International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

# Deep Learning for Myocardial Infarction Detection Using Echocardiography: A Comprehensive Survey of Segmentation and Classification Approaches

**A. Shobana[1], K. P. Malarkodi[2]**

[1]Research Scholar, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-8, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-8, Tamil Nadu, India
Corresponding Author Email: *malarkodiphd2017[at]gmail.com*

**Abstract:** ***Myocardial infarction (MI) arises when blood circulation to a segment of cardiac muscle is suddenly obstructed, resulting in tissue ischemia and irreversible myocardial damage. Imaging techniques, including Electrocardiography (ECG), Echocardiography, Coronary Angiography, and Computed Tomography (CT) play a vital role in assessing cardiac structure and function for MI diagnosis. Among these, echocardiography is widely used due to its non-invasive nature and ability to evaluate Regional Wall Motion Abnormalities (RWMA) and ventricular performance. However, manual evaluation of these images is labor-intensive and susceptible to inter-observer fluctuation, especially in detecting subtle wall motion changes. To overcome these challenges, Machine Learning (ML) techniques were initially introduced to support MI detection through automated segmentation and classification. But ML approaches relied heavily on manual feature engineering and often struggled with high-dimensional and complex patterns in echocardiographic sequences. In the last few years, Deep Learning (DL) techniques have surfaced as a formidable alternative which can autonomously learning classification characteristics using raw data. This survey reviews existing DL-based approaches for MI detection using echocardiography, thereby analyzing their methodologies, datasets, and performance metrics, and provides a comparative evaluation to identify the most reliable and clinically applicable frameworks.***

**Keywords:** Myocardial infarction, Deep Learning, Echocardiography, Classification, Segmentation

## 1. Introduction

MI, generally known as a heart attack, is a severe cardiovascular disorder resulting from the abrupt blockage of blood supply to a segment of the cardiac muscles, causing tissue ischemia and eventual apoptosis. It occurs primarily due to Coronary Artery Disease (CAD), where lipomas lipids, and immune cells accumulate along the arterial walls, forming atherosclerotic plaques. When these plaques rupture, blood clots form and block the coronary artery, depriving the myocardium of oxygen and nutrients. If not treated promptly, this results in irreversible myocardial damage and life-threatening complications [1]. Fig. 1 shows the pathophysiological mechanism of myocardial infarction caused by coronary artery blockage.

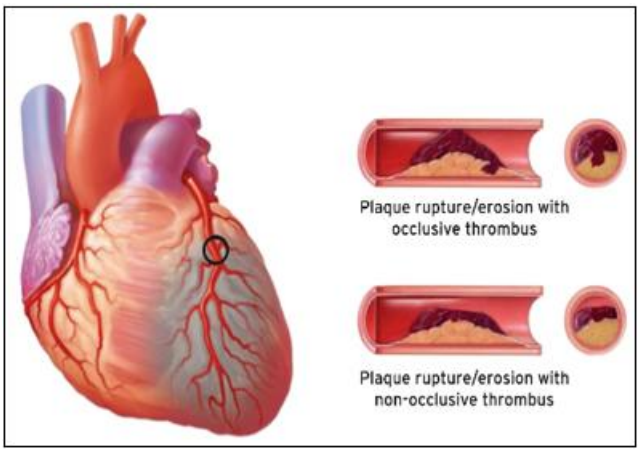


**Figure 1:** Visual representation of myocardial infarction [1]

Common symptoms of MI include persistent chest discomfort or pressure, pain extending to the left arm, neck, or jaw, dyspnea of breath, nausea, dizziness and excessive sweating. However, atypical or "silent" MIs are also observed, particularly in elderly individuals, women, and diabetic patients, making early diagnosis challenging [2,3]. Clinically, MI is classified as ST-elevation MI (STEMI) as well as Non-ST-elevation MI (NSTEMI), determined by the degree of coronary obstruction and ECG findings [4]. Despite advances in healthcare, MI remains to be a primary factor of global death, particularly in populations with sedentary lifestyles, unhealthy diets, smoking habits, and limited access to preventive healthcare [5].

Early predictions and timely diagnosis of MI are critical to reducing mortality, minimizing cardiac damage, and improving long-term patient survival. Rapid intervention can restore blood flow and preserve heart function, thereby reducing the burden of cardiovascular disease. Several clinical tests are routinely used for MI diagnosis:

- Electrocardiogram (ECG) – Monitors cardiac electrical activity and identifies irregularities such as ST-segment elevation, T-wave inversion, or arrhythmias. It is the primary and fastest diagnostic tool [4]. Fig. 2 illustrates the sample image of electrocardiography.
- Echocardiography – Uses ultrasound waves to visualize heart structure and function, helping to detect wall motion abnormalities and reduced pumping capacity [6]. Fig. 3 displays the sample image of echocardiography.
- Coronary Angiography – Imaging technique that visualizes coronary arteries to identify blockages and guide interventional procedures [7]. Fig. 4 displays the sample image of coronary angiography.

- CT Scans – Provide detailed structural and functional imaging of the heart, assessing infarct size, tissue viability, and complications [8]. Fig. 5 shows a representative image of CT scans.

