RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26515

Lightweight MIDepnet: an Efficient Deep Learning Approach for Myocardial Infarction Detection using Enhanced Image Segmentation

B. Arockia Valanrani [a,*], S. Devi Suganya [a]

[a] Department of Computer Science, Vellalar College for Women, Erode-638012, Tamil Nadu, India

* Corresponding Author Email: valansajeev17phd@gmail.com

DOI: https://doi.org/10.54392/irjmt26515

Received: 17-03-2026; Revised: 10-08-2026; Accepted: 15-09-2026; Published: 24-09-2026

Abstract: Timely identification and detection of Myocardial Infarction (MI) are essential for mitigating cardiac tissue injury and reducing mortality risk. For accurate MI prediction, many algorithms based on Deep Learning (DL) have been created. Amongst, The Encoder-Decoder Convolutional Neural Network (E-D CNN) for image segmentation and a separate CNN model for MI prediction were used in the development of the deep network model for MI detection (MIDepnet). However, the E-D CNN involves many parameters and heavy architectures, hindering real-time image segmentation and detection. To address this, Lightweight E-D CNN (L-E-D CNN) segmentation model is developed for effective MI prediction. In this model, split blocks are introduced in L-E-D CNN which constitutes to two channels. The first channel consists of two parallel paths, one with standard convolution and the other with Multi Dense Dilated Depthwise Separable Convolution (MDDDSC) to capture both local and global features. The use of MDDDSC increase the receptive field with less parameters size. The second channel uses two factorized convolutions with an adaptable dilation rate to increase the receptive field without increasing parameters size. Additionally, a segmented mask training, morphological operations and post-processing pipeline are performed to improve segmentation. The segmented mask training is adopted to learn the inner region with and without papillary muscles and the outer ventricular region to enhance the boundary delineation. The predicted inner mask is subtracted from the outer mask to obtain a refined and separated ventricular region. Morphological operations are then applied to smooth boundaries, fill small gaps and remove minor artifacts. Then, the post-processing stage are applied to improve the boundary clarity, reduces noise and enhances segmentation accuracy. Lastly, the CNN-LSTM model is used for MI prediction using the features obtained from the L-E-D CNN. The complete model is termed as Lightweight MIDepnet (LMIDepnet). Finally, the outcomes show that the LMIDepnet algorithm on the Hamad Medical Corporation, Tampere University and Qatar University (HMC-QU) dataset and Realistic Synthetic 2D Ultrasound dataset obtains the accuracy of 95.97% and 94.89%, respectively, compared to the UNet + Support Vector Machine (SVM), UNet + CNN, DeepLabV3 + ResNet50, UNet + DenseNet121, SegNet + Saimese Neural Network (SNN) models.

Keywords: Myocardial Infarction, Deep Learning (DI), Image Segmentation, Parameter Reduction, Post Training.

1. Introduction

Myocardial Infarction (MI) is a severe cardiovascular disorder, which occurs when the coronary blood supply is suddenly diminished or completely blocked, leading to myocardial ischemia and irreversible injury unless treated quickly [1]. This disruption may cause serious harm to the heart muscle, which might be lethal. Shortness of breath, anxiety, sweating, dizziness, and chest and body discomfort are some of the symptoms [2]. Reducing MI mortality requires prompt identification and treatment. MI can be detected by a variety of imaging modalities, including CT scans, MRI scans, Echocardiography, Cardiac Magnetic Resonance imaging (CMR), and Chest X-Rays (CXR) [3]. Echocardiography is a commonly used imaging technique, easy to access, cost effective, and non-invasive, which can evaluate cardiac structure, Real-time anomalies in regional wall motion and Left ventricular (LV) function, and so is widely used in the assessment of MI [4].

Echocardiography uses ultrasound to visualize cardiac anatomy and motion, while Doppler modalities are used to assess blood-flow velocities and hemodynamics [5]. Modern examinations are commonly acquired as dynamic cine loops spanning one or more

cardiac cycles, enabling joint analysis of ventricular structure and temporal motion [6]. Echocardiographic studies also include multiple standardized views and recording modes that can be analyzed from still frames or video sequences [7]. Despite its non-invasive nature and broad clinical utility, echocardiography remains operator dependent, and manual interpretation can be time-consuming and variable.

Artificial Intelligence (AI), particularly Deep Learning (DL), has therefore been increasingly investigated for automated echocardiographic interpretation [8]. Several DL frameworks have been created to use echocardiography data to predict and categorize cardiovascular diseases. Amongst, a three-stage model [9] was proposed for prompt identification of MI using echocardiography. The initial step involves developing the E-D CNN method for LV wall segmentation in every echo frame. The following step is the creation of a feature-engineering block using the predicted segmentation masks; the third step is the identification of MI using the SVM algorithm. However, these models depend on annotated echocardiographic data, and the preparation of high-quality ground truth can be labor intensive. By producing additional echocardiogram image samples, Generative Adversarial Networks (GANs) [10] have been investigated to augment training data for MI diagnosis. Nevertheless, GAN-generated images may show limited diversity or anatomical inconsistency, which can restrict their usefulness for MI prediction.

So, MIDepnet [11] created a model for image augmentation called CGADDPM-GAN, which employs a Contrastive Guided Adversarial Denoising Diffusion Probabilistic Model. to generate a variety of high-quality echo images. This GAN model synthesizes and performs domain translation on echo visuals while keeping crucial anatomical details through a Diffusion Probabilistic Model (DPM) with a reverse denoising approach for improved image quality. Variational inference is used to train optimal transition parameters, aligning data with the target distribution. When training with smaller batches, Contrastive Learning (CLL) improves representations even more, leading to better performance with less time spent training. LV wall pictures were segmented using the E-D CNN paradigm, and a separate CNN model was utilized to predict MI from these segmented images. This E-D CNN model has shown very promising technique for segmentation in ultrasound echocardiogram images. But medical image segmentation for fast the evaluation and treatment in real-time is challenging because to its significant complexity and large quantity of network settings.

Hence, in this work, LMIDepNet is developed to overcome the limitations of the conventional E-D CNN by efficiently segmenting echo images for MI prediction. In this framework, Lightweight Encoder–Decoder CNN (L-E-D CNN) is proposed that uses split block for parameter reduction and post-processing pipeline for increasing segmentation performance. In the new split block structure, two channels are performed. In the first channel, a dual-path structure is utilized where standard convolution operates alongside with MDDDSC to effectively record each local and global contextual features through an expanded receptive field while maintaining a reduced parameter size. In the second channel, in order to enhance the receptive field while minimizing the number of parameters, two successive factorized convolutions are utilized, with an adjustable dilation rate in the second convolution. Outputs from both channels are integrated using a pointwise convolution to form the final feature representation which preserves detailed spatial information required for accurate segmentation while reducing model's complexity.

In addition, post-processing with segmented mask training strategy and morphological operations are employed to improve boundary delineation. As a first step, we use a segmented mask training technique to improve boundary learning. This involves training the model independently for the inner area (with and without papillary muscles) and the outside portion of the left ventricle. A refined ventricular area with unambiguous structural separation is obtained by subtracting the estimated inner mask from the outer mask after prediction. The morphological operations are applied to smooth boundaries, fill small gaps, and remove minor artifacts. The outputs from different training strategies are combined within the post-processing pipeline to enhance segmentation accuracy. Finally, the refined segmented images are then passed into feature engineering stage and the CNN model of MIDepnet for final MI prediction.

The study is structured as follows: Part II investigates relevant research. Part III details the LMIDepnet technique, while Part IV assesses it and Part V finishes the analysis and makes recommendations for future upgrades.

2. Literature Survey

Deng *et al.* [12] developed DL model for MI segmentation and movement assessment to produce stress rates on echo films. In this model, 3D cardiac segmentation network (3D-CSN) utilizes multiple frames of echo images to segment and obtain the centerline of a target setting. It estimates the pixel motion field, generates velocity vectors, propagates the centerline position and calculates global and local line arc dimensions for MI prediction. However, it needs huge amount of fake data for training and may be liable to overfitting.

Lin *et al*. [13] designed a 3D-CNN framework for the recognition of Regional Wall Motion Abnormalities (RWMAs) in MI sufferers utilizing echocardiography

data. The first step of view selection for the myocardium was conducted utilizing the Xception approach. Subsequently, the LSTM-U-Net model performs segmentation on each frame derived from outcomes of the Xception architecture. Finally, both the segmented and original video were summed and fed into the 3D-CNN model to detect the cardiac function in myocardial infarction. But this model obtains relatively lower Intersection over Union (IOU) to the distinct boundaries nearby the edge of imagery region.

Hamila *et al.* [14] depicted an automatic 2D and 3D-CNN model for MI identification and segmentation in echo films. The LV chamber was divided using the Apical four-Chamber (A4C) viewpoint, and 2D-CNN was used to pre-process the data in this approach. Then, the 3D-CNN was used to categorize and detect MI. On the other hand, it results with extreme computational period and memory. In order to extract real-time echoes, segment the heart, and measure cardiac parameters from echocardiogram pictures, Zamzmi *et al.* [15] created the Trilateral Attention Network (TaNet). The model uses a Spatial Converter System module to confine areas of concern, encodes rich features, aggregates feature embeddings for cardiac region segmentation and jointly segments, quantifies, and localizes cardiac classification systems. On the other hand, training was non-static, and the converged ratio was poor.

Evain *et al*. [16] established the DL algorithm for motion assessment using echo. In this model, a normalized cross-correlations was computed from the 2DE images and performed cost volume operation for patch judgements amongst two feature maps. The features were fed into a Convolutional Neural Sub-Network (CNSN) estimator to which predict dense displacement map. The network's parameters were optimized using a multi-layer loss function to estimate the motion from 2DE images. However, there is high computational time and lower learning rate.

By using Pre-Trained Recurrent CNNs (PRCNNs), Balakrishnan *et al*. [17] proposed an Internet of Things (IoT)-driven classification of echo pictures for the purpose of estimating the likelihood of cardiac disorders. The collected data was acquired from archaic cardiac individuals as well as current data through IoT connected devices. Fuzzy C-means algorithm (FCM) was employed for image pre-processing and segmentation. Finally, PRCNN was used for echo classification and DCNN for disease estimate. However, the parameters of these models were not optimized suitably causing ambiguity.

An ensemble learning method for MI displacement identification in echo was developed by Nguyen *et al.* [18]. Initially, the collected images of left ventricle (LV) were pre-processed and segmented using U-Net-model. Then, the attributes were retrieved from segmented images. Finally, the support vector machine (SVM) was applied for MI recognition and categorization. However, it was trained using a small dataset and was not fully automated.

For the purpose of MI prediction, Deepika and Jaisankar [19] created an Efficient Cardiac Ventricle 3D Network (ECV-3DN) and an Enhanced CNN Algorithm (ECNN). In this method, the collected echo images were pre-processed and augmented to increase the image quality. The ECNN model leverages CNN structure, transfer learning (TL) and attention mechanism for attribute retrieval and representation learning from echo films. Then, ECV-3DN was used to devise the spatial and temporal features from echo and ECG data for effective prognosis for MI. To overcome the problem of data interpretation from different sources, however, this approach needs a sophisticated feature fusion model.

An automated system for the identification of heart disorders combining DL and a self-supervised contrastive learning method was created by Holste *et al.* [20] and is called EchoCLR. The pre-training of EchoCLR involves contrastive learning to distinguish between patient videos and a temporal-reordering task in which the model predicts the correct order of shuffled video frames. This self-supervised strategy learns echocardiographic representations that can subsequently support downstream cardiovascular classification. More recently, Jamthikar *et al.* [21] investigated ultrasonic texture analysis (ultrasomics) for acute MI, using echocardiographic image features to characterize infarct-related myocardial texture and morphology across multisource data. Their findings support ultrasound-based tissue characterization as a complementary approach for identifying and localizing infarcted myocardium; however, broader validation remains important for establishing generalizability and the relationship of these image signatures to infarct extent. Table 1 represents a comparison of existing MI detection models based on segmentation backbone, temporal modeling, classifier, and data scale.

