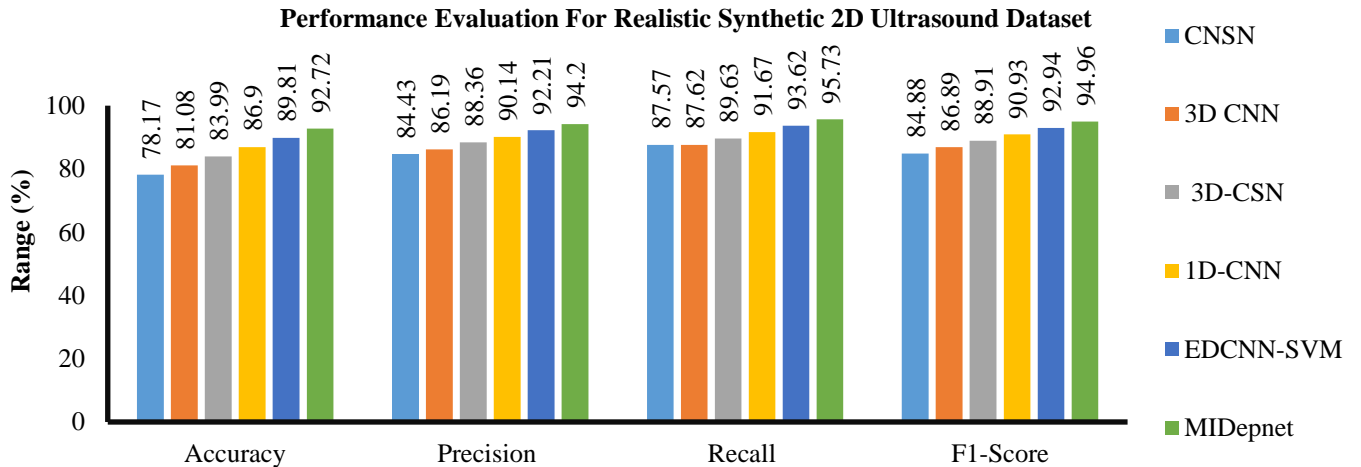

Fig. 7 Performance analysis of the proposed and existing model for MI detection by realistic synthetic 2D ultrasound dataset

Table 2. List of optimal hyperparameters

Models	Parameter	Range
CNSN [12]	Filter Size	96
	Regularization Factor	0.0004
	Learning rate	0.0001
	Activation function	Sigmoid
	Epochs	15
	Loss Function	MSE
	Batch Size	24
	Optimizer	Adam
3D-CNN [13]	No. of Conv. Layer	3
	Output Dimension	150
	Learning rate	0.001
	Activation function	ReLU
	Epochs	100
	Batch Size	8
	Optimizer	RMSProp
	Loss Function	Binary CE
3D-CSN [15]	No. of Conv. Layer	3
	Learning rate	0.0001
	Activation function	ReLU

	Loss Function	CE
	Batch Size	12
	Optimizer	Adam
1D-CNN [19]	No. of Conv. Layer	3
	Kernel Size	15
	Learning rate	0.1
	Filter Size	11
	Activation function	ReLU
	Loss Function	Categorical CE
	Batch Size	64
	Optimizer	Adam
E-D CNN-SVM [10]	Kernel Size	3
	Learning rate	0.001
	Activation function	Tanh
	Epochs	25
	Loss Function	CE
	Batch Size	32
	Optimizer	Adam
Proposed MIDepnet	No. of Residuals	18
	No. of Conv. Blocks	6
	Kernel Size	3
	Regularization Factor	L_2 (0.08)
	Filter Size	32
	Learning Rate	0.0001
	Activation function	ReLU
	Epochs	150
	Loss Function	CE
	Batch Size	64
	Optimizer	Adam
	Dropout	0.5