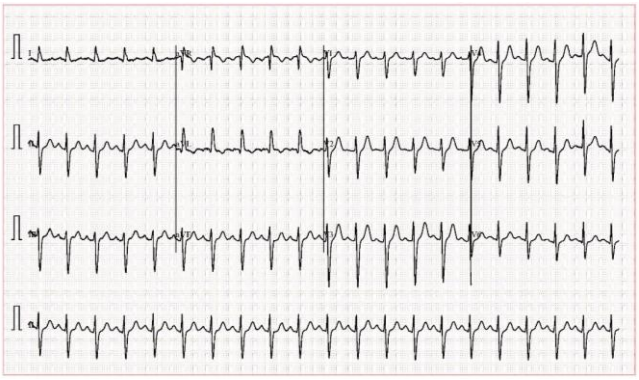

**Figure 2:** Sample image of electrocardiography [4]

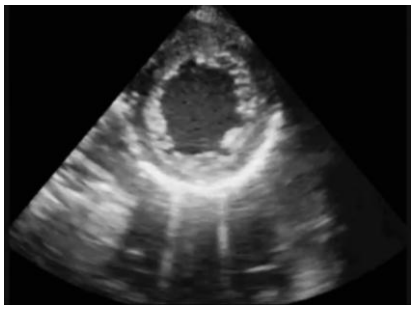

**Figure 3:** Sample image of echocardiography [6]

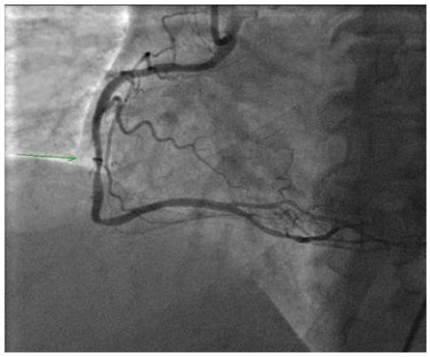

**Figure 4:** Sample image of coronary angiography [7]

**Figure 5:** Sample image of CT scans [8]

Among these modalities, echocardiography is considered one of the most valuable tools for detecting acute MI. It enables rapid assessment, visualization of cardiac structure and function, identification of RWMA, evaluation of ventricular performance including left ventricular ejection fraction (LVEF), risk stratification, and early detection of life-threatening complications such as ventricular rupture, thrombus formation, or cardiogenic shock [6, 9]. Because it is non-invasive, radiation-free, and widely available, echocardiography plays a central role in managing critically ill cardiac patients.

However, despite its clinical importance, echocardiographic assessment has several inherent limitations. Image quality is highly dependent on operator expertise and patient-specific factors such as acoustic window availability [9]. Manual analysis of echocardiographic images is labor- intensive along with subject to inter-observer variation and diagnostic bias. Subtle wall motion abnormalities may be missed, particularly in early-stage MI or in cases with complex cardiac motion patterns. Furthermore, accurate segmentation of cardiac chambers and quantification of ejection fraction require skilled cardiologists and advanced equipment, which may not always be available in resource-limited settings.

To overcome these limitations, Artificial Intelligence (AI) has grown up as an effective tool to augment precision, consistency, and efficiency of echocardiographic analysis [10]. AI encompasses computational techniques designed to mimic human cognitive abilities, including pattern recognition and decision-making from complex medical data. In cardiac imaging, AI systems assist clinicians in automating image interpretation, detecting abnormalities, and supporting clinical decisions. Within AI, Machine Learning (ML) and Deep Learning (DL) are the two primary subfields applied to echocardiographic MI detection.

ML methods operate by training algorithms to recognize patterns in structured features extracted from images. In echocardiography, ML models analyze parameters such as ventricular dimensions, motion vectors, texture descriptors, and Doppler measurements to classify cardiac function or detect MI-related abnormalities. Common ML methodologies include Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (KNN). Although ML approaches have demonstrated effectiveness, they rely heavily on manual feature engineering and expert-defined parameters [11]. Feature selection bias and poor generalization across different imaging devices and clinical settings further limit traditional ML models. This dependence limits scalability and may lead to performance degradation when handling high-dimensional imaging data or subtle pathological changes.

To address these drawbacks, researchers increasingly employ DL, a subset of ML that automatically learns hierarchical feature representations directly from raw imaging data. DL architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) models can record spatial and temporal cardiac motion patterns without explicit feature design [12, 13]. These models can directly learn myocardial motion patterns and spatiotemporal dynamics from echo sequences. Their ability to process large-scale imaging datasets improves diagnostic accuracy, robustness, and reproducibility compared to conventional methods.

This survey therefore focuses on systematically reviewing and critically analysing existing DL-based approaches for MI

detection using echocardiographic imaging. The objective is to examine the architecture, datasets, evaluation metrics, and clinical applicability. A comparative analysis is presented to highlight strengths and limitations. The subsequent sections of this work are structured as follows: Section 2 provides a comprehensive analysis AI-based models for MI detection using echocardiography. Section 3 presents an in-depth examination of methodologies, datasets, and performance measures. Section 4 evaluates those models in terms of accuracy on benchmark datasets. Ultimately, Section 5 summarizes the survey findings and discusses future study potential opportunities in intelligent cardiac diagnosis.