2.1 Research Gap

The literature reveals several advancements in DL based MI detection from echocardiogram images. However, key challenges remain in terms of high computational complexity, long training time and limited generalizability. Some models also require advanced feature fusion to address data interpretation issues across multiple modalities. Additionally, small dataset sizes and the need for improved interpretability in MI prediction hinder model performance. Models with large network parameters face difficulties in real-time applications due to slow processing and substantial data requirements. To overcome these challenges, new split blocks to reduce parameters and optimized training strategies are offered to enhance segmentation effectiveness and maintaining high precision for real-time medical applications.

Table 1. Comparison of Existing MI Detection Methods

Ref. No.	Segmentation Backbone	Temporal Modeling	Classifier	Data scale
[12]	3D-CSN	Optical flow network for motion	-	In Shantou University Medical College's First Affiliated Hospital, 150 echocardiography films from 50 patients
[13]	LSTM-Unet	LSTM	3D CNN	4,142 examinations (training/internal testing) and 2,811 examinations (external testing)
[14]	2D CNN	Sliding window: Spatial windowing, Temporal windowing	3D CNN	HMC-QU Dataset 162 A4C view recordings of 2D echocardiography
[15]	TaNet	Representation of Echo-Specific Data under Self-Supervision	View Classification	EchoNet-Dynamic - NIH IVC comprises 268 recordings of intravenous contrast (IVC) taken from 264 patients, 10,036 B-mode (A4C) videos taken from 10,036 patients at random, and 68 PLAX videos taken from 60 patients, Doppler Dataset - Doppler flows collected from patients from Clinical Center at NIH
[16]	RoI, GLS	PWC-Net	View Classification	CAMUS dataset contains 500 patients from the University Hospital of St-Etienne (France)
[17]	FCM	GRU	PRCNN, Softmax classifier	UCI Machine Learning Repository - The entire dataset is multivariate, and it has 132 instances with 12 unique attributes.
[18]	UNet, Net++, LinkNet, DeepLabV3, and PAN	Myocardial segment displacement tracking	SVM, LR, DT, and KNN	The HMC-QU dataset has 109 echocardiograms, which are used for training and validation purposes. For independent testing, a local clinical site in Vietnam provided the E-Hospital dataset, which contains 60 echocardiograms.

This study focuses on reducing computational complexity and enhancing model generalization for more effective MI diagnosis.

3. Proposed Methodology

This segment, the whole description for LMIDepnet is provided to effective segmentation task of echo images to predict MI. Figure 1 depicts the workflow of LMIDepnet model.

3.1 Data Preprocessing and Augmentation

All raw echocardiographic frames were first preprocessed by a standardized pipeline. All raw images and ground truth binary masks were resized with bilinear interpolation for raw images and nearest neighbor interpolation for ground truth binary masks to preserve precise mask boundaries. Lastly, the intensity values of the pixels were quantized in the range [0, 1] to make the numbers stable and also to help training the network to propagate the gradients optimally. Data augmentation was only carried out during the training stage to avoid overfitting and make the segmentation and classification more robust. In order to maintain exact spatial alignment both the raw frames and the binary ground-truth masks were simultaneously and identically processed with affine spatial transformations (random rotations (±15^∘), translation (±10%), horizontal flips (p=0.5) . Furthermore, the raw ultrasound frames were subjected to zero-mean Gaussian/speckle noise addition and contrast modulation of ±15%, solely in order to mimic actual imaging artifacts without changing the target mask contours.

3.2 Proposed L-E-D CNN for Image Segmentation

The L-E-D CNN model achieves an efficient balance between model parameters and segmentation accuracy, simplifying the complexities of MI prediction. Figure 2 depicts the visualization of the proposed L-E-D CNN model.

3.2.1 Split Block Structure

The parameter reduction is achieved by introducing a split block in E-D CNN to extract most relevant information for segmentation task. Figure 3 depicts the proposed channel split block.

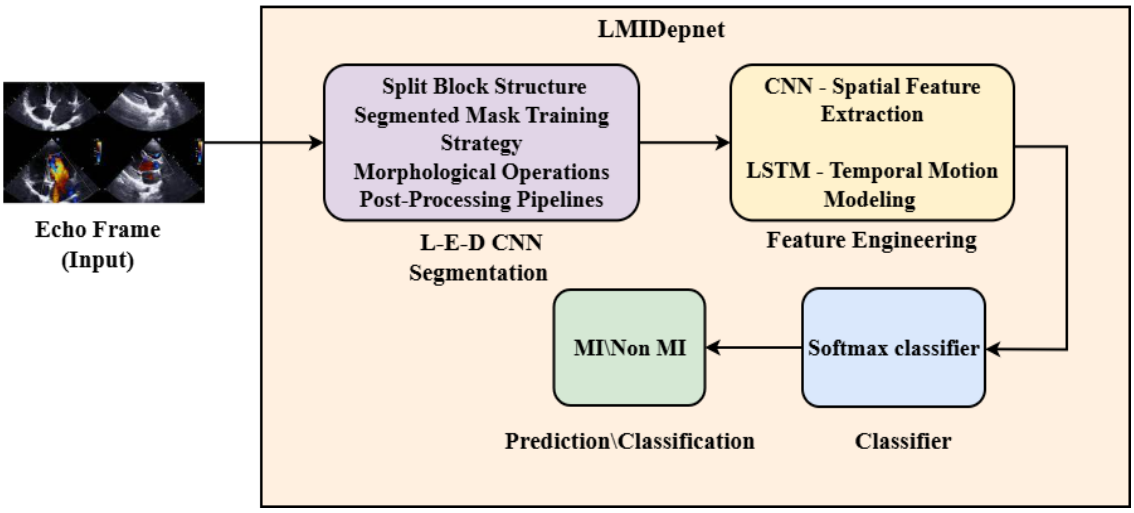

Figure 1. Schematic illustration of the LMIDepnet Model.

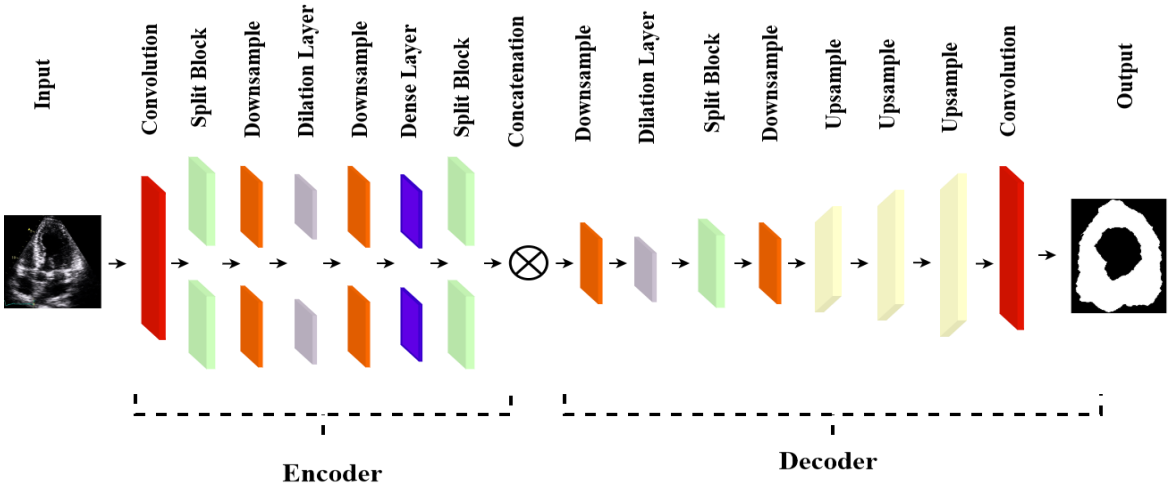

Figure 2. Depiction of Split block structure in L-E-D CNN.

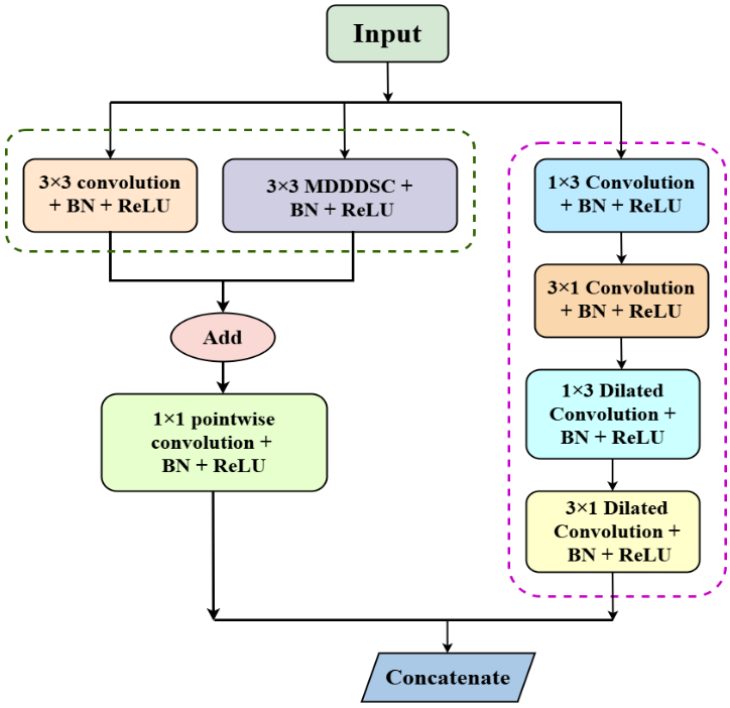

Figure 3. Structure of Proposed Split Block Model.

The newly constructed split block takes its cues from the Lightweight Encoder-Decoder Network known as LEDNet [22]. Two separate channels are established inside this block to record distinct aspects of the echo images, allowing more efficient extraction of relevant details. The first channel combines standard convolution with MDDDSC, whereas the second channel uses factorized and dilated convolutions. The design of the first channel also follows the general principle of densely connected atrous convolution, as exemplified by DenseASPP [23], where dense connectivity promotes feature reuse and atrous convolutions enlarge the receptive field without reducing spatial resolution. Together, both channels capture complementary local and global features to enhance segmentation performance.

3.2.1.1 Design of First channel

The first channel is divided into two parallel convolutional layers like standard convolutional and MDDDSC. This channel is designed to extract spatial features efficiently while improving feature propagation and reducing computational complexity.

The standard convolution plays a crucial role in capturing local spatial edges, textures, and structural details are examples of patterns present in the echocardiographic images. Initially, the input feature map is passed through 3×3 convolutional layer activation, then Rectified Linear Unit (ReLU) and Batch Normalisation (BN) activation. In this convolutional operation, each convolutional filter slides over the input feature map and performs element-wise multiplication and summation to generate a new feature representation. The network can learn significant things during this approach spatial relationships between neighbouring pixels. To further improve feature propagation and reuse previously learned features, dense connectivity is adopted in this branch. Dense connection works by feeding the feature maps from all the layers below it into the current layer. The network is able to learn more discriminative features with less redundancy because to this technique, which also improves gradient flow and increases feature reuse. The dense connectivity can be mathematically expressed as in Eq. (1)

$$A_x = H_x([A_0, A_1, \dots, A_{x-1}]) \quad (1)$$

Where, $[A_0, A_1, \dots, A_{x-1}]$ stands for the combined feature maps from all the layers before it, while H_x is a composite function that includes BN, ReLU, convolution, and dropout. In order to manage the model's complexity, the growth rate parameter is usually set to 16, and each layer's output produces z feature maps.