The statement asserts that the MIDepnet model outperforms than all previous models in terms of precision, recall, f1-score, and accuracy; accordingly, it realizes that the accuracy of the MIDepnet is 14.2%, 11.36%, 9.99%, 8.52% and 2.84%, higher than CNSN, 3D CNN, 3D-CSN, 1D-CNN and EDCNN-SVM models. The precision of the MIDepnet is 10.46%, 8.37%, 6.28%, 4.19% and 2.09% higher when compared to the existing models accordingly. The recall of the MIDepnet model is 10.55%, 8.44%, 6.33%, 4.22% and 2.11% higher than the other classical models correspondingly. Similarly, the F-measure of the MIDepnet model is 10.51%, 8.41%, 6.3%, 4.2% and 2.1% higher than the other traditional models, respectively. Figure 7 shows the efficacy of various models on the Realistic Synthetic 2D Ultrasound Dataset for MI identification. It claims that the MIDepnet model has a higher success rate than previous models. Accordingly, it determines that the accuracy of the MIDepnet is 14.55%, 11.64%, 8.73%,5.82 and 2.91% higher than CNSN, 3D CNN, 3D-CSN, 1D-CNN and EDCNN-SVM models. The precision of the MIDepnet is 9.77%, 8.01%, 5.84%, 4.06%, and 1.99% improved when compared to the other prediction models accordingly. The recall of the MIDepnet model is 8.16%, 8.11%, 6.1%, 4.06% and 2.11% higher than the other models correspondingly. Similarly, the F-measure of the MIDepnet

model is 10.08%, 8.07%, 6.05%, 4.03% and 2.02%, increased higher than the other detection models, respectively. In the literature, EDCNN-SVM [10], 3D-CNN [13], and 1D-CNN [19] models have utilized the HMC-QU dataset. Similarly, CNSN [12] and 3D-CSN [15] utilized the Realistic Synthetic 2D Ultrasound Dataset. Hence, this framework validates both datasets' proposed and existing models using the parameters in Table 2. The above comparison proves that the proposed model determines efficient results on both datasets for the MI prediction using echo images. This is because the proposed model efficiently synthesizes the echo images with diverse diversity, improving clinical relevance for MI detection. In addition, it uses the DPM in GAN to generate sample data with less scaling time and enhance MI recognition.

5. Conclusion
In this paper, MIDepnet is proposed to address data insufficiency and provide an accurate prediction of MI. This method uses CGADDPM-GAN to synthesize the efficient image instances with more significant variances, ensuring relevant anatomical structures are preserved. The model uses a DPM for network parameterization and incorporates CLL for efficient synthetic data generation. Also, this model employs representation learning, resulting in smaller batch

sizes and lower training time. The features from CGADDPM-GAN are fed into an E-D CNN for segmentation, feature innovation and CNN for MI identification. Finally, the test findings demonstrate that the MIDepnet model achieves an accuracy of 92.72% and 93.75% in comparison to other diagnosis models such as ED- CNN, CNSN, 3D-CNN, 3D-CSN and 1D-CNN when applied on HMC-QU and Realistic Synthetic 2D Ultrasound Dataset. The proposed model effectively addresses the insufficiency of echocardiographic images for MI detection.

But, the large number of network parameters and high computational complexity hinders real-time medical image segmentation for rapid therapy and diagnosis. So, future work will focus on parameter reduction techniques to enhance segmentation accuracy, reduce complexity and enable faster and 0 efficient clinical applications.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Nader Salari et al., "The Global Prevalence of Myocardial Infarction: A Systematic Review and Meta-Analysis," *BMC Cardiovascular Disorders*, vol. 23, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Lei Lu et al., "Myocardial Infarction: Symptoms and Treatments," *Cell biochemistry and biophysics*, vol. 72, no. 2, pp. 865-867, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[3] Lei Li et al., "Multi-Modality Cardiac Image Computing: A Survey," *Medical Image Analysis*, vol. 2023, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Petros Nihoyannopoulos, and Jean Louis Vanoverschelde, "Myocardial Ischaemia and Viability: The Pivotal Role of Echocardiography," *European Heart Journal*, vol. 32, no.7, pp. 810-819, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[5] Maryam Esmaeilzadeh, Mozhgan Parsaee, and Majid Maleki, "The Role of Echocardiography In Coronary Artery Disease and Acute Myocardial Infarction," *The Journal of Tehran University Heart Center*, vol. 8, no. 1, pp. 1-13, 2013. [Google Scholar] [Publisher Link]