## 2. Review of Dl Models for Acute Myocardial Infarction Detection

Degerli et al. [14] presented a three-stage pipeline to prompt MI diagnosis from inadequate echocardiography by combining segmentation with motion-driven feature engineering and classification. First, the complete LV wall was delineated in all frames using an encoder–decoder CNN. Secondly, attribute engineering was conducted on the segmented LV wall. For each segment, three motion-related descriptors were computed over one cardiac cycle: (1) endocardial border displacement, (2) segment center-of-mass displacement, and (3) segment region change. Finally, in the MI detection stage, these engineered features were input into classifiers like Decision Tree (DT), Linear Discriminant Analysis (LDA), SVM as well as RF.

Li et al. [15] devised a myocardial segmentation framework for MCE sequences. An encoder–decoder architecture was employed, where a U-Net–based encoder extracted multi-resolution spatial features from MCE frames, and a hierarchical ConvLSTM decoder integrated temporal information across consecutive frames. The encoder generated feature maps at multiple scales to capture both global and fine myocardial details, which were then passed to the ConvLSTM layers that incorporated current frame features, previous layer outputs, and hidden states from prior frames. To further reduce error propagation caused by poor-quality frames, a bi-directional training strategy was introduced, in which the network was trained sequentially in both upward and downward downward orientations to enforce temporal robustness. The segmented myocardium was subsequently used for perfusion quantification by generating time–intensity replenishment curves and fitting them to an exponential model to estimate parameters such as plateau intensity, microbubble velocity, and myocardial blood flow.

Deng et al. [16] developed a DL based 3D-CSN model for automated myocardial segmentation from echocardiographic videos. The model provided frame-level cardiac segmentation, where consecutive frames were processed to gather both spatial and temporal information. This network generated segmentation masks to describe the Region of Interest (ROI) and extract the LV myocardium (LVM) to establish a reference coordinate system. Next, an optical flow–based motion estimation network was employed to estimate pixel-wise myocardial motion between consecutive frames. Finally, the updated centreline positions were used to calculate arc length changes throughout the cardiac cycle, from which Global Longitudinal Strain (GLS) and Regional Longitudinal Strain (RLS) were derived.

Hamila et al. [17] devised a two-stage DL framework for early MI detection from echocardiography videos. First, a 2D CNN segmentation network with an encoder–decoder structure was trained using manually created masks that delineated the left ventricle (LV) from the apical four-chamber (A4C) perspective. As video frames varied in size, a spatial sliding window approach was applied to standardize inputs and increase the number of training samples. The network predicted pixel-wise probability masks, which were reconstructed into full-frame masks using an inverse sliding-window process and then refined using a minimum bounding box to isolate the LV region. Each original frame was multiplied by this bounding box to generate segmented frames, and these frames were reassembled into LV-focused videos. In the second stage, the segmented videos were resized to a uniform resolution and divided into temporal windows of 5, 7, or 9 frames using a sliding-window technique with a one-frame stride, creating spatiotemporal clips. These clips were fed into a 3D CNN, which learned motion dynamics and structural variations of the LV across cardiac cycles. The final sigmoid output neuron performed binary MI classification, where the 3D convolutions captured wall-motion abnormalities and temporal cardiac patterns associated with MI.

Laumer et al. [18] developed a comprehensive computer vision process to differentiate takotsubo syndrome (TTS) with AMI alongside unprocessed echocardiographic videos. First, each video frame was semantically segmented using a CNN to label cardiac structures such as LV myocardium and chambers. Second, view-specific convolutional autoencoders were trained to reconstruct these segmentation masks. The encoder parts of these autoencoders generated low-dimensional latent features for each frame, producing multivariate temporal sequences that captured myocardial motion dynamics. Third, the latent feature sequences were processed using a temporal CNN. Multiple sub sequences were sampled and resampled to a fixed length, passed through temporal encoders for each view, concatenated, and fed to dense layers to output probabilities for TTS vs AMI.

Lin et al. [19 introduced a comprehensive, end-to-end DL system for automatic detection of RWMAs and identification heart rhythm assessment from echocardiography. First, a view classification model (XceptionNet) automatically identified the three apical views (A4C, A2C, ALX) from DICOM videos, followed by an image quality assessment model that excluded frames where ventricular borders were not clearly visible. Next, a segmentation stage used separate LSTM-U-Net models for each view to delineate endocardial and epicardial edges of the LV and segment the myocardium into three anatomical regions (A, F, L) according to ASE guidelines. The segmented masks were concatenated with original video frames to guide downstream analysis. For RWMA detection, the combined video–mask inputs were processed using a 3D CNN based on the R(2+1)D architecture, which factorizes spatiotemporal convolutions to efficiently learn motion features.

Muraki et al. [20] developed an automated method to detect acute MI from echocardiography employing a CNN– Long Short-Term Memory (LSTM) architecture. Echocardiographic cine loops were collected in two views: the left ventricular long-axis and the short-axis papillary muscle level. During preprocessing, the ECG trace embedded in the images was removed, frames were cropped, and one complete cardiac cycle per patient was extracted using R–R interval timing. Each cycle was interpolated to 30 frames using linear interpolation to standardize temporal length. For feature extraction, each of the 30 frames was passed through a pre-trained VGG16 model. To capture temporal dynamics, the feature sequences were fed into a LSTM network, which then analysed sequential variations in ventricular wall motion and output a probability of AMI vs. normal myocardium through a sigmoid classifier.