MDDDSC improves computational efficiency while maintaining strong feature representation capability. In this sub-branch, BN and ReLU activation are applied after processing the input feature map using a 3×3 dilated convolution. Dilated convolution allows the network to acquire more contextual information from the input picture by expanding the receptive area without increasing the number of learnable parameters. In MDDDSC, the convolution kernels are decomposed to reduce parameters without sacrificing model performance. In the dense dilated layer, the conventional 3×3 convolution is decomposed into two sequential convolutions of size 3×1 and 1×3. Figure 4 shows the MDDDSC design. The dense dilated layer (Figure 5a and b) and the conventional dense layer are shown.

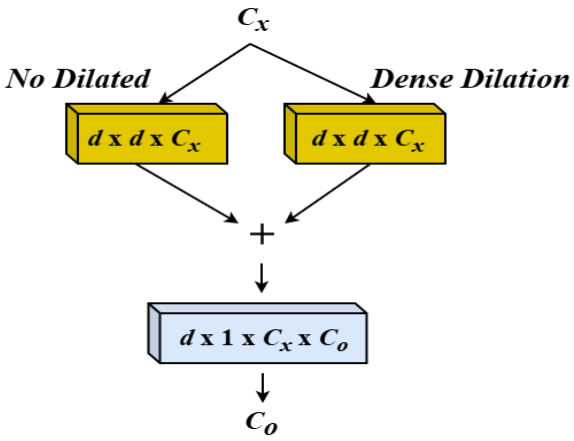

Figure 4. MDDDSC with number of parameters in each step of convolution.

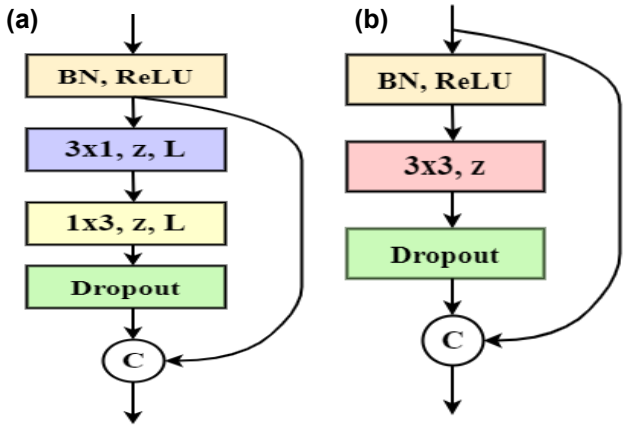

Figure 5 (a). Dense Dilated Layer **(b)** Dense Dilated Layer.

Reducing the amount of parameters without sacrificing good spatial feature extraction is achieved using this factorised convolution, as well as combined with BN, ReLU activation and dropout for improved training stability. Eq. (1) can therefore be reformulated as in Eq. (2).

$$A_x = H_x^2\left(H_x^1([A_0, A_1, \dots, A_{x-1}])\right) \quad (2)$$

In semantic segmentation, down-sampling expands the receptive area of the convolution kernel, which in turn allows it to collect more context. On the other hand, precise edge forms and other spatial

features might be lost. To minimize down-sampling and further enlarge the receptive field without losing spatial resolution, dilated convolution is incorporated. In dilated convolution, the spacing between kernel elements is controlled using a dilation rate r, allowing the convolution to cover a larger region of the image without increasing the kernel size or the number of parameters. Thus, it makes it possible to keep the same resolution while flexibly aggregating contextual information across scales. A 5×5 receptive field may be achieved with a conventional 3×3 convolution kernel and a dilation rate of 2, while a dilation rate of 3 produces a 7×7 receptive field. The mathematical formulation of dilated convolution is expressed as in Eq. (3)

$$B(u,v) = \sum_{x=1}^{u} \sum_{y=1}^{v} A(u + r \times x, v + r \times y)\, w(x,y) \quad (3)$$

Here, A is the input feature map, $B(u,v)$ is the output at dilated from the input, and $w(x,y)$ is a filter with U's length and V's width. To further increase the receptive field size, the Dense Dilated Block employs a sequence of dilated convolution layers with dilation rates $r = 2,4,8$ in a consecutive fashion. The spatial resolution of the feature maps is preserved by this structure, which allows for effective aggregation of multi-scale contextual information.

In combination with the depthwise operations in MDDDSC, it requires only $C_x \times d \times d$ additional parameters compared with standard depthwise separable convolution, resulting in approximately 87% reduction in learnable parameters compared to conventional convolution layers. This parameter-efficient convolution mechanism is particularly beneficial for semantic segmentation, as it reduces model complexity while expanding the receptive field to capture a larger portion of the input image. Consequently, the proposed structure effectively preserves important spatial details while enabling faster and more accurate segmentation.

The dense dilated block with three convolutions and the dilated convolution rate in layer is shown in Figure 6. Results from the standard convolution branch and the MDDDSC branch are added and then passed to a pointwise convolution layer. Pointwise convolution performs a 1 × 1 convolution, which is used to project feature maps across channels while reducing the number of parameters. This operation enables efficient integration of the extracted features from both branches and improves the network capability. The pointwise convolution helps the model learn both local and global spatial relationships while maintaining computational efficiency.

3.2.1.2 Design of Second Channel

By using two successive factorised convolutions in the second channel, we may increase the field of vision without changing the number of parameters and extract more important information. The process of factorising the convolutional kernel begins with two convolutional layers that follow one another. The learning parameter of the 2D convolutional layer, denoted as $W \in \mathcal{R}^{C_x \times d \times d \times C_o}$, is defined as the product of the number of input channels C_x and the number of output channels C_o. Each convolution's kernel dimension is denoted by $d \times d$. The $2 \times d \times C_x \times C_o$ learning parameters are determined by the two convolutions of $1 \times d$ and $d \times 1$ kernels that make up this factorised layer. A typical h×w convolution kernel, for instance, can be split into two one-dimensional convolutions of size $h \times 1$ and $1 \times w$, in that order. A 3×3 convolution is substituted by a 3×1 convolution and a 1×3 convolution in this structure. The number of parameters in a conventional 3×3 convolution is reduced from 9 to 6 via factorisation, resulting in approximately 33% reduction in parameters. The configuration of factorized convolution is depicted in Figure 7.

Second, dilated factorized convolution adjusts dilation rate r, without changing the amount of parameters, expand the convolutional kernel's field of view. A convolution's dilation rate determines the distance between kernel elements, which in turn increases the receptive field without boosting the kernel's size or parameters. The parameters of a dilated convolution determine the rate of dilatation and the distance between kernel elements.

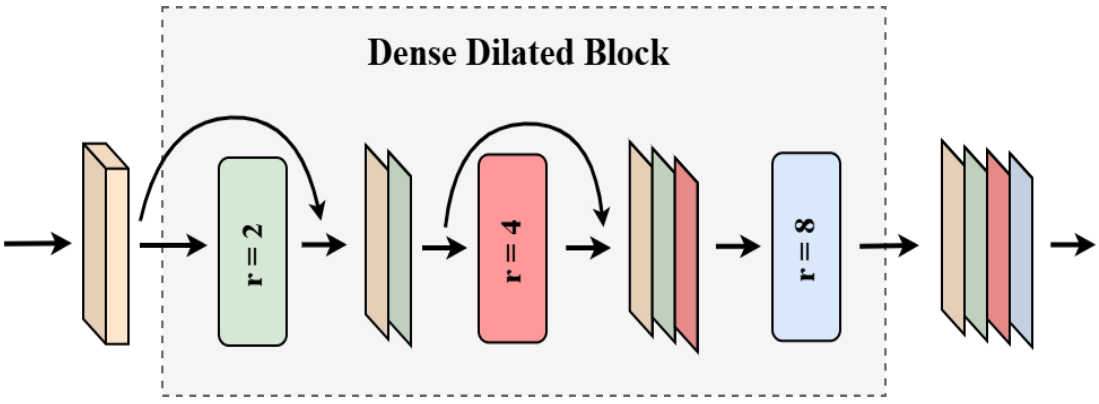

Figure 6. A Dense Dilated Block using a rate of dilation and three convolutions.

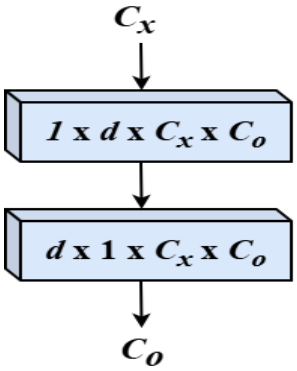

Figure 7. Structure of Factorized Convolution

Table 2. Layer Configuration of Split-Block structure of L-E-D CNN model

Layer	Input	Type	Filters	Output
Input	—	—	—	256×256×1
1	Input Image	3×3 Convolution	16	256×256×16
1s2	1	Splitblock (r=2)	32	256×256×32
1s3	1s2	Downsample	32	128×128×32
1s4	1s3	Dilation Layer	32	128×128×32
1s5	1s4	Downsample	32	64×64×32
1s6	1s5	Denseblock	32	64×64×32
1s7	1s6	Splitblock (r=3)	32	64×64×32
1s2.1	1	Splitblock (r=2)	16	256×256×16
1s3.1	1s2.1	Downsample	16	128×128×16
1s4.1	1s3.1	Dilation Layer	32	128×128×32
1s5.1	1s4.1	Downsample	64	64×64×64
1s6.1	1s5.1	Denseblock	16	64×64×16
1s7.1	1s6.1	Splitblock (r=3)	32	64×64×32
1s8	1s7.1 ⊗ 1s7	Concatenation	—	64×64×64
1s9	1s8	Downsample	16	32×32×16
1s10	1s9	Dilation Layer (r=2)	32	32×32×32
1s11	1s10	Splitblock (r=6)	128	32×32×128
1s12	1s11	Dilation Layer (r=2)	128	32×32×128
1s13	1s12	Upsample	64	64×64×64
1s14	1s13	Upsample	32	128×128×32
1s15	1s14	Upsample	16	256×256×16
1s16	1s15	3×3 Convolution	128	256×256×128
1s17	1s16	1×1 Convolution	16	256×256×16
Output	1s17	Pointwise Convolution	1	256×256×1

The dilation rate establishes the distance between neighbouring kernel values; so, it misses certain pixels rather than applying the kernel to every neighbouring pixel. A $3x3$ kernel with $r = 2$ uses 9 parameters instead of 25 for the same field of vision as a $5x5$ kernel. At $r = 1$, the process acts like a regular convolution, and dilation does not occur. On the other hand, semantic segmentation tasks might suffer from a loss of local information if dilation rates are too high.

The features extracted from both channels are concatenated to form the final feature representation, reducing parameters while maintaining detailed features. The L-E-D CNN architecture is shown in Table 2 with the layer types, each layer input, number of filters and output for respective layers.

3.2.2 Post-Processing Pipelines with Segmented Mask Training Strategy and Morphological Operations

3.2.2.1 Segmented Mask Training Strategy

In this process, two training approaches for LV segmentation is performed. In the first approach, the model is trained directly using the echo images as input and the complete segmentation mask as output. This training strategy is applied separately for segmentation

including papillary muscles and segmentation excluding papillary muscles. In the second approach, a different strategy is introduced to improve segmentation performance. The inner and outer regions are trained independently of each other so that the model can forecast the entire region instead of all at once. The inner region trained both with and without papillary muscles, while the outer region is trained independently. After prediction, the estimated inner mask is subtracted from the estimated outer mask to obtain the final segmented region. Any negative values produced after subtraction are mapped to zero. This strategy allows the model to focus more specifically on inner and outer structures before combining them. The results from the direct training and the mask subtraction strategy are used for refinement in the post-processing stage.