[6] James N. Kirkpatrick et al., "Echocardiography In Heart Failure: Applications, Utility, and New Horizons, *Journal of the American College of Cardiology*, vol. 50, no. 5, pp. 381-396, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sachin B. Malik, "Transthoracic Echocardiography: Pitfalls and Limitations as Delineated at Cardiac CT And MR Imaging", *Radiographics*, vol. 37, no. 2, pp. 383-406, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Konstantinos C. Siontis et al., "Artificial Intelligence-Enhanced Electrocardiography in Cardiovascular Disease Management," *Nature Reviews Cardiology*, vol. 18, no. 7, pp. 465-478, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] M. Tokodi, B. Magyar, A. Soos, M. Takeuchi, M. Tolvaj, B. K. Lakatos, ... and A. Koács, "Deep Learning-Based Prediction of Right Ventricular Ejection Fraction using 2D Echocardiograms," *JACC: Cardiovascular Imaging*, vol. 16, no. 8, pp. 1005-1018, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Aysen Degerli et al., "Early Detection of Myocardial Infarction In Low-Quality Echocardiography," *IEEE Access*, vol. 9, no. 1, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yanhui Guo et al., "Automatic Myocardial Infarction Detection In Contrast Echocardiography Based on Polar Residual Network," *Computer Methods and Programs in Biomedicine*, vol. 198, no. 10, pp. 1-39, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ewan Evain et al., "Motion Estimation by Deep Learning in 2D Echocardiography: Synthetic Dataset and Validation," *IEEE Transactions on medical imaging*, vol. 41, no. 8, pp. 1911-1924, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Oumaima Hamila et al., "Fully Automated 2D and 3D Convolutional Neural Networks Pipeline for Video Segmentation and Myocardial Infarction Detection in Echocardiography", *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37417-37439, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ghada Zamzmi et al., "Real-Time Echocardiography Image Analysis and Quantification of Cardiac Indices," *Medical image analysis*, vol. 80, no. 1, pp. 1-51, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Yinlong Deng et al., "Myocardial Strain Analysis of Echocardiography Based on Deep Learning," *Frontiers in Cardiovascular Medicine*, vol. 9, no. 1, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] X. Lin, F. Yang, Y. Chen, X. Chen, W. Wang, X. Chen, ... and K. He, "Echocardiography-Based Ai Detection of Regional Wall Motion Abnormalities and Quantification of Cardiac Function in Myocardial Infarction", *Frontiers in Cardiovascular Medicine*, Vol. 9, No. 1, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tuan Nguyen et al., "Ensemble Learning of Myocardial Displacements for Myocardial Infarction Detection In Echocardiography," *Frontiers in Cardiovascular Medicine*, vol. 10, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Chitra Balakrishnan, and V. D. Ambeth Kumar, "IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks," *Diagnostics*, vol. 13, no. 4, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Aysen Degerli et al., "Early Myocardial Infarction Detection Over Multi-View Echocardiography," *Biomedical Signal Processing and Control*, vol. 87, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Shengrong Li et al., "Seismic Fault Detection Using an Encoder–Decoder Convolutional Neural Network with a Small Training Set," *Journal of Geophysics and Engineering*, vol. 16, no. 1, pp. 175-189, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Kenya Kusunose et al., "Clinically Feasible and Accurate View Classification of Echocardiographic Images Using Deep Learning," *Biomolecules*, vol. 10, no. 5, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Aysendegerli, HMC-QU Dataset , Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/aysendegerli/hmcqu-dataset

[23] Martino Alessandrini et al., "Realistic Vendor-Specific Synthetic Ultrasound Data for Quality Assurance of 2-D Speckle Tracking Echocardiography: Simulation Pipeline and Open Access Database," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 3, pp. 411-422, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[24] Cristiana Tiago et al., "A Domain Translation Framework with an Adversarial Denoising Diffusion Model to Generate Synthetic Datasets of Echocardiography Images," *IEEE Access*, vol. 11, no. 1, pp. 17594-17602, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Long Teng, Zhongliang Fu, and Yu Yao, "Interactive Translation in Echocardiography Training System with Enhanced Cycle-GAN," *IEEE Access*, vol. 8, pp. 106147-106156, 2022. [CrossRef] [Google Scholar] [Publisher Link]