Adalioglu et al. [21] presented a multi-view MI detection framework called Self-Attention Fusion Network (SAF-Net) that fused information from A2C and A4C echocardiography. For each recording, three reference frames (first, middle, last) from one cardiac cycle were extracted per view. These frames were fed into three pretrained CNN backbones such as, Inception-v3, ResNet-50 and DenseNet-121 to obtain complementary high-level spatial features. The resulting feature vectors from the three networks were concatenated into a 5120-dimensional representation per view, and features from both views formed a 5120×2 feature matrix. SAF-Net then performed feature embedding using a fully connected layer to reduce dimensionality and project features into a latent space (64×2). A self-attention mechanism was applied to this latent representation to model interdependencies between the two views. Finally, the fused feature vector passed through dense layers with sigmoid activation for binary MI classification.

Liu et al. [22] developed a DL-based AIEchoDx for automated classification of echocardiographic videos into four cardiovascular conditions like atrial septal problem, dilation of the heart, cardiac hypertrophy, and previous MI along with typical cases. Initially, meticulously documented echo videos and associated clinical files were collected. A pretrained Inception-V3 system was retrained to obtain spatial characteristic vectors across individual echo frames, whose features were combined to form temporal feature matrices. These matrices were then processed using a multi-layer 1DCNN thereby learning dynamic cardiac motion patterns and perform disease classification. In addition, echocardiographic features were integrated with electronic medical record data to further stratify dilated cardiomyopathy patients into clinically distinct phenogroups.

Nguyen et al. [23] introduced an ensemble learning framework for MI detection that integrates motion features derived from multiple LV segmentation models to improve robustness and diagnostic accuracy. The pipeline had trio phases: LV wall segmentation, motion-based feature engineering, and MI classification. In the first stage, several U-Net–style CNNs were employed to segment the LV wall from A4C echocardiography. In the second stage, the endocardial boundary was divided into six functional segments, and segment-wise myocardial displacement was computed using L1 distances between sampled boundary points in reference and subsequent frames. To combine complementary strengths of different segmentation models without increasing feature dimensionality, a performance-based weighting scheme was applied, where feature vectors from each model were weighted and summed into a single compact representation. In the final stage, classifiers such as Logistic Regression (LR), SVM, DT, and KNN were trained for detection.

Deepika and Jaisankar [24] devised a multi-modal DL framework for MI detection using enhanced CNN algorithm and an ECV-3D network. It integrated echocardiogram images, ECG signals, and clinical patient data into a unified diagnostic system. Initially, echocardiographic frames underwent image pre-processing and augmentation to enhance structural and textural cardiac features, while ECG signals were processed to extract meaningful electrical activity patterns and clinical data were normalized and encoded into structured vectors. An enhanced CNN, incorporating transfer learning and attention mechanisms, was used to extract high-level spatial features from the echo images. These image features were then fused with ECG and clinical features and fed into the ECV-3D network to learn complex spatiotemporal and cross-modal relationships.

Degerli et al. [25] suggested a multi-view MI detection framework that fused A4C and A2C echocardiography to capture RWMA across 12 myocardial segments. Initially, the endocardial boundary of the LV was separated using Active Polynomials (APs) to achieve smooth and noise-resistant boundary tracking. The extracted boundary was then divided into standardized myocardial segments, and segment-wise displacement curves were computed over one cardiac cycle by tracking multiple boundary points. From these curves, maximum normalized displacement features were derived for each segment to quantitatively represent wall motion. Feature vectors from A4C and A2C views were then concatenated to form a unified multi-view representation. Finally, these motion features were fed into classifiers including SVM, KNN, DT, RF, and a compact 1DCNN to perform binary MI detection.

Kasim et al. [26] suggested a multi-view, motion-driven framework for detecting RWMA from echocardiography by combining segmentation, motion estimation, deep temporal modeling, and classical classifiers. First, A2C and A4C echo loops were processed using a pre-trained U-Net to segment cardiac structures and isolate the LV myocardium. The LV boundary was tracked across frames to preserve myocardial structure. Next, dense optical flow was applied to consecutive segmented frames to compute pixel-wise motion fields, yielding horizontal (u) and vertical (v) displacement components that quantified myocardial wall motion. These optical flow stacks were then fed into a Temporal Convolutional Network (Temporal ConvNet) based on a ResNet-101 backbone and influenced by the Temporal Segment Network (TSN) framework. Features from A2C and A4C views were concatenated into a 2048-dimensional representation capturing complementary motion information. Finally, this feature vector was classified using several classifiers like SVM, RF, DT, KNN, and MLP).

Soe and Iwata [27] developed an MI detection system that primarily leveraged LV wall motion dynamics as the core diagnostic feature within an LSTM-based DL framework. Expert-annotated LV wall motion served as ground truth, and three complementary feature groups including image pixel data, LV dimensional changes, and wall motion displacement were extracted through preprocessing and segmentation of echocardiographic images. These temporal features were fed into an LSTM network, which modeled sequential myocardial motion patterns across the cardiac cycle to capture abnormalities associated with MI. The LSTM output was subsequently passed to a SVM classifier for final binary decision-making (MI vs non-MI).