3.2.2.2 Morphological Operations

To refine the segmentation results and improve structural consistency, morphological image processing operations are applied during the post-processing stage. A circular structuring element of size 7 × 7 is used to preserve anatomical smoothness while correcting small irregularities in the predicted masks. Two morphological operations are performed sequentially like morphological closing followed by morphological opening. Filling tiny holes and gaps inside the segmented left ventricle area is the initial step in morphological closure, which involves dilatation followed by erosion. This operation strengthens boundary continuity, particularly in the outer epicardial region, and helps recover small unsegmented areas that may appear due to weak edge prediction. Subsequently, morphological opening, in order to eliminate tiny, isolated false positive areas, a process that involves erosion and dilatation is carried out. This step eliminates tiny incorrectly segmented areas either inside the ventricular cavity or outside the true boundary. The combined effect of closing and opening improves boundary smoothness, reduces segmentation noise, and ensures a more coherent and anatomically consistent final segmentation mask.

3.2.2.3 Post-Processing Pipelines

Using 2D echocardiographic pictures as inputs and segmentation masks as supervisory outputs, the proposed L-E-D CNN model is trained. Figure 8 depicts the post-processing pipeline with morphological operation. In this training approach, the network learns to predict the complete left ventricle region from the echo image, considering both cases including papillary muscles and excluding papillary muscles. In addition to this conventional training method, a second segmentation strategy is employed to improve boundary delineation. An inner (endocardial) and outside (epicardium) region prediction model is trained independently in this method.

The inner region is learned both with and without papillary muscles, while the outer region is trained independently. After inference, by subtracting the expected outer mask from the predicted inner mask, we may get the final ventricular region, and any negative values resulting from this subtraction are set to zero. For the proposed work, the outputs obtained from the direct training approach and the subtraction-based approach are subsequently combined using a union operation to enhance segmentation robustness. Finally, morphological closing followed by morphological opening is applied using a 7 × 7 circular structuring elements to fill small gaps and remove minor incorrectly segmented regions, resulting in a refined and anatomically consistent final LV segmentation.

3.3 Feature Engineering and Classification

The L-E-D CNN generates segmented LV wall images. Feature engineering [9] is performed to remove the inner boundary of the segmented LV wall to identify the Endocardial Boundary (EB). The EB is divided into six segments, and echocardiographic data are used to calculate displacement for each segment. These features can be represented using colour-coded LV wall and EB segments, displacement arcs, segment-area curves, and extreme-displacement snapshots to support MI assessment. Next, a CNN-LSTM classifier is used for MI prediction from the derived LV features.

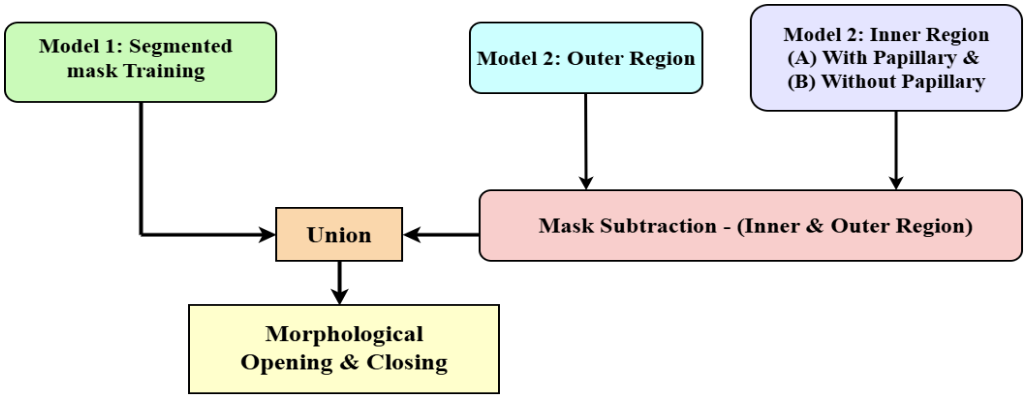

Figure 8. Post-processing pipeline with training strategy and morphological operation.

This design is supported by prior echocardiography work showing that CNNs can extract spatial image features while LSTMs model temporal wall-motion information across a cardiac cycle for MI classification [24]. In LMIDepnet, this stage combines spatial and temporal information before the final Softmax decision. The LMIDepnet workflow is summarized below.

Algorithm: LMIDepnet model for MI Prediction

Input: Combination of 2D echocardiography data sets

Begin

1. Augment the collected echo frames using to increase their diversity and generate high-quality echocardiography images.
2. Divide the input image into two parallel channels for feature extraction.
3. **\\First Channel**
4. The input image is processed through two parallel convolution paths.
5. Apply 3×3 convolution in first path to extract local spatial features.
6. Perform 3×3 MDDDSC in the second path to capture wider contextual information.
7. The outputs of these two paths are combined using an add operation.
8. Perform 1×1 pointwise convolution to fuse the combined features and reduce channel dimensionality.
9. **\\ Second Channel**
10. Perform the input feature map using factorized convolutions to improve efficiency.
11. Complete the factorisation of a typical 3×3 convolution by applying 1×3 convolution and then a 3×1 convolution.
12. Perform dilated factorized convolutions using 1×3 dilated convolution and 3×1 dilated convolution to capture spatial context without significantly increasing parameters.
13. Concatenate the combined outputs to form the final feature representation, reducing parameters while maintaining detailed features.
14. **\\Post-processing with segmented mask training strategy and morphological operation**
15. To improve boundary delineation, use a segmented mask training approach to learn the inner and outer ventricular regions, as well as the region inside and without papillary muscles.
16. After predicting the inner mask, subtract it from the outer mask to obtain a refined and separated ventricular region.
17. Apply morphological operations to smooth boundaries, fill small gaps and remove minor artifacts.
18. Perform post-processing stage to improve the boundary clarity, reduces noise and enhances segmentation accuracy.
19. Feed the segmented image features into feature engineering for MI prediction.
20. Use an individual CNN model to classify the segmented features for MI prediction.

End

Output: Segmented image regions for final MI prediction outcome

4. Results and Discussion

An Intel® CoreTM i7-4210 CPU running at 1.7 GHz base frequency and 2.70 GHz boosted frequency accompanied by 4GB of RAM, NVIDIA 3090 GPU (24GB VRAM) and a 1TB hard drive, running Windows 1 64-bit, is utilised to develop the L-E-D CNN model in a Python module. The F1-score, accuracy, recall, and precision constitute the basis of performance analysis. The hyperparameter settings for both the current and proposed models are presented in Table 3.

Table 3. Hyperparameter Settings

Hyperparameters	**Values**
UNet+SVM, UNet+CNN, DeepLabV3+ResNet50, UNet+DenseNet121, SegNet+SNN, and proposed LMIDepnet	
Learning rate	0.001
Optimizer	Adam
Epochs	250
Batch size	32
UNet+SVM	
Kernel	Linear
Regularization factor	0.01
Gamma	0.0001
UNet+CNN, DeepLabV3+ResNet50, UNet+DenseNet121, SegNet+SNN and proposed LMIDepnet	
No. of convolution layers	5
No. of filters in convolution layer	128
Kernel size	3
Pool size	2

4.1 Dataset Description

This research employs the following dual datasets:

HMC-QU Dataset [25]: The HMC-QU benchmark was created through collaboration among Hamad Medical Corporation, Tampere University, and Qatar University using de-identified echocardiographic recordings acquired in 2018–2019. The released dataset contains 322 videos from 162 unique patients, including 93 patients with MI and 69 without MI. It includes 162 apical four-chamber (A4C) recordings and 160 apical two-chamber (A2C) recordings. A subset of 109 A4C recordings provides frame-by-frame LV segmentation masks. The recordings have a temporal resolution of 25 frames/s, with spatial resolution varying across acquisitions. The source cohort from which the benchmark was curated comprised more than 10,000 echocardiographic examinations and more than 800 acute ST-elevation MI admissions.

The data was split into two sets: one with 129 individuals for training and another with 33 subjects for held-out testing. A subject-stratified 5-fold cross-validation scheme was carried out only on the 80% training cohort to optimize model hyper-parameters and to monitor training convergence to prevent data leakage, and a pseudo-random seed (seed = 42) was fixed to assure the absolute fold reproducibility. Four sub-folds of the fold (103 subjects) were trained with augmented samples and the other sub-fold of the fold (26 subjects) was used as the unaugmented validation fold. The augmentation was performed as described in section 3.1. After training, the model of each fold level was tested on the 20% held-out test set (33 subjects / 730 raw frames). The average value of held-out test set is used to perform confusion matrix.

For the left ventricle segmentation task, an average of approximately 21.5 frames per subject were used with a total subject cohort of 109 (2,346 frames) consisting of 72 MI subjects (1,548 frames) and 37 non-MI control subjects (798 frames); a subject-wise stratified 80/20 train/test split was used. To ensure strict isolation of patients and to prevent leaking data from different time frames of cardiac cycles, all the frames of a patient were chosen to be used either in the training split or in the testing split. The training split (80%) consisted of 87 subjects (58 MI and 29 Non-MI) containing 1,870 raw frames (1,247 MI and 623 Non-MI), whereas the held-out test split (20%) comprised 22 subjects (14 MI and 8 Non-MI) containing 476 raw frames (301 MI and 175 Non-MI). The distribution of patients/sequences for training and testing for HMC-QU dataset is given in Table 4, respectively.

Realistic Synthetic 2D Ultrasound (RS-2DU) Dataset [26]: To evaluate robustness across vendor-specific image characteristics, the model was also tested on the realistic vendor-specific synthetic ultrasound benchmark. The resource provides simulated echocardiographic sequences generated for multiple ultrasound vendors and cardiac motion patterns, with ground-truth motion information designed for quality assurance of 2-D speckle-tracking echocardiography. In this study, 105 simulated sequences (5,934 2-D frames) were used, with sequence lengths ranging from 50 to 72 frames per cardiac cycle.

A temporal resampling pipeline was applied in the sequences to normalize each cine-loop sequence to a fixed frame duration of $T = 60$ frames per cardiac cycle, The RS-2DU benchmark (105 sequences / 6,300 frames) was then divided into an 80% training set (84 sequences / 5040 frames) and a 20% held-out test set (21 sequences / 1260 frames) for 100% experimental reproducibility and leakage prevention, in which the experimental random seed was fixed to 42. Model optimization was performed only in the training cohort (67 sub-training sequences and 17 validation sequences per fold, stratified 5-fold cross-validation). Augmentation as per described in section 3.1 was only applied to the sub-training folds to have 335 effective augmented sequences (20,100 frames) per fold. The 21 sequences (1,260 frames) from the 20% held out test set were left completely raw and unaugmented for final evaluation. The average value of held-out test set is used to perform confusion matrix.

Healthy Motion Pattern, Ischemic Motion Patterns based on specific coronary artery occlusion territories; "LADdist" stands for Left Anterior Descending Artery Distal Occlusion, LADprox (Proximal occlusion of Left Anterior Descending artery), LCX (Occlusion of Left Circumflex artery), RCA (Occlusion of Right Coronary Artery) are the classes of this dataset. The original cases of RS-2DU biomechanical simulation were grouped into five regional wall motion classes based on the clinical regional pathology: Healthy, Basal Septal, Apical, Mid-Lateral, Diffuse Hypokinesia. In Table 5, after augmentation on held-out RS-2DU test set is mentioned.

Table 4. Training-testing split details of HMC-QU dataset

Class / Diagnosis	Total Subjects	Original Frames	Train Set (80%) Subjects / Raw Frames	Train Set (3× Augmentation) Total Augmented Frames	Test Set (20%) Subjects / Raw Frames (No Aug)
MI	93	1,957	74 / 1,557	4,671	19 / 400
Non-MI	69	1,624	55 / 1,294	3,882	14 / 330
Total	162	3,581	129 / 2,851	8,553	33 / 730

Table 5. Training-testing split details of RS-2DU test set

Pathology / Motion Class	Original Sequences (Before Augmentation)	Original Frames (≈60 frames/seq)	Augmented Sequences (5× Effective Multiplier)	Effective Augmented Frames
Class 1: Healthy / Normal	17	1,020	85	5,100
Class 2: Basal Septal Infarction	17	1,020	85	5,100
Class 3: Apical Infarction	17	1,020	85	5,100
Class 4: Mid-Lateral Infarction	17	1,020	85	5,100
Class 5: Diffuse Hypokinesia	16	960	80	4,800
Total Training Set (80%)	84	5,040	420	25,200
Held-Out Test Set (20%)	21	1,260	21 (Unaugmented)	1,260 (Unaugmented)
Total Benchmark Cohort	105	6,300	441	26,460

4.2 Evaluation Metrics

In order to determine how well the system calculates MI using the datasets that were collected, the following metrics were used.

4.2.1 Segmentation Analysis Metrics

Dice Similarity Coefficient (DSC)

$$DSC = \frac{2\,|Y \cap \bar{Y}|}{|Y| + |\bar{Y}|} \quad (4)$$

Eq. (4) calculates a spatial overlap between the predicted mask ($\bar{Y}$) and the ground truth (Y).

Intersection over Union (IoU)

$$IoU = \frac{|Y \cap \bar{Y}|}{|Y \cup \bar{Y}|} = \frac{DSC}{2 - DSC} \quad (5)$$

Eq. (5) Gets the result by dividing the union area by the overlap area.

Boundary F1-Score (BF-Score)

It compares the similarity of the predicted contour CP with the ground truth contour CG. In contrast to region-based measures, from the boundary's recall and precision, it calculates the harmonic mean in a given distance threshold θ , and it is very sensitive to structural boundary detail errors.

$$BF - Score = \frac{2 \times P_{boundary} \times R_{boundary}}{P_{boundary} + R_{boundary}} \quad (6)$$

In Eq. (6) $P_{boundary}$ and $R_{boundary}$ are the boundary precision and recall, calculated between the CP and CG.

Boundary Precision $P_{boundary}$

The precision of the boundary is the number of points in the contoured boundary that are close to the actual (ground truth) boundary.

$$P_{\text{boundary}} = \frac{|CP \cap_\theta CG|}{|CP|} \quad (7)$$

In Eq. (7), $|CP \cap_\theta CG|$ denotes number of predicted boundary points that lie within a distance tolerance θ of the closest ground truth boundary point. $|CP|$ is the total count of points on the predicted boundary.

Boundary Recall $R_{boundary}$

The recall of the boundary is the number of points in the actual (ground truth) boundary that are close to the predicted boundary contour. It evaluates the completeness of the contour extraction based on how much of the actual contour boundary is predicted.

$$R_{\text{boundary}} = \frac{|CG \cap_\theta CP|}{|CG|} \quad (8)$$

In Eq. (8), $|CG|$represents as many points as there are along the ground truth border, and $|CG \cap_\theta CP|$ denotes number of ground truth boundary points within a distance tolerance θ (in pixels) from the predicted boundary points.

Hausdorff Distance (HD)

HD_{95} is a variant of the HD that incorporates the 95th percentile of the distances to the boundary rather than the absolute maximum distance, which makes it less sensitive to outlying pixels like a few mis-predicted ones.

$$H_{95}(CG, CP) = \max\left(P_{95_{cg \in CG}}\left\{\min_{cp \in CP} D(cg, cp)\right\}, P_{95_{cp \in CP}}\left\{\min_{cg \in CG} D(cg, cp)\right\}\right) \quad (9)$$

In Eq. (9), cg and cp are individual boundary points belonging to Ground truth contour (CG) and predicted contour (CP); $D(cg, cp)$ denotes the Euclidean distance between point cg and ; and $P_{95}\{.\}$ represents the 95th percentile of the given set of minimum distances, rather than the absolute maximum used in the standard HD.

4.2.2 Classification Analysis Metrics

- *Accuracy:* Part of the instances that were predicted with reasonable accuracy are identified (MI sufferers) versus the overall incidences (both MI and non-MI individuals) as,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

When the model accurately predicted MI individuals, it is shown by True Positive (TP) in Eq. (10). When the model correctly identified non-MI participants, it is represented by a True Negative (TN). A False Positive (FP) occurs when the model mistakenly identifies healthy individuals as MI cases. The quantity of MI that is incorrectly forecasted as normal is called False Negative (FN).

- *Precision:* A proportion of TPs is calculated to the entire amount of optimistic forecasts produced by the simulation, as delineated in Eq. (11).

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

- *Recall:* It computes a proportion of TPs towards the overall quantity of noticed satisfactory occurrences within dataset, as indicated in Eq. (12).

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

- *F1-score:* As in Eq. (13), it represents the average of accuracy and recall, which are balanced.

$$F1 - score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (13)$$

- *ROC-AUC:* It graphs the TP Rate (TPR) compared to FP Rate (FPR) at various barrier levels. It assesses the framework capability to differentiate among good and bad categories. The TPR and FPR is defined as in Eq. (14) and Eq. (15).

$$TPR = \frac{TP}{TP+FN} \quad (14)$$

$$FPR = \frac{FP}{FP+TN} \quad (15)$$

4.3 Performance Analysis of L-E-D CNN for Segmentation and Comparison with Baseline Models

This segment estimates the effectiveness of L-E-D CNN for segmentation compared to the baseline models, such as SegNet, U-Net, and DeepLabV3. Figure 9 shows the results of segmentation incorporating and omitting papillary muscle for one sample image.

All quantitative segmentation metrics (DSC, IoU, BF1, and HD_{95}) were calculated at the subject level. Predictions were first averaged over each individual subject and then the overall performance is reported as the mean ± standard deviation over all subjects in the test set. In Table 6 Quantitative evaluation of LV wall segmentation performance on the HMC-QU (22 subjects containing 476 frames) and RS-2DU test set (21 sequences / 1,260 frames) reported at the subject level.

Standard direct-mask predictions in Table 6, comparison with the classic baseline architectures (SegNet, U-Net, DeepLabV3), show that they reach a functional performance limit with ultrasound loops with low contrast and complex trabeculations, causing high degradation of the boundaries. In the proposed L-E-D CNN framework, the spatial errors both in HMC-QU and RS-2DU datasets are overcome by explicitly separating the regional feature learning into isolated inner and outer regions with post processing. This structural isolation ensures high boundary continuity over cardiac cycles with up to 0.95 (DSC) and an HD_{95} error of less than 3.15 mm.

In Table 7, a detailed comparison of the proposed L-E-D CNN framework with different learning strategies, anatomical position, and post-processing methods is presented. The conventional direct-mask training resulted in the worst performance (DSC $= 0.88 \pm 0.04$ on HMC-QU and 0.89 ± 0.03 on RS-2DU) and was associated with the highest boundary displacement errors ($HD_{95} = 5.20$ mm and 4.85 mm).

The use of the proposed inner–outer subtraction strategy significantly improved spatial overlap (DSC increased from 0.92 to 0.94) and also significantly reduced the boundary errors ($HD_{95} < 3.52$ mm). This is evidence that decoupling of predictions for inner and outer mask in the region prevents topological disconnections due to low contrast myocardial walls.

When papillary structures are added to the ground truth, the raw boundary distance errors (HD_{95}) for the HMC-QU and RS-2DU datasets increased by 0.11 mm and 0.15 mm respectively. In contrast, the overlap metrics (DSC= 0.92--0.94) were found to be stable for the L-E-D CNN even in the presence of complex intracavity structures.

Adding morphological post-processing resulted in the best segmentation results on all metrics. It further refined myocardial boundaries, driving peak regional overlap (DSC = 0.94 ± 0.02 on HMC-QU; $.95 \pm 0.02$ on RS-2DU) and minimizing boundary distance errors (HD_{95} = 3.15 ± 0.82 mm on HMC-QU; 2.84 ± 0.65 mm on RS-2DU). The overall results demonstrate the optimal pipeline configuration for LV wall segmentation in clinical echocardiograms that uses inner–outer regional subtraction and morphological post-processing.

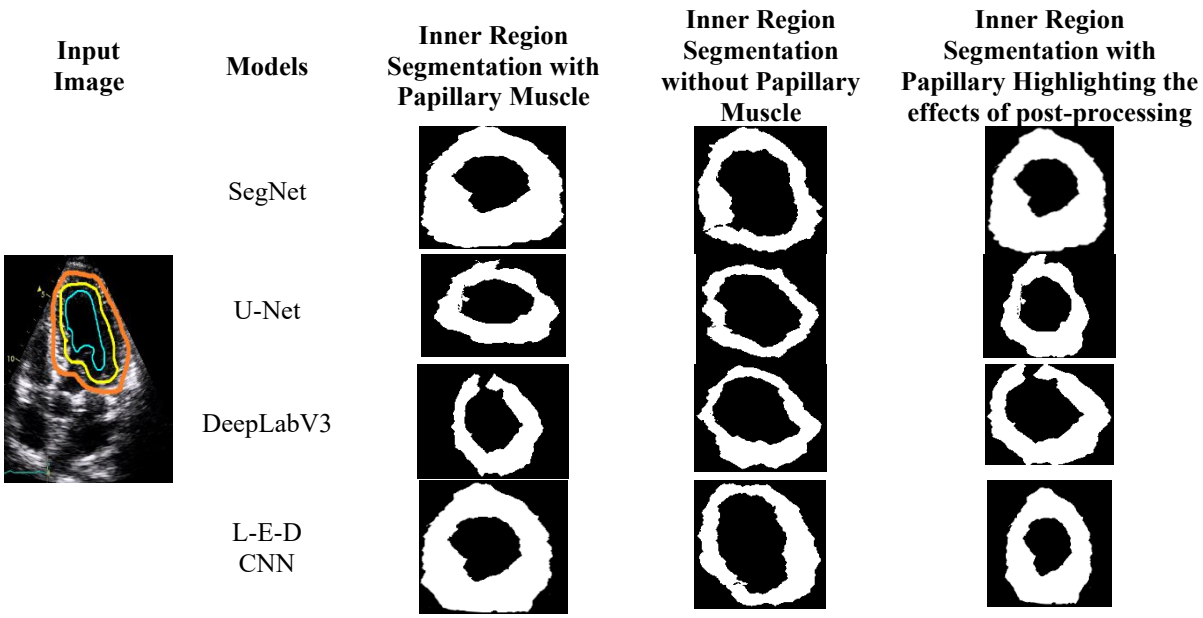

Figure 9. Segmentation results including papillary muscle and excluding papillary muscle for one example image. Orange - Segmentation of outer region boundary; Yellow – Inner region without Papillary Muscles; Blue – Inner region with Papillary Muscles.