Tabuco & Naval [28] devised an MI detection framework designed for low-quality 2D echocardiograms. Initially, echocardiographic frames underwent image enhancement using Gaussian blurring, gray-level slicing, and Otsu thresholding to suppress noise and standardize intensity variations. The enhanced frames were then fed into a U-Net++ segmentation network, which utilized dense skip connections to accurately delineate the LV wall. The predicted LV masks have been used to the actual photographs to obtain separated myocardial regions. From these regions, motion-based handcrafted features were extracted using Harris corner detection with homography matching to track key points across frames and compute maximum segment displacement. In addition, Histogram of Oriented Gradients (HOG) descriptors were derived from uniformly sampled segmented frames to capture structural and motion information. Meanwhile, a pretrained ResNeSt50 model with extensive data augmentation was employed to learn robust deep spatial features despite limited and noisy data. Finally, the handcrafted motion descriptors and CNN-learned representations were used for segment-level classification to identify infarcted and normal LV segments, enabling automated MI detection.

Valanrani and Suganya [29] introduced a DL–based framework to improve MI detection from echocardiography. It employed a hybrid generative approach combining a diffusion probabilistic model with a generative adversarial network (GAN) to synthesize diverse and realistic echocardiographic images, thereby enlarging the training dataset. The generated and original images were then processed using an encoder–decoder CNN to automatically segment the LV wall in every frames. From the segmented masks, endocardial boundary was extracted and divided into six anatomical segments, and motion-related features such as segment displacement, segment area variation, and maximum boundary movement over the cardiac cycle were calculated to capture RWMA. These clinically relevant motion features were subsequently used as inputs to a CNN classifier, which learned to distinguish among MI and non-MI cases.

Gomez et al. [30] introduced a cascade framework for automated MI diagnosing and localizing from echocardiography by combining DL segmentation with ML classification. The pipeline consisted of three stages. First, LV wall segmentation was performed utilizing U-Net with an EfficientNet-B4 backbone, trained from scratch on preprocessed frames. The segmented LV wall was then divided into anatomical segments to isolate the six motion-relevant regions. Second, feature extraction translated segmentation outputs into clinically interpretable motion descriptors. Global features included estimated LVEF (derived from area changes using an area–volume approximation) and global longitudinal strain (GLS). Segment-level features included regional longitudinal strain and normalized maximum segment displacement between end-diastole and end-systole. In total, 14 features (2 global + 2 per segment × 6 segments) were computed per patient. Third, MI recognition and customization were performed applying a dual-phase RF cascade. The first RF classifier conducted patient-specific MI detection using all features. If positive, six segment-specific RF classifiers were applied, each using the dual worldwide characteristics plus duo attributes from the segment that are relevant to localize infarcted regions.

## 3. Comparative Analysis

This section compares the previously reviewed models, highlighting their efficiency by providing its merits, demerits and the Evaluation metrics of respective approaches, as in Table 1.

**Table I**. Summary of recent deep learning models thyroid nodule detection

| Ref No. | Techniques Used | Merits | Demerits | Dataset | Performance Metrics |
|---|---|---|---|---|---|
| [14] | CNN, LDA, DT, RF, SVM | The number of features is efficiently minimized which makes it valuable in both engineering and medical perspective | This model is computationally intensive | HMC-QU dataset | For SVM (best): 6-segment features: Sensitivity = 85.97%, Specificity = 74.03% Precision = 86.85%. 5-segment features: Sensitivity = 83.09%, Specificity = 74.03% Precision = 86.85%, F1-score = 84.83%, Accuracy = 80.24% |
| [15] | U-Net–based encoder, Hierarchical ConvLSTM, | The model is less complex due to its simple architecture | This investigation conducted cross-verification using a restricted cohort of individuals at a single site. | MCE dataset | Dice coefficients = 0.81 ± 0.07, IoU = 0.68 ± 0.09, Hausdorff distance (HD)= 27.59 ± 12.82, Mean ± Standard Deviation = 54.63 ± 4.72 |
| [16] | 3D-CSN | The computational timing of the method was less for both segmentation and motion estimation | The model results heavily relied on synthetic samples | Publicly available dataset with simulated echo images | A2C Dice = 0.826, A3C Dice = 0.808, A4C Dice = 0.833 |
| [17] | 2D CNN, 3D CNN | The model is less complex due to its lightweight architecture | It requires more processing time | HMC-QU dataset | 2D CNN: Accuracy = 97.18%, 3D CNN: Accuracy = 90.9%, Precision = 100%, Recall = 97.2%, F1-score = 97.2% |