Table 6. Comparative Performance Evaluation of LV Wall Segmentation

Model	DSC	IoU	BF1	HD_{95} , mm
HMC-QU dataset				
SegNet	0.84±0.05	0.73±0.06	0.79±0.05	7.24±1.85
U-Net	0.90±0.04	0.82±0.05	0.85±0.04	4.86±1.21
DeepLabV3	0.89±0.04	0.81±0.05	0.84±0.04	5.12±1.34
L-E-D CNN	0.94±0.02	0.89±0.03	0.91±0.03	3.15±0.82
RS-2DU dataset				
SegNet	0.86 ± 0.05	0.75±0.06	0.80±0.05	6.85±1.62
U-Net	0.91±0.03	0.84±0.04	0.87±0.04	4.21±1.05
DeepLabV3	0.91±0.04	0.83±0.04	0.86±0.04	4.48±1.18
L-E-D CNN	0.95±0.02	0.91±0.03	0.93±0.02	2.84±0.65

Table 7. Performance Evaluation of L-E-D CNN Configurations

Techniques / Measures	DSC	IoU	BF1	HD_{95} (mm)
HMC-QU Dataset				
Direct-mask training	0.88 ± 0.04	0.80 ± 0.05	0.83 ± 0.04	5.20 ± 1.10
Inner–outer subtraction (without papillary muscle)	0.93 ± 0.01	0.88 ± 0.03	0.90 ± 0.02	3.41 ± 0.88
Inner–outer subtraction (with papillary muscle)	0.92 ± 0.03	0.87 ± 0.04	0.89 ± 0.03	3.52 ± 0.96
Inner–outer subtraction (with papillary + post-processing)	0.94 ± 0.02	0.89 ± 0.03	0.91 ± 0.03	3.15 ± 0.82
RS-2DU Dataset				
Direct-mask training	0.89 ± 0.03	0.81 ± 0.04	0.84 ± 0.04	4.85 ± 1.02
Inner–outer subtraction (without papillary muscle)	0.94 ± 0.02	0.90 ± 0.02	0.92 ± 0.02	3.10 ± 0.72
Inner–outer subtraction (with papillary muscle)	0.94 ± 0.02	0.89 ± 0.03	0.91 ± 0.03	3.25 ± 0.80
Inner–outer subtraction (with papillary + post-processing)	0.95 ± 0.02	0.91 ± 0.03	0.93 ± 0.02	2.84 ± 0.65

Table 8. Component-Wise Ablation Study of L-E-D CNN model (HMC-QU Dataset)

Model Variant	Split-Block Design	Inner–Outer Subtraction	Post-Processing	DSC	IoU	BF1	HD_{95} (mm)
Baseline	×	×	×	0.85 ± 0.05	0.76 ± 0.06	0.79 ± 0.05	6.15 ± 1.50
+ Split-block	√	×	×	0.88 ± 0.04	0.80 ± 0.05	0.83 ± 0.04	5.20 ± 1.10
+ Subtraction strategy	√	√	×	0.92 ± 0.03	0.87 ± 0.04	0.89 ± 0.03	3.52 ± 0.96
+ Post-processing	√	√	√	0.94 ± 0.02	0.89 ± 0.03	0.91 ± 0.03	3.15 ± 0.82

The individual contribution of each key module in the proposed L-E-D CNN framework was systematically analysed on the HMC-QU dataset (Table 8) by focusing on the component-wise ablation study. Progressively testing performance starting from a standard CNN baseline to the full framework gives a few insights:

The performance improvement in the proposed Split-block architecture over the conventional convolutional blocks was observed right away with the DSC increasing from 0.85 ± 0.05 to 0.88 ± 0.04 and HD_{95} decreasing from 6.15 mm to 5.20 mm. This again supports that the split feature-extraction paths improve the multi-scale spatial representation of complex LV boundaries.

The significant performance improvement was obtained when using the inner–outer subtraction strategy which was integrated with split-block. DSC increased significantly by $+0.04$ (reaching 0.92 ± 0.03), while HD_{95} dropped sharply from 5.20 mm to 3.52 mm. It shows that by decoupling the inner and outer regional topology prediction there will be no boundary disconnections if there is any in direct-mask prediction.

With the addition of the final morphological post-processing pipeline (Full L-E-D CNN Model), the segmented contours were further sharpened, which led to the highest performance of the final DSC (0.94 ± 0.02), IoU (0.89 ± 0.03) and the lowest boundary error $HD_{95}(3.15 \pm 0.82$ mm). Overall, these findings validate the individual contribution of each of the proposed modules to optimal LV wall segmentation, and that it does so in a distinctive, complementary and statistically significant way.

4.4 Performance Analysis of MI Classification

4.4.1 Performance Analysis of MI Diagnosis on the HMC-QU Dataset

Confusion matrix for HMC-QU dataset (testing set) mentioned in Figure 10. This part calculates the LMIDepnet model's performance with baseline models, including the UNet+SVM, UNet+CNN, DeepLabV3+ResNet50, UNet+DenseNet121, SegNet+SNN on the HMC-QU Dataset.

Using the HMC-QU dataset, Figure 11 displays the ROC curves for the LMIDepnet and baseline approaches for MI detection. It is analyzed that the LMIDepnet model surpasses them in every metrics, all of which are at 0.96 for the HMC-QU respectively. This improvement stems from the model's ability to effectively balance parameter count with MI classification performance, enabling efficient MI detection.

On the HMC-QU dataset, Figure 12 shows how well baseline methods and LMIDepnet identify MI. Claiming to outperform all previous models in terms of accuracy, precision, recall, and F1-score, it presents the LMIDepnet model. The LMIDepnet's precision is comprehended 23.26%, 19.69%, 14.45%, 11.37%, and 6.93% higher to SegNet+SNN, UNet+SVM, UNet+CNN, UNet+DenseNet121, and DeepLabV3+ResNet50. The precision of the LMIDepnet is 19.86%, 14.7%, 10.56%, 7.19%, and 5.02% outperforms the other methods. The recall of the LMIDepnet model is 21.85%, 17.54%, 13.98%, 8.92%, and 4.89% surpasses the other techniques. The F1-score of the LMIDepnet model is 20.27%, 15.35%, 11.08%, 8.07%, and 4.33% greater than the same algorithms, correspondingly.

To assess the statistical significance of the proposed LMIDepnet a paired t-test was performed in the HMC-QU dataset. Using the p-value, values < 0.05 were considered to be statistically significant improvement. From Table 9, it is seen that the proposed LMIDepnet resulted in minimum p values compared with the baseline models, which indicates the statistical significance of the proposed model and its effectiveness in segmenting and classification of MI.

4.4.2 Performance Analysis of MI Diagnosis on the RS-2DU dataset

This part calculates the LMIDepnet model's performance with baseline models, including the UNet+SVM, UNet+CNN, DeepLabV3+ResNet50, UNet+DenseNet121, SegNet+SNN on the the RS-2DU dataset. Confusion matrix for RS-2DU dataset (testing set) is mentioned in Figure 13.

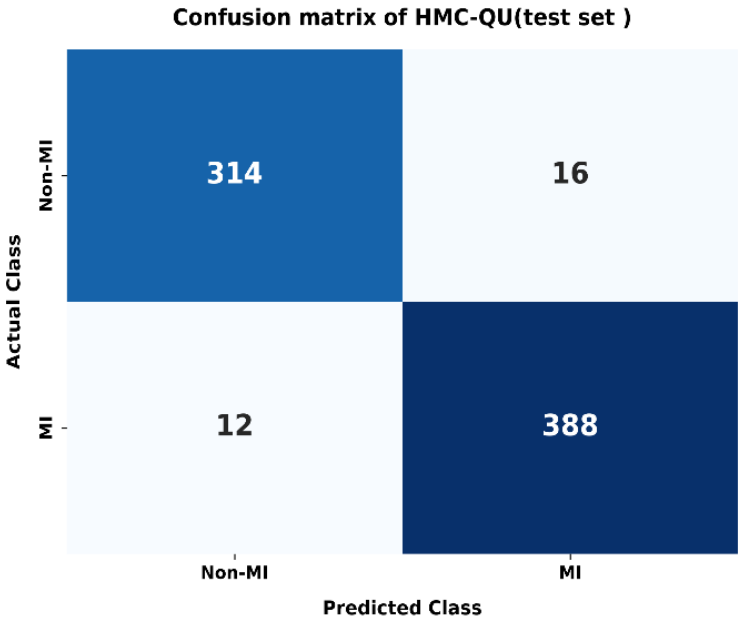

Figure 10. Confusion matrix for HMC-QU Dataset (testing set).

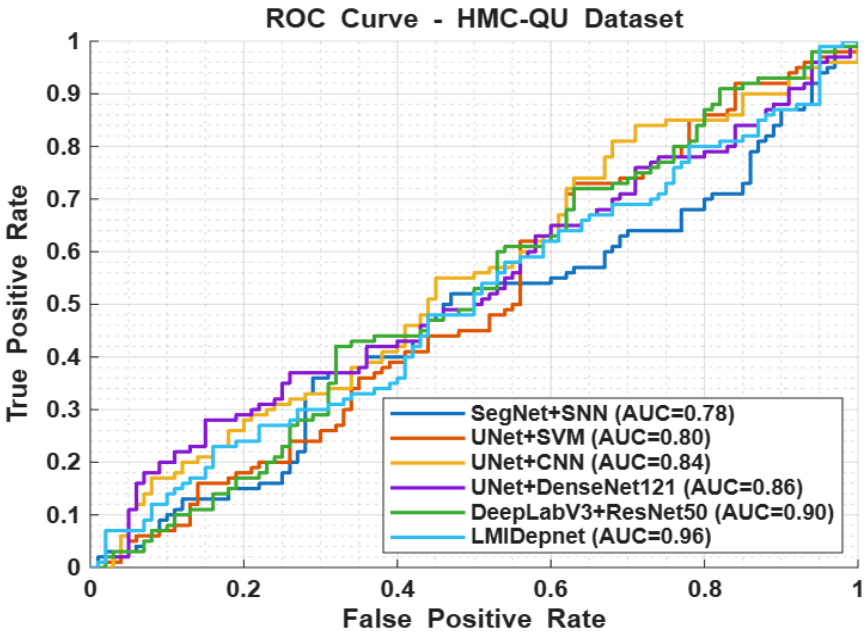

Figure 11. ROC curve of proposed LMIDepnet and baseline models on HMC-QU dataset.

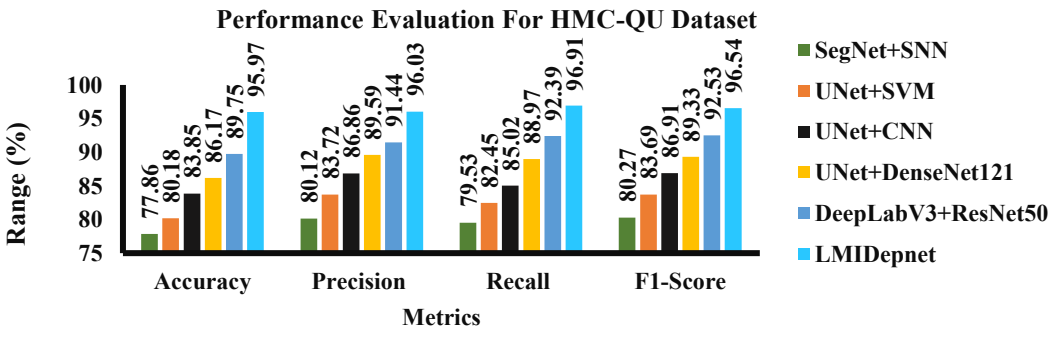

Figure 12. Comparison analysis of proposed LMIDepnet and baseline models on HMC-QU Dataset.

Table 9. Paired 5-fold cross-validation statistical comparison on Accuracy (%) and F1-Score (%) comparing LMIDepnet for HMC-QU Dataset

Baseline Model Compared	Baseline Mean (%)	Mean Difference (%)	Standard Error (SE)	t-Statistic (t)	p-Value	95% Confidence Interval (CI)
Accuracy (%)						
SegNet + SNN	77.86	18.11	0.27	67.07	< 0.001	[+17.36%, +18.86%]
UNet + SVM	80.18	15.79	0.408	38.7	< 0.001	[+14.66%, +16.92%]
UNet + CNN	83.85	12.12	0.356	34.04	< 0.001	[+11.13%, +13.11%]
UNet + DenseNet121	86.17	9.80	0.324	30.25	< 0.001	[+8.90%, +10.70%]
DeepLabV3 + ResNet50	89.75	6.22	0.217	28.66	< 0.001	[+5.62%, +6.82%]
F1-Score (%)						
SegNet + SNN	80.27	16.27	0.27	60.26	< 0.001	[+15.52%, +17.02%]
UNet + SVM	83.69	12.85	0.408	31.5	< 0.001	[+11.72%, +13.98%]
UNet + CNN	86.91	9.63	0.356	27.05	< 0.001	[+8.64%, +10.62%]
UNet + DenseNet121	89.33	7.21	0.324	22.25	< 0.001	[+6.31%, +8.11%]
DeepLabV3 + ResNet50	92.53	4.01	0.217	18.48	< 0.001	[+3.41%, +4.61%]

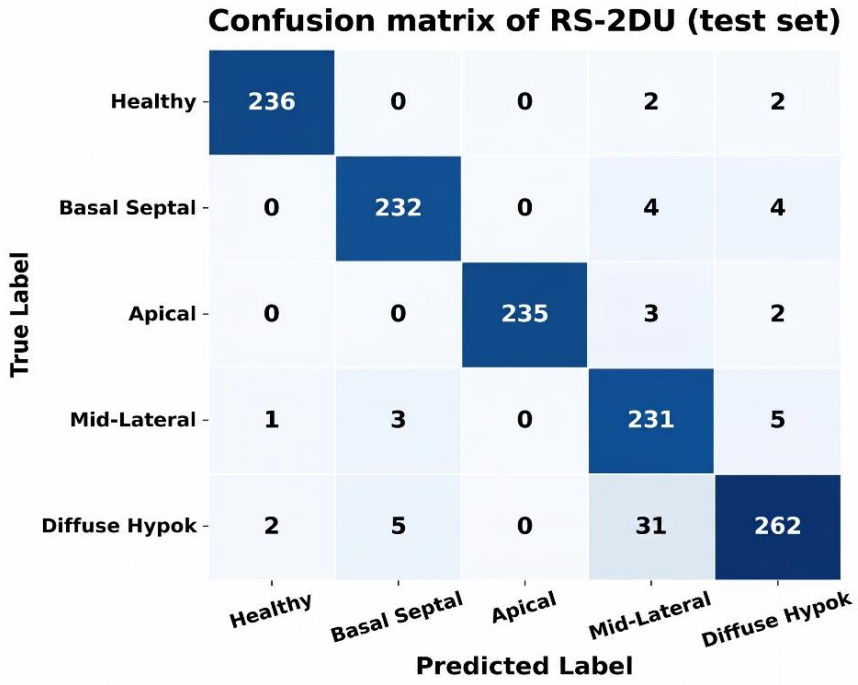

Figure 13. Confusion matrix for RS-2DU Dataset (testing set)

On the basis of the RS-2DU dataset, Figure 14 displays the ROC curves for the LMIDepnet and baseline approaches for MI detection. All of the metrics for the HMC-QU dataset show that the LMIDepnet model is better than them, with an average score of 0.95. This enhancement is a result of the model's skill in efficiently detecting MI by balancing the number of parameters with MI classification performance.

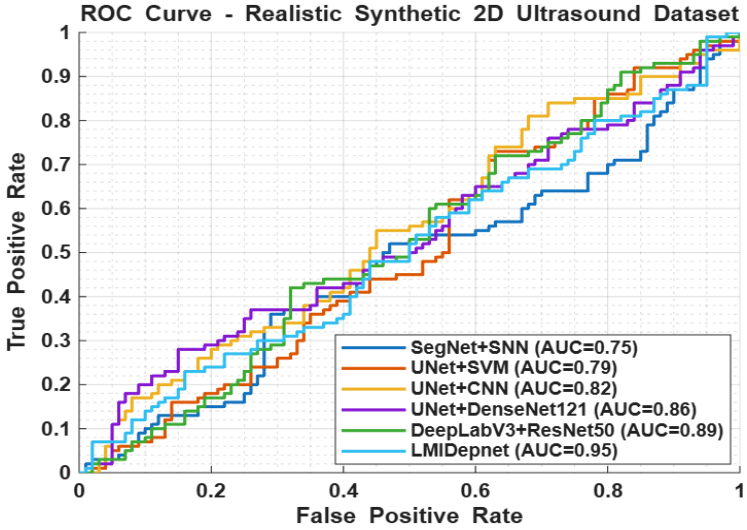

Figure 14. ROC curve of proposed LMIDepnet and baseline models on Realistic Synthetic 2D Ultrasound Dataset.

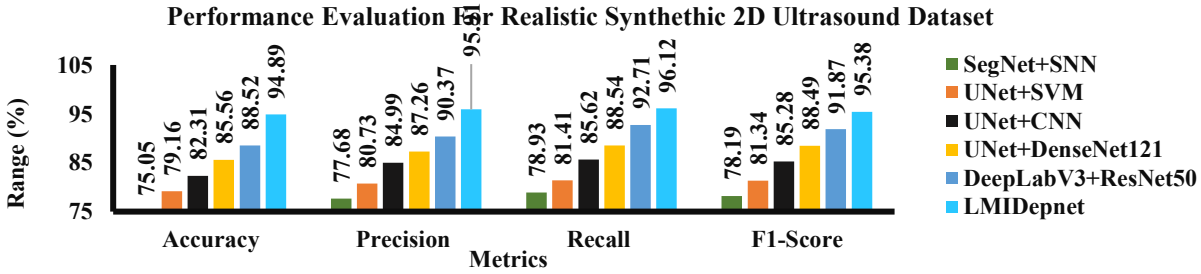

Figure 15. Performance analysis of proposed LMIDepnet and baseline models on Realistic Synthetic 2D Ultrasound Dataset.

Table 10. Paired 5-fold cross-validation statistical comparison on Accuracy (%) and F1-Score (%) comparing LMIDepnet for RS-2DU Dataset

Baseline Model Compared	Baseline Mean (%)	Mean Difference (%)	Standard Error (SE)	t-Statistic (t)	p-Value	95% Confidence Interval (CI)
Accuracy (%)						
SegNet + SNN	75.05	19.84	0.34	58.37	<0.001	[+18.89%, +20.78%]
UNet + SVM	79.16	15.73	0.178	88.49	<0.001	[+15.23%, +16.22%]
UNet + CNN	82.31	12.58	0.451	27.87	<0.001	[+11.32%, +13.83%]
UNet + DenseNet121	85.56	9.33	0.364	25.6	<0.001	[+8.32%, +10.34%]
DeepLabV3 + ResNet50	88.52	6.37	0.19	33.56	<0.001	[+5.84%, +6.89%]
F1-Score (%)						
SegNet + SNN	78.19	17.19	0.501	34.28	<0.001	[+15.80%, +18.58%]
UNet + SVM	81.34	14.04	0.124	113.1	<0.001	[+13.70%, +14.38%]
UNet + CNN	85.28	10.10	0.258	39.22	<0.001	[+9.38%, +10.82%]
UNet + DenseNet121	88.49	6.89	0.649	10.62	<0.001	[+5.09%, +8.69%]
DeepLabV3 + ResNet50	91.87	3.51	0.636	5.52	0.0052	[+1.75%, +5.27%]

Figure 15 presents the effectiveness of proposed LMIDepnet, and baseline models applied to RS-2DU Dataset for MI recognition.

It is analysed that the LMIDepnet model beats baseline models, achieving in every metrics. Specifically, the accuracy of LMIDepnet is 26.44%, 19.87%, 15.28%, 10.9%, and 7.19% higher than that of SegNet+SNN, UNet+SVM, UNet+CNN, UNet+DenseNet121, and DeepLabV3+ResNet50 respectively. Precision is also improved by 23.47%, 18.8%, 12.85%, 9.91%, and 6.13% compared to the same models.

Similarly, the recall of LMIDepnet surpasses these models by 21.78%, 18.07%, 12.26%, 8.56%, and 3.68%. In terms of F-measure, LMIDepnet achieves gains of 21.98%, 17.26%, 11.84%, 7.79%, and 3.82% over each of the baseline model.

A paired t-test was used for the RS-2DU data set to determine the statistical significance of the proposed LMIDepnet. Using the p-value, values < 0.05 were considered statistically significant. It is seen that for all the models, the proposed LMIDepnet model shows statistically significant and low p values as compared to other models as seen from Table 10 which is statistically significant and proves assessment of the proposed model's segmentation and classification of MI on RS-2DU dataset.

The class-wise distribution of the dataset namely RS-2DU, used for multi-view MI classification is presented in Table 11, which shows the original training samples, augmented samples, and the whole held out test set. Data was divided into an 80% training set (84 sequences, 5,040 frames), and a 20% testing set (21 sequences, 1,260 frames), to enhance the model's generalizability to new data, but not to the test set, which needed objective evaluation, thanks to augmentation performed just to the training set. Table 12 summarizes the results of the view-wise categorization on the RS-2DU test set.

4.5 Computational Efficiency and Operational Deployment Analysis

In order to assess the operational feasibility and real-time deployment potential of our proposed methodologies, an extensive computational efficiency benchmark was undertaken in both the execution environment (GPU and CPU-only modes). Table 13 compares the memory consumption, structural complexity, and per-frame latency of the standalone segmentation methods with the proposed L-E-D CNN and standard backbones (SegNet, U-Net, and DeepLabV3). In addition, Table 14 shows the computational time of the end-to-end MI classification models, comparing the full LMIDepnet pipeline with the standard hybrid classifiers.

Table 13 shows that L-E-D CNN has the lowest number of parameters, 13.4M, compared to all other baseline segmentation architectures, and it has the lowest number of FLOPs, 4.3 GFLOPs, as well as the smallest model size 52 MB, which is 21.6% and 28.3% reduction, respectively, then SegNet (17.1M parameters, 6.0 GFLOPs). Additionally, Table 9 shows that L-E-D CNN has the lowest run time memory footprint 1,954 MB GPU / 3,642 MB CPU) and the best per frame GPU inference latency (0.05 seconds / 20.0 FPS), which is suitable for lightweight clinical use.

Table 11. Multi view training and test split details of RS-2DU dataset

Apical View	Total Sequences	Training Set (80%) sequences	Frames	Held-Out Test Set Sequences (20%)	Test Frames
A4C (Apical 4-Chamber)	35	28	1680	7	420
A2C (Apical 2-Chamber)	35	28	1680	7	420
A3C (Apical 3-Chamber)	35	28	1680	7	420
Total	105	84	5040	21	1,260

Table 12. View-wise classification performance on the RS-2DU dataset

Apical View	Test Sequences/frames	Accuracy	Precision	Recall	F1-Score
A4C	7 / 420	95.24%	96.10%	96.30%	0.958
A2C	7 / 420	94.52%	95.50%	95.80%	0.951
A3C	7 / 420	94.88%	96.00%	96.10%	0.952
Combined	21 / 1,260	94.89%	95.91%	96.12%	0.954

Table 13. Computational Efficiency Evaluation of LV Segmentation Models

Model	Parameters (M)	FLOPs (G)	Model Size (MB)	Training Time (s) [CPU / GPU]	Inference Time (s) [CPU / GPU]	GPU Throughput (FPS)	Memory Usage (MB) [CPU / GPU]
SegNet	17.1	6	66	4138 / 1176	2.2 / 0.12	8.3	4215 / 2264
U-Net	15.6	5.3	60	3872 / 1058	1.8 / 0.09	11.1	3972 / 2116
Deep LabV3	14.7	4.9	57	3694 / 1012	1.7 / 0.08	12.5	3861 / 2068
L-E-D CNN	13.4	4.3	52	3365 / 901	1.3 / 0.05	20.0	3642 / 1954

Table 14. Computational Efficiency Evaluation of MI Classification Frameworks

Model	Parameters (M)	FLOPs (G)	Model Size (MB)	Training Time (s) [CPU / GPU]	Inference Time (s) [CPU / GPU]	GPU Throughput (FPS)	Memory Usage (MB) [CPU / GPU]
SegNet + SNN	22.4	8.1	88	5420 / 1580	2.8 / 0.16	6.3	4850 / 2620
UNet + SVM	18.2	6.2	72	4810 / 1390	2.3 / 0.12	8.3	4320 / 2310
UNet + CNN	24.5	8.5	96	5930 / 1720	2.9 / 0.15	6.7	5120 / 2780
UNet + DenseNet121	30.1	10.4	118	6840 / 1980	3.4 / 0.19	5.3	5980 / 3240
DeepLabV3 + ResNet50	38.2	12.8	150	7950 / 2310	4.1 / 0.24	4.2	6850 / 3710
LMIDepnet	17.8	5.9	70	4250 / 1180	1.8 / 0.08	12.5	4050 / 2180

As shown in Table 14, the proposed LMIDepnet framework significantly improves the efficiency of the computing resources: 17.8M parameters, 5.9 GFLOPs, and a storage space of 70 MB, compared with over a 53% reduction in the number of parameters and FLOPs of the conventional end-to-end MI classification frameworks using deep backbones, such as DeepLabV3+ResNet50 (38.2M parameters, 12.8 GFLOPs). Furthermore, Table 10 shows that LMIDepnet provides operation latency of 0.08 seconds per sequence (12.5 FPS) on GPU and 1.8 seconds per sequence on CPU, with a small memory footprint (2,180 MB GPU / 4,050 MB CPU), which is well suited for real-time diagnostic workflows.