| | | | | | |
|---|---|---|---|---|---|
| [18] | CNN, Autoencoder | This model utilized real-time data making the results trustable and reliable. | The study population is small | Cohort study | Sensitivity = 75.5%, Specificity = 74.1%, Accuracy = 74.8%, AUC = 0.79 |
| [19] | XceptionNet, LSTM-U-Net, 3D CNN, | This model can efficiently recognize even when the image quality is low. | This model faced black box issue with the ambiguous elements and their contributions to the outcome | Internal test and exterior testing datasets | Internal test dataset: AUC = 0.913, Sensitivity = 85.4%, Specificity = 83.2%; External test dataset: Sensitivity = 81.6%, Specificity = 83.7%, AUC = 0.897, |
| [20] | CNN–LSTM, VGG16, | The model significantly avoided the problem of overfitting | This method is only performed offline, thereby lacking real-time processing | Fujita health university hospital dataset | Long-axis view images: Accuracy = 85.1%. Short-axis papillary muscle level images: Accuracy = 83.2% |
| [21] | SAF-Net, DenseNet-121, Inception-v3, and ResNet-50 | The architecture was compact and computationally efficient | The dataset employed was limited which may introduce overfitting when unseen dataset is applied | HMC-QU-TAU dataset | Accuracy = 78%, Precision = 88.26%, Specificity = 79%, Geometric mean = 77.19% |
| [22] | AIEchoDx, Inception-V3, multi-layer 1DCNN | The model performed efficiently on multi-class diseases which are usually difficult in this application | The model was evaluated with only a small subset of data which may limit its generalizability | There were 1,807 echo movies from 1276 patients with various pathologies, including usual, ASD, HCM, DCM, and previous MI. | AUC = 0.99 |
| [23] | U-Net–style CNNs, SVM, DT, LR, KNN | The model could capture multiple features at once without increasing the complexity | The results yielded comparatively lower values when tested externally on local datasets than public datasets. | HMC-QU and E-Hospital datasets | HMC-QU: Precision = 85.2%, Specificity = 70.1%, Sensitivity = 85.9%, Accuracy = 85.5%, F1-score = 80.2%. E-Hospital: F1 score = 0.8, Accuracy = 76.7%, Sensitivity = 77.8%, Specificity = 75.0%. |
| [24] | CNN, ECV-3D | This model avoids the problem of overfitting | To make the model feasible, the initial cost of implementation remains high | Physikalisch-Technische Bundesanstalt (PTB) and Cleveland HD datasets | PTB: Accuracy = 95.2%, Sensitivity = 93.8%, Specificity = 96.5%. Cleveland HD: Accuracy = 92.7%, Sensitivity = 91.5%, Specificity = 93.8%. |
| [25] | 1DCNN, SVM, DT, RF | The APs can efficiently identify the actual endocardial limit, even in suboptimal perspectives. | The model demands more computational time | HMC-QU dataset | For 1DCNN: Sensitivity = 83.75%, Specificity = 78.00%, Precision = 85.90 %, F1−Score = 84.81%, F2−Score = 84.17%, Accuracy = 81.54% |
| [26] | Pre-trained U-Net, Temporal ConvNet, TSN, SVM, RF, DT, KNN, MLP | The module was enhanced with the capacity to analyze a wider range of motion characteristics | The architecture is computationally heavy | HMC-QU dataset | For MLP: Sensitivity = 80.60%, Specificity= 81.40%, Precision = 79.83%, F1-score = 80.24%, Accuracy = 87.19% |
| [27] | LSTM | In addition to the used data, novel datasets are tested to avoid generalizability | The model takes more training time | Custom dataset | Accuracy = 95%, Sensitivity = 96%, Specificity = 94%, AUC = 0.98 |
| [28] | HOG, U-Net++, ResNeSt50, Harris corner detection, CNN | The model was fine-tuned to enhance the accuracy. | When any changes in LV size, the data collection became challenging which may decrease the accuracy | HMC-QU dataset | Specificity = 99.76%, Precision = 94.68%, F1-score = 94.64%, Accuracy = 99.54% |
| [29] | GAN, encoder-decoder CNN | This model consumes less time even when diverse datasets are employed | The model may reduce the accuracy when low quality images are applied | HMC-QU and Realistic Synthetic 2D Ultrasound datasets | HMC-QU: MSE = 8.6%, Accuracy = 93.75%, Precision = 94.98%, Recall = 95.78%, F1-score = 95.38%. Realistic Synthetic 2D Ultrasound: MSE = 9.2%, Accuracy = 92.72%, Precision = 94.2%, Recall = 95.73%, F1-score = 94.96%. |
| [30] | EfficientNet-B4, two-stage RF cascade | The model takes minimal computational time | A system operates solely using the A4C vision, hence limited its gathering capabilities to this perspective. | HMC-QU dataset | Accuracy = 94.9%, Sensitivity = 100%, Specificity = 89.8% |

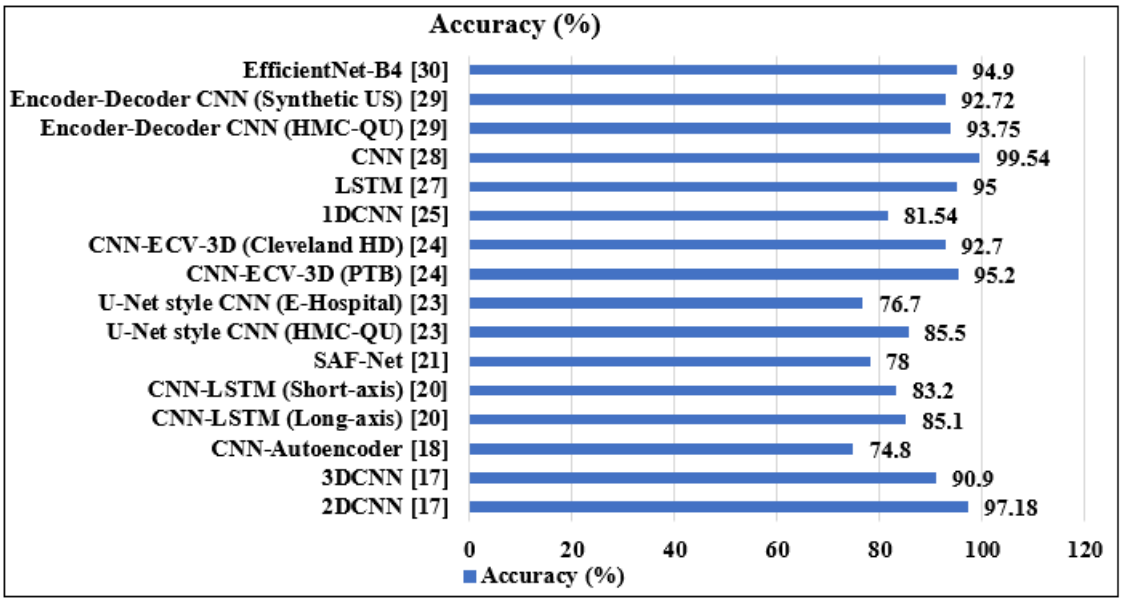


**Figure 6:** Comparison of various Dl-based MI segmentation and detection models

## 4. Performance Evaluation

To further understand the usefulness of existing DL approaches for MI identification using echocardiography, a performance comparison is conducted based on classification accuracy, as summarized in Table 1. Since these studies employ accuracy as a common evaluation metric, a direct quantitative comparison is possible despite differences in datasets, imaging protocols, and architectural complexity. Fig. 6 illustrates the graphical comparison of accuracy among the reviewed models.

From the comparison, the EfficientNet-B4 with two-stage RF cascade model [30] demonstrates the most balanced and practically reliable performance, achieving 94.9% accuracy. Although some models such as [28] report higher raw accuracy (99.54%), their performance may be influenced by dataset-specific conditions and handcrafted feature dependencies, potentially limiting generalizability. In contrast, model [30] integrates deep segmentation with clinically interpretable motion features such as LVEF, GLS, and segmental displacement, improving both diagnostic reliability and clinical relevance.

Other strong-performing methods include the multimodal CNN–ECV–3D framework [24] (95.2% and 92.7%), the LSTM-based temporal model [27] (95%), and the encoder–decoder CNN [29] (93.75% and 92.72%), all of which highlight the value of spatial–temporal modeling. The 2D CNN model [17] achieved 97.18%, outperforming its 3D counterpart, indicating that well-optimized spatial CNNs can sometimes exceed more complex volumetric models. According to the analyses, while several architectures achieve high accuracy, EfficientNet-B4 with two-stage RF cascade model [30] stands out as the most clinically dependable approach due to its combination of high accuracy, lower computational demand, and use of physiologically meaningful features, making it more suitable for real-world deployment.

## 5. Conclusion

Echocardiography plays a crucial role in MI assessment due to its real-time imaging capability and ability to evaluate RWMA. However, manual interpretation is often limited by time consumption and variability in diagnostic outcomes. Early ML approaches provided initial support through feature-based classification but were constrained by manual feature engineering and limited adaptability to complex patterns. DL has significantly enhanced MI detection by enabling automatic extraction of features directly from imaging data. This survey examined various DL models, highlighting their strengths, limitations, and performance across different datasets. The comparative analysis revealed that models combining segmentation, motion-based features, and deep spatial learning generally achieve superior results, while lightweight and clinically interpretable frameworks show greater potential for real-world deployment. Despite promising outcomes, challenges such as limited dataset diversity, computational complexity, and interpretation of models persist. Consequently, forthcoming studies ought to concentrate on creating resilient and computationally efficient spatial–temporal models validated on large, heterogeneous datasets to ensure reliable integration of AI-assisted MI detection into routine clinical practice.

## References

[1] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, and Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction, "Fourth universal definition of myocardial infarction (2018)," Journal of the American College of Cardiology, vol. 72, no. 18, pp. 2231–2264, 2018.

[2] J. G. Canto, R. J. Goldberg, M. M. Hand, R. O. Bonow, G. Sopko, C. J. Pepine, and T. Long, "Symptom presentation of women with acute coronary syndromes:

Myth vs reality," Archives of Internal Medicine, vol. 167, no. 22, pp. 2405–2413, 2007.
[3] H. A. DeVon, C. J. Ryan, A. L. Ochs, and M. Shapiro, "Symptoms across the continuum of acute coronary syndromes: Differences between women and men," American Journal of Critical Care, vol. 17, no. 1, pp. 14–24, 2008.
[4] B. Ibanez, S. James, S. Agewall, M. J. Antunes, C. Bucciarelli-Ducci, H. Bueno, and P. Widimský, "2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation," European Heart Journal, vol. 39, no. 2, pp. 119–177, 2018.
[5] World Health Organization, "Cardio vascular diseases fact sheet," World Health Organization, 2023.
[6] R. M. Lang, L. P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, and J. U. Voigt, "Recommendations for cardiac chamber quantification by echocardiography in adults," European Heart Journal–Cardiovascular Imaging, vol. 16, no. 3, pp. 233–271, 2015.
[7] P. J. Scanlon, D. P. Faxon, A. M. Audet, B. Carabello, G. J. Dehmer, K. A. Eagle, and S. C. Smith, "ACC/AHA guidelines for coronary angiography," Journal of the American College of Cardiology, vol. 33, no. 6, pp. 1756–1824, 1999.
[8] B. Vattay, S. Borzsák, M. Boussoussou, M. Vecsey-Nagy, Á. L. Jermendy, F. I. Suhai, and B. Szilveszter, "Association between coronary plaque volume and myocardial ischemia detected by dynamic perfusion CT imaging," Frontiers in Cardiovascular Medicine, vol. 9, pp. 974805, 2022.
[9] P. A. Pellikka, A. Arruda-Olson, F. A. Chaudhry, M. H. Chen, J. E. Marshall, T. R. Porter, and S. G. Sawada, "Guidelines for performance, interpretation, and application of stress echocardiography in ischemic heart disease," Journal of the American Society of Echocardiography, vol. 33, no. 1, pp. 1–41, 2020.
[10] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
[11] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," Journal of the American College of Cardiology, vol. 69, no. 21, pp. 2657–2664, 2017.
[12] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C. P. Langlotz, and J. Y. Zou, "Video-based AI for beat-to-beat assessment of cardiac function," Nature, vol. 580, no. 7802, pp. 252–256, 2020.
[13] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," npj Digital Medicine, vol. 1, no. 1, pp. 6, 2018.
[14] A. Degerli, M. Zabihi, S. Kiranyaz, T. Hamid, R. Mazhar, R. Hamila, and M. Gabbouj, "Early detection of myocardial infarction in low-quality echocardiography," IEEE Access, vol. 9, pp. 34442–34453, 2021.
[15] M. Li, D. Zeng, Q. Xie, R. Xu, Y. Wang, D. Ma, and H. Fei, "A deep learning approach with temporal consistency for automatic myocardial segmentation," International Journal of Cardiovascular Imaging, vol. 37, no. 6, pp. 1967–1978, 2021.
[16] Y. Deng, P. Cai, L. Zhang, X. Cao, Y. Chen, S. Jiang, and B. Wang, "Myocardial strain analysis of echocardiography based on deep learning," Frontiers in Cardiovascular Medicine, vol. 9, pp. 1067760, 2022.
[17] O. Hamila, S. Ramanna, C. J. Henry, S. Kiranyaz, R. Hamila, R. Mazhar, and T. Hamid, "Fully automated 2D and 3D convolutional neural networks pipeline for video segmentation and myocardial infarction detection," Multimedia Tools and Applications, vol. 81, no. 26, pp. 37417–37439, 2022.
[18] F. Laumer, D. Di Vece, V. L. Cammann, M. Würdinger, V. Petkova, M. Schönberger, and C. Templin, "Assessment of artificial intelligence in echocardiography diagnostics in differentiating takotsubo syndrome from myocardial infarction," JAMA Cardiology, vol. 7, no. 5, pp. 494–503, 2022.
[19] X. Lin, F. Yang, Y. Chen, X. Chen, W. Wang, X. Chen, and K. He, "Echocardiography-based AI detection of regional wall motion abnormalities," Frontiers in Cardiovascular Medicine, vol. 9, pp. 903660, 2022.
[20] R. Muraki, A. Teramoto, K. Sugimoto, K. Sugimoto, A. Yamada, and E. Watanabe, "Automated detection scheme for acute myocardial infarction using convolutional neural network and long short-term memory," PLoS One, vol. 17, no. 2, pp. e0264002, 2022.
[21] I. Adalioglu, M. Ahishali, A. Degerli, S. Kiranyaz, and M. Gabbouj, "SAF-Net: Self-attention fusion network for myocardial infarction detection using multi-view echocardiography," in Proceedings of Computing in Cardiology (CinC), 2023, vol. 50, pp. 1–4.
[22] B. Liu, H. Chang, D. Yang, F. Yang, Q. Wang, Y. Deng, and K. He, "A deep learning framework assisted echocardiography with diagnosis and lesion localization," Scientific Reports, vol. 13, no. 1, pp. 3, 2023.
[23] T. Nguyen, P. Nguyen, D. Tran, H. Pham, Q. Nguyen, T. Le, and H. Pham, "Ensemble learning of myocardial displacements for myocardial infarction detection," Frontiers in Cardiovascular Medicine, vol. 10, pp. 1185172, 2023.
[24] S. Deepika and N. Jaisankar, "Detecting and classifying myocardial infarction in echocardiogram frames," IEEE Access, vol. 12, pp. 51690–51703, 2024
[25] A. Degerli, S. Kiranyaz, T. Hamid, R. Mazhar, and M. Gabbouj, "Early myocardial infarction detection over multi-view echocardiography," Biomedical Signal Processing and Control, vol. 87, pp. 105448, 2024.
[26] S. Kasim, J. Tang, S. Malek, K. S. Ibrahim, R. E. R. Shariff, and J. K. Chima, "Enhancing regional wall abnormality detection accuracy," PLoS One, vol. 19, no. 9, pp. e0310107, 2024.
[27] H. T. Soe and H. Iwata, "Enhancing myocardial infarction diagnosis: LSTM-based deep learning approach," Journal of Medical and Biological Engineering, vol. 44, no. 5, pp. 704–710, 2024.
[28] F. C. A. Tabuco and P. C. Naval, "Myocardial infarction detection using video frame key-point and gradient matching," in Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2024, pp. 1–7.

[29] B. A. Valanrani and S. D. Suganya, "Improving myocardial infarction detection from echo images," International Journal of Engineering Trends and Technology, vol. 72, no. 12, pp. 285–297, 2024.
[30] C. Gomez, A. Letizia, V. Tufano, F. Molinari, and M. Salvi, "A cascade approach for the early detection and localization of myocardial infarction in 2D-echocardiography," Medical Engineering & Physics, pp. 104400, 2025.