5. Conclusion

The LMIDepnet framework was developed for optimal echocardiographic image analysis and improve the diagnostic accuracy for the evaluation of MI. Lightweight L-E-D CNN segmentation module was proposed with parameter reduction strategy and optimized regional training strategy to improve the accuracy of segmentation while decreasing computational burden, respectively. As input, this model uses images and passes it through two parallel channels of feature extraction to extract local structural details and global context simultaneously, followed by fusion of the two channels using pointwise convolutions, with factorized convolutions. The model processes an image via two parallel feature-extraction channels, applying simultaneously both local structural details and global context, which is then fused together using pointwise convolutions, with factorized convolutions. In addition, an inner–outer regional training approach is able to target specific cardiac structures to successfully reduce speckle noise and provide precise endocardial and epicardial delineation in the presence of indistinct boundaries, with the help of a morphological post-processing pipeline. As a consequence, the spatial representations extracted are highly tolerant to image quality changes and provide good segmentation performance on the HMC-QU benchmark and on the RS-2DU benchmark.

Based on these segmented feature maps, the entire LMIDepnet classification system uses the L-E-D CNN segmentation backbone, to learn spatial features

including temporal cardiac motion using a CNN-LSTM network hybrid, and a Softmax classifier to make final decisions. The proposed end-to-end framework is well suited for real-time clinical MI detection, as it is sufficiently informative for diagnosis and also efficient enough to be used for computational purposes. Experimental results show that LMIDepnet significantly outperforms the conventional baselines (SegNet+SNN, UNet+SVM, UNet+CNN, UNet+DenseNet121, and DeepLabV3+ResNet50) yields impressive 95.97% and 94.89% classification rates on the HMC-QU and RS-2DU datasets, respectively.

References

[1] D. Yan, S. Zhan, C. Guo, J. Han, L. Zhan, Q. Zhou, X. Wang, The Role of Myocardial Regeneration, Cardiomyocyte Apoptosis in Acute Myocardial Infarction: A Review of Current Research Trends and Challenges. Journal of Cardiology, 85(4), (2024) 283-292. https://doi.org/10.1016/j.jjcc.2024.09.012

[2] M.A. Matter, F. Paneni, P. Libby, S. Frantz, B.E. Stähli, C. Templin, A. Mengozzi, Y.J. Wang, T.M. Kündig, L. Räber, F. Ruschitzka, C.M. Matter, Inflammation in Acute Myocardial Infarction: the Good, the Bad and the Ugly. European Heart Journal, 45(2), (2024) 89–103. https://doi.org/10.1093/eurheartj/ehad486

[3] M. Milosevic, Q. Jin, A. Singh, S. Amal, Applications of AI in Multi-Modal Imaging for Cardiovascular Disease. Frontiers in Radiology, 3, (2024) 1294068. https://doi.org/10.3389/fradi.2023.1294068

[4] E.G. Akramova, E.V. Vlasova, A.A. Saveliev, E.B. Zakirova, Information Value of Echocardiography in Inferior Myocardial Infarction at Different Stages of Observation. Journal of Clinical Practice, 15(1), (2024) 17–25. https://clinpractice.ru/clinpractice/article/view/580663

[5] C. Mitchell, P.S. Rahko, L.A. Blauwet, B. Canaday, J.A. Finstuen, M.C. Foster, K. Horton, K.O. Ogunyankin, R.A. Palma, E.J. Velazquez, Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography. Journal of the American Society of Echocardiography, 32(1), (2019) 1–64. https://doi.org/10.1016/j.echo.2018.06.004

[6] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C.P. Langlotz, P.A. Heidenreich, R.A. Harrington, D.H. Liang, E.A. Ashley, J.Y. Zou, Video-Based AI for Beat-to-Beat Assessment of Cardiac Function. Nature, 580(7802), (2020) 252–256. https://doi.org/10.1038/s41586-020-2145-8

[7] A. Madani, R. Arnaout, M. Mofrad, R. Arnaout, Fast and Accurate View Classification of Echocardiograms using Deep Learning. NPJ Digital Medicine, 1, (2018) 6. https://doi.org/10.1038/s41746-017-0013-1

[8] J. Zhang, S. Gajjala, P. Agrawal, G.H. Tison, L.A. Hallock, L. Beussink-Nelson, M.H. Lassen, E. Fan, M.A. Aras, C. Jordan, K.E. Fleischmann, M. Melisko, A. Qasim, S.J. Shah, R. Bajcsy, R.C. Deo, Fully Automated Echocardiogram Interpretation in Clinical Practice: Feasibility and Diagnostic Accuracy. Circulation, 138(16), (2018) 1623–1635. https://doi.org/10.1161/CIRCULATIONAHA.118.034338

[9] A. Degerli, M. Zabihi, S. Kiranyaz, T. Hamid, R. Mazhar, R. Hamila, M. Gabbouj, Early Detection of Myocardial Infarction in Low-Quality Echocardiography. IEEE Access, 9, (2021) 34442–34453. https://doi.org/10.1109/ACCESS.2021.3059595

[10] C. Tiago, S.R. Snare, J. Šprem, K. McLeod, A Domain Translation Framework with an Adversarial Denoising Diffusion Model to Generate Synthetic Datasets of Echocardiography Images. IEEE Access, IEEE, 11, (2023) 17594–17602. https://doi.org/10.1109/ACCESS.2023.3246762

[11] B.A. Valanrani, S.D. Suganya, Improving Myocardial Infarction Detection from Echo Images using Contrastive Guided Adversarial Denoising Diffusion Probabilistic Model. International Journal of Engineering Trends and Technology, 72(12), (2024) 285–297. https://doi.org/10.14445/22315381/IJETT-V72I12P124

[12] Y. Deng, P. Cai, L. Zhang, X. Cao, Y. Chen, S. Jiang, Z. Zhuang, B. Wang, Myocardial Strain Analysis of Echocardiography Based on Deep Learning. Frontiers in Cardiovascular Medicine, 9, (2022) 1067760. https://doi.org/10.3389/fcvm.2022.1067760

[13] X. Lin, F. Yang, Y. Chen, X. Chen, W. Wang, X. Chen, Q. Wang, L. Zhang, H. Guo, B. Liu, L. Yu, K. He, Echocardiography-Based AI Detection of Regional Wall Motion Abnormalities and Quantification of Cardiac Function in Myocardial Infarction. Frontiers in Cardiovascular Medicine, 9, (2022) 903660. https://doi.org/10.3389/fcvm.2022.903660

[14] O. Hamila, S. Ramanna, C.J. Henry, S. Kiranyaz, R. Hamila, R. Mazhar, T. Hamid, Fully Automated 2D and 3D Convolutional Neural Networks Pipeline for Video Segmentation and Myocardial Infarction Detection in Echocardiography. Multimedia Tools and Applications, 81(26), (2022) 37417–37439. https://doi.org/10.1007/s11042-021-11579-4

[15] G. Zamzmi, S. Rajaraman, L.Y. Hsu, V. Sachdev, S. Antani, Real-Time Echocardiography Image

Analysis and Quantification of Cardiac Indices. Medical Image Analysis, 80, (2022) 102438. https://doi.org/10.1016/j.media.2022.102438

[16] E. Evain, Y. Sun, K. Faraz, D. Garcia, E. Saloux, B.L. Gerber, M. Craene, O. Bernard, Motion Estimation by Deep Learning in 2D Echocardiography: Synthetic Dataset and Validation. IEEE Transactions on Medical Imaging, IEEE, 41(8), (2022) 1911–1924. https://doi.org/10.1109/TMI.2022.3151606

[17] C. Balakrishnan, V.D. Ambeth Kumar, IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks. Diagnostics, 13(4), (2023) 775. https://doi.org/10.3390/diagnostics13040775

[18] T. Nguyen, P. Nguyen, D. Tran, H. Pham, Q. Nguyen, T. Le, H. Van, B. Do, P. Tran, V. Le, T. Nguyen, H. Pham, Ensemble Learning of Myocardial Displacements for Myocardial Infarction Detection in Echocardiography. Frontiers in Cardiovascular Medicine, 10, (2023) 1185172. https://doi.org/10.3389/fcvm.2023.1185172

[19] S. Deepika, N. Jaisankar, Detecting and classifying myocardial infarction in Echocardiogram Frames with an Enhanced CNN Algorithm and ECV-3D Network. IEEE Access, IEEE, 12, (2024) 51690-51703. https://doi.org/10.1109/ACCESS.2024.3385787

[20] G. Holste, E.K. Oikonomou, B.J. Mortazavi, Z. Wang, R. Khera, Efficient Deep Learning-Based Automated Diagnosis from Echocardiography with Contrastive Self-Supervised Learning. Communications Medicine, 4(1), (2024) 133. https://doi.org/10.1038/s43856-024-00538-3

[21] A.D. Jamthikar, Q.A. Hathaway, K. Maganti, Y. Hamirani, S. Bokhari, N. Yanamala, P.P. Sengupta, Ultrasonic Texture Analysis for Predicting Acute Myocardial Infarction. JACC: Cardiovascular Imaging, 18(11), (2025) 1185–1199. https://doi.org/10.1016/j.jcmg.2025.06.018

[22] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, L.J. Latecki, (2019) LedNet: A Lightweight Encoder-Decoder Network for Real-Time Semantic Segmentation. In 2019 IEEE International Conference on Image Processing (ICIP), IEEE, Taipei, Taiwan. https://doi.org/10.1109/ICIP.2019.8803154

[23] M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang, (2018). DenseASPP for Semantic Segmentation in Street Scenes. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, USA. https://doi.org/10.1109/CVPR.2018.00388

[24] R. Muraki, A. Teramoto, K. Sugimoto, K. Sugimoto, A. Yamada, E. Watanabe, Automated Detection Scheme for Acute Myocardial Infarction using Convolutional Neural Network and Long Short-Term Memory. PLOS ONE, 17(2), (2022) e0264002. https://doi.org/10.1371/journal.pone.0264002

[25] A. Degerli, HMC-QU echocardiography dataset, Kaggle Dataset, 2021. Available at: https://www.kaggle.com/datasets/aysendegerli/hmcqu-dataset (accessed 15 September 2026).

[26] M. Alessandrini, B. Chakraborty, B. Heyde, O. Bernard, M. De Craene, M. Sermesant, J. D'Hooge, 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, 65(3), (2017) 411–422. https://doi.org/10.1109/TUFFC.2017.2786300

Authors Contribution Statement

B. Arockia Valanrani: Conceptualization, Methodology, Software, Validation, Resources, Data Curation, Writing—Original Draft Preparation, Writing—Review and Editing. S. Devi Suganya: Formal Analysis, Investigation, Visualization, Supervision. Both authors read and agreed to the published version of the manuscript

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes