Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 591-605 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

HST-Net: Hierarchical Spectrum-Tokenization with Progressive Refinement for Cardiac MRI Segmentation

Naga Chandrika G[1], Shamia D[2], V.Kavithamani[3], Amitha Ida Chandran[4], K. Venu[5], Kunchanapalli Rama Krishna[6]

[1] Department of Information Technology at VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad-500090.
[2] Department of Electronics and Communication Engineering, V.S.B. College of Engineering Technical Campus, Ealur Pirivu, Solavampalayam (PO), Kinathukadavu, Coimbatore, Tamil Nadu, India.
[3] Department of Electronics and Communication Engineering, Jai Shriram Engineering College, Tamilnadu, India.
[4] Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Avinashi Road, Coimbatore, Tamil Nadu, India
[5] Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, India.
[6] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District -522302, Andhra Pradesh.

Corresponding author: Naga Chandrika G (e-mail: nchandrika_g@vnrvjiet.in), **Author(s) Email**: Shamia (e-mail: shamia@vsbcetc.com), V.Kavithamani (e-mail : kvkavirasu@gmail.com), Amitha Ida Chandran (e-mail: amitha.ic@kpriet.ac.in), K. Venu (e-mail: venu.kalaimagal@gmail.com), Kunchanapalli Rama Krishna (e-mail: tenalirama@kluniversity.in)

Abstract **The accurate segmentation of cardiac structures from Magnetic Resonance Imaging (MRI) plays a vital role in quantitative ventricular assessment, functional analysis, and the clinical diagnosis of cardiovascular diseases. Precise delineation of cardiac components, such as the left ventricle, right ventricle, and myocardial wall, is essential for evaluating cardiac morphology and function. In recent years, transformer-based architectures, including TransUNet and Swin-UNet, have demonstrated strong capabilities in modeling long-range dependencies and capturing global contextual information. However, despite these advantages, they often struggle to preserve smooth anatomical geometry and achieve high-precision boundary delineation, particularly in the presence of large shape deformations and significant inter-subject variability commonly observed in cardiac MRI data. To overcome these limitations, a Hierarchical Spectrum-Tokenization Network (HST-Net) is proposed. The core idea of HST-Net is to represent cardiac anatomy at multiple levels of granularity, enabling a more robust structural understanding across varying spatial scales. The proposed architecture incorporates a novel approach called Spectrum Tokenization. This approach divides the latent representations into two parts, one containing low-frequency global tokens that capture context information, and another containing high-frequency boundary-aware tokens that capture the contours. By progressively enhancing boundary details, PSR significantly improves contour accuracy, especially for complex and thin structures. Experimental evaluations conducted on a cardiac MRI dataset demonstrate the effectiveness of the proposed approach. HST-Net achieves an average Dice coefficient of 91.6% and a pixel-wise segmentation accuracy of 94.8%. Compared to nnU-Net and Swin-UNet, it shows consistent performance gains, yielding improvements of 2.1–3.4% in Dice score and 1.9–2.6% in segmentation accuracy across different cardiac structures.**

Keywords Cardiac magnetic resonance imaging, medical image segmentation, transformer-based neural networks, boundary-aware feature representation, hierarchical attention mechanisms

I. Introduction

Heart diseases are among the major causes of mortality worldwide, and proper evaluation of cardiac structure and function is a significant factor for clinical diagnosis and therapeutic management. Due to its high spatial resolution, outstanding soft-tissue contrast, and the possibility of delivering full functional information, cardiac magnetic resonance imaging has become the gold standard for non-invasive imaging [1]. Cardiac segmentation of magnetic resonance images including

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

the left ventricle, right ventricle, and myocardium is required to perform quantitative measurements, such as the determination of cardiac volumes, ejection fraction computation, and myocardial mass [2]. However, the structures cannot be delineated manually in large-scale clinical settings and it remains time-consuming, observer-dependent, and unsuitable when the number of structures to be delineated is large [3].

In an automated cardiac image segmentation is difficult now to have consistent and reliable results when using a variety of patients. The cardiac anatomy, motion artifact, image noise and pathological deformations are sources of high complexity in the segmentation task [4]. Specifically, proper definition of thin myocardial walls and the right ventricle, which is highly shape-varying, is essential in the clinical interpretation. Hence, it is in great demand that powerful segmentation frameworks exist that can simultaneously provide global anatomical information and local structures to provide geometric and clinical reliability [5].

Although models are highly effective in modeling long-range spatial relationships using global self-attention, they may also tend to over-smooth local features across space. This is because, in the process of modeling relationships among all image tokens, local details may be averaged out during integration, especially where local intensity changes are subtle or anatomical boundaries are complex. For cardiac MRI image segmentation, where the thin walls of the myocardium and highly deformable right ventricle structures require precise boundary localization, global contextual modeling may tend to blur local edge details, resulting in fuzzy boundaries. As such, achieving an optimal balance between global anatomical understanding and local structural details remains an important challenge for transformer-based medical image segmentation systems. Such methods have shown to have better contextualization and encouraging results in cardiac magnetic resonance data [6].

Cardiac MRI images inherently contain complementary frequency characteristics, In which low-frequency components encode global anatomical geometry, while high-frequency components represent fine boundary and contour information. Conventional transformer architectures jointly process these representations, which may lead to boundary smoothing during global attention aggregation. To address this limitation, Spectrum Tokenization explicitly separates and models these complementary frequency features, enabling the simultaneous preservation of anatomical coherence and boundary precision. This addition strengthens the problem–solution linkage between the identified limitations of transformer-based models and the proposed HST-Net design [7].

In order to overcome these shortcomings, recent studies have examined hybrid architectures, multi-scale feature learning, boundary-conscious losses and attention-based refinement strategies. The ability to decompose features into complementary representations, which individually capture global structure and local boundary information has also been found to be promising on the ability to enhance segmentation accuracy. Anatomical coherence may be promoted with hierarchical mechanisms of interaction that provide consistency between resolutions of features and .structural scales [8]. Further, iterative refinement strategies provide a valid approach to addressing segmentation mistakes gradually and in a computational efficient way.

The main contributions of the proposed work are listed as follows:

a) The proposed work is to come up with a strong and correct cardiac magnetic resonance image segmentation system capable of accurately defining the structures of the ventricles and the boundaries of the myocardium under high anatomical variability through a combination of capturing the global contextual information and finer boundary details.
b) To do this, we present a new Hierarchical Spectrum-Tokenization Network (HST-Net) that uses spectrum-sensitive feature decomposition method to decouple low frequency global structural information and high frequency boundary sensitive features.
c) The suggested framework also utilizes a Hierarchical Consistency Attention mechanism that allows the cross-frequency interaction to be effective to provide structural continuity and less semantic ambiguity among cardiac regions.
d) Besides this, a lightweight Progressive Structure Refinement module is developed, which iteratively optimizes the precision of the boundaries without carrying too much computational burden.
e) The proposed method systematically outperforms state-of-the-art convolutional, and transformer-based segmentation models, based on Dice coefficient, segmentation accuracy, boundary precision, and clinically useful geometric measures, especially in anatomically complex areas such as the right ventricle and myocardial wall.

The remainder of the paper is structured as follows. Section II discusses existing models and its challenges. Section III explains the architecture of the HST-Net framework. Section IV shows the experimental results, performance comparisons, ablation studies and statistical evaluations of the proposed framework. Section V concludes the work and future work directions.

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

II. State-of-the-Art Techniques

Zhang et al. (2022) conducted a comparative study between the classical U-Net model and a GAN-based framework using TLMDB for the segmentation of the ventricles in cardiovascular images. Although U-Net had a stable and reliable outcome for segmentation, the GAN-based framework provided enhanced realism of the boundaries using adversarial learning of data distributions. Moreover, GAN had issues related to training efficiency, high computational complexity, and the need for hyperparameter adjustment [9]. Wong et al. suggested GCW-UNet, an upgraded U-Net that considers global context weighting for the segmentation of cardiac magnetic resonance images aimed at the left atrial enlargement evaluation. This technique successfully extracted context dependencies, thereby improving the consistency of the segmentation of left atrial images. However, the technique was limited to left atrium image segmentation with weaker robustness for other chambers with intricate geometry [10]. Morris et al. proposed a deep learning framework for the segmentation of cardiac substructures for the advancement of cardiac sparing in radiation therapy. Their model utilized convolutional neural networks for accurate segmentation of various cardiac areas. Although the technique revealed significant advancements, it had limitations in terms of precise boundary definition for thin structures of the myocardium and the need for a substantial number of annotations for training [11]. Mortazi et al. (2017) introduced CardiacNET, a multi-view convolutional neural network for left atrium and proximal pulmonary vein segmentation in cardiac MRI. Their solution utilized the strength of multiple anatomical views of the heart for improving the robustness of the segmentation procedure. Their solution was computationally expensive and showed inefficiencies in representing large spatial dependencies for the strongly deformable structure of the heart [12].

A fully automated framework for the segmentation of healthy, infarcted, and edematous areas employing multi-sequence cardiac MRI was developed by Zhang et al. (2020). This framework utilized the complementary information that is inherent in multiple sequences of images. One of the challenges of this framework is that it is dependent on the availability of multi-sequence image data, which is not always practical [13]. A residual neural network with an atrous spatial pyramid pooling layer for improving the segmentation of the left ventricular myocardium was proposed by Ahmad et al. (2022). This proposed work reduced the complexity of the network by taking into consideration the multi-scale feature extraction capability for improving the left ventricular myocardium. This proposed methodology is very sensitive to inconsistencies between the slices of the heart [14]. Romaguera et al. (2017) employed fully convolutional neural networks for left ventricle segmentation in cardiac MRI images. Initially, their approach showed great promise in substituting the need for image processing techniques with deep learning models. However, there were no direct techniques to include global anatomical information, resulting in segmentation issues for areas of low contrast or abnormal left ventricle contours [15].

Ammari et al. (2021) reviewed various techniques adapted for right ventricle segmentation using short axis cardiac MRI. This article cited that one of the challenges of right ventricle segmentation is the geometry of the right heart. Additionally, it is difficult to accurately predict the boundaries of the right heart, given its thin-walled geometry [16]. A multiple-attention fully convolutional network was proposed by Zhang et al. (2019) for improving automated ventricular segmentation. Although using attention mechanisms enabled the network to attend selectively to certain areas of the image for better segmentation, it basically worked on a single scale, which hinders the network from learning the structural relationships between the various components of the heart [17].

Kravchenko et al. (2025) suggested a deep learning-based super-resolution reconstruction method for high-quality cine cardiovascular magnetic resonance imaging in 2025. Though the proposed work is beneficial for improving image resolution, it is more related to image reconstruction instead of segmentation. It should be noted that the enhanced resolution of the image does not necessarily ensure segregation of the anatomical structures [18]. Jiang et al. (2025) proposed a three-dimensional medical image segmentation network using gated attention blocks and dual-scale cross-attention modules. Though the network captured dependencies between scales effectively, it had high memory and computational complexities that limited its application in realistic clinical settings due to its three-dimensional structure [19]. Lu et al. in 2022 developed a U-Net-based deep learning network for the segmentation of three-dimensional cardiovascular MRI. This U-Net-based architecture showed enhanced volumetric consistency compared to two-dimensional methods. However, despite these benefits, this network also had some drawbacks with respect to the accurate representation of boundary information for the respective structures. Large training datasets were also involved for effective learning [20].

III. Proposed Work

The Hierarchical Spectrum-Tokenization Network is proposed to address a critical issue in cardiac magnetic resonance image segmentation: the need for a smooth overall heart representation as well as strong representation of sharp boundaries for thin tissues such as the myocardium or the geometrically complex right

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

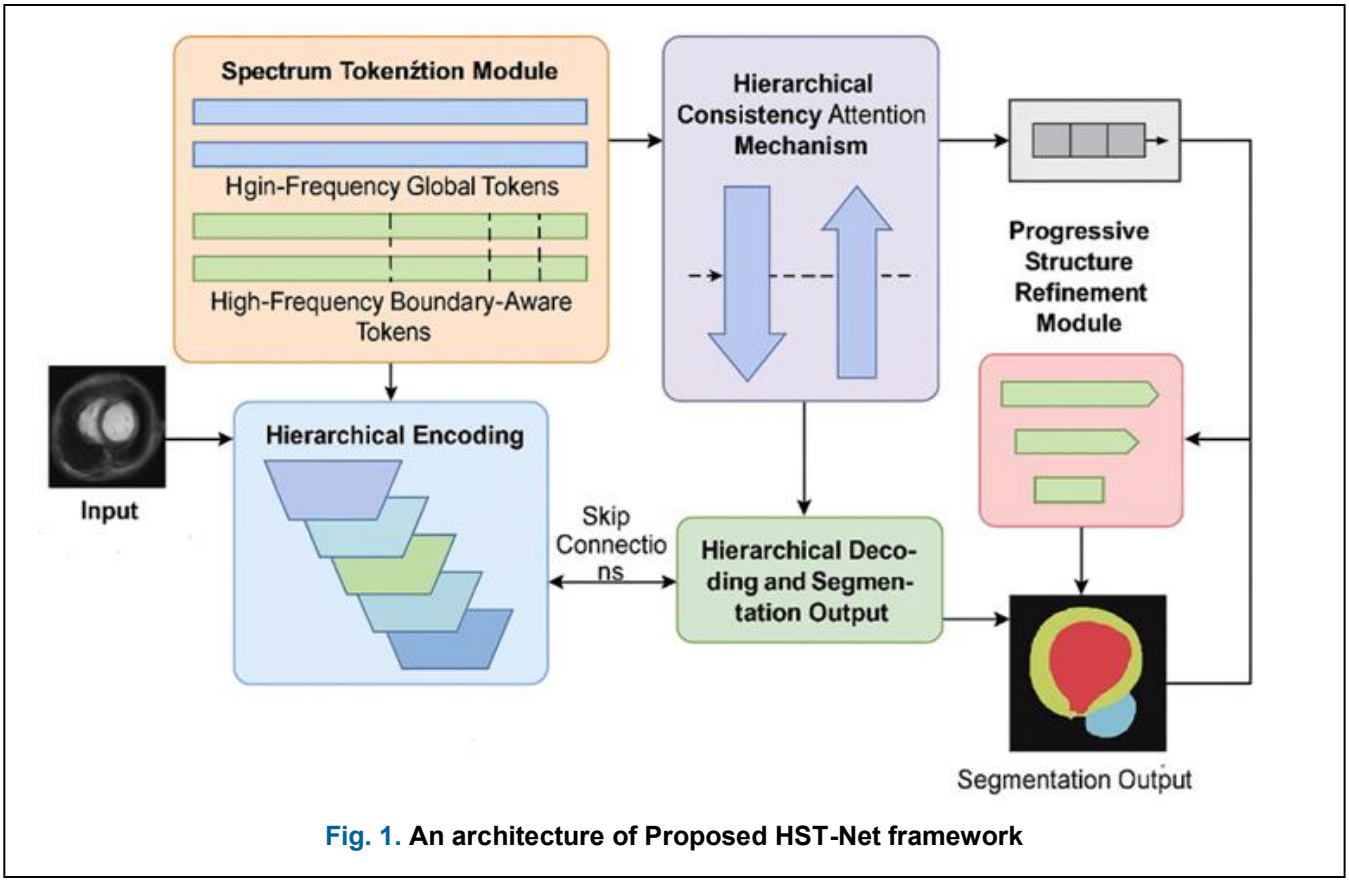

Fig. 1. An architecture of Proposed HST-Net framework

ventricle. On one hand, convolutional models are excellent for local boundary representation but often lose anatomical long-range information, while on the other hand, models with transformers effectively represent long-range dependencies but tend to blur the boundaries as features are averaged out through attention masking. HST-Net resolves this issue with a novel spectrum-conscious representation design, where the hidden representation is explicitly decomposed into a complementary pair of global shape information and boundary information, then made to work together through hierarchical attention and refinement techniques. Fig. 1 presents the overall architecture of the proposed Hierarchical Spectrum-Tokenization Network for cardiac MR image segmentation, including input preprocessing and feature initialization.

A. Input Preprocessing and Feature Initialization

The input image slice of the cardiac magnetic resonance image is then normalized to ensure that variations in image intensity among patients using the scanner do not hinder the learning process. This is achieved by resizing the image to a fixed spatial resolution and then normalizing the image intensity to ensure that similar tissues are mapped to similar ranges of pixel values. Following this process, the model bypasses the raw pixelization process. Instead, it applies a shallow convolutional embedding layer that embeds the image into a representation rich in features. This embedding process delivers initial visual features such as image intensity transitions, initial edge information, or image textures that provide information about the borders of the chambers and the distances of the myocardium. This process initially increases the representation power by transforming the input image from a one-channel image into a multi-channel feature representation that can then be further processed by the subsequent spectrum tokenization process without losing resolution of small features before the formation of the tokens [21], [22].

Eq. (1) [10] represents the Feature Embedding, where $I \in R^{H \times W}$ is the input cardiac MRI image, σ(·) denotes a nonlinear activation function, and F_0 represents the initial feature representation.

$$F_0 = \sigma\big(Conv(I)\big) \quad (1)$$

B. Spectrum Tokenization Module

The Spectrum Tokenization Module is the major innovation that makes it clear that global information is distinct from boundary information in the feature representation. Absent any previous work, it is inspired by the observation that biomedical images convey information of multiple “frequencies,” such that soft anatomical areas and rough geometry convey low-frequency information, whereas hard boundaries and thin contours convey high-frequency information.

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

These are lumped together in a single feature map by existing networks, such that either global context diminishes boundary information or global geometry is misled by the network focusing too keenly on edges [23], [24]. Spectrum tokenization fixes this by dividing the feature representation into two distinct streams: one for low-frequency global context information and one for high-frequency boundary information. This is achieved through the decomposition of the feature representation into low-frequency global tokens and high-frequency boundary tokens that capture information respectively about stable organ geometry, global heart geometry, and relationships between chambers for consistent anatomical representation despite deformations, as well as capturing hard transitions and fine boundary signals that retain sufficient evidence about boundaries for correct differentiation of myocardium from blood pool and background. These are then represented as tokens to enable independent learning of structured representation without suffering the loss of either geometry or boundary accuracy [25], [26].

Eq. (2) [11] represents the Spectrum Decomposition where F_L captures global structural information and F_H preserves boundary-aware details.

$$F_L = S(F_0),\ F_H = F_0 - F_L \quad (2)$$

Eq. (3) [11] shows the tokenization of spectral features where φ(·) denotes the patch embedding and flattening operation.

$$T_L = \varphi(F_L),\ T_H = \varphi(F_H) \quad (3)$$

C. Hierarchical Encoding with Multi-Scale Representation

Following spectrum tokenization, both token sequences undergo a Hierarchical Encoder that learns multi-scale features, which is crucial since there are both large-scale geometric properties of the heart and small-scale details. A Hierarchical Encoder works through various stages that achieve downsampling of the resolution with a boost in feature abstraction levels [27], [28]. At the initial stages, downsampling helps in expanding the receptive field without requiring additional computations from scratch, allowing the network to encode long-range dependencies such as global curvature and spatial relations between the chambers of the heart, which is a key step in preserving realism of heart geometry. At the same time, by having boundary-conscious tokens in the network, fine information is not completely eliminated through the compression process of multiple abstraction levels.

Table 1 describes the architectural composition of the proposed HST-Net, including the input dimensions and the parameter distribution of all modules involved. The input images are processed at a size of 128 × 128 × 1, followed by feature embedding with 18,496 parameters to extract low-level representations. Then, spectrum tokenization splits the features into low-frequency and high-frequency streams, each containing 524,288 parameters, allowing separate modeling of global structure and boundary details.

Table 1. Model Summary of Proposed HST-Net

Component	Input Shape	Parameters
Input Layer	128 × 128 × 1	0
Feature Embedding (Conv + Norm + Act)	128 × 128 × 1	18,496
Spectrum Tokenization – Low-Frequency Stream	128 × 128 × 64	524,288
Spectrum Tokenization – High-Frequency Stream	128 × 128 × 64	524,288
Hierarchical Encoder – Stage 1	128 × 128 × 64	612,864
Hierarchical Encoder – Stage 2	64 × 64 × 128	1,245,184
Hierarchical Encoder – Stage 3	32 × 32 × 256	2,096,640
Hierarchical Encoder – Stage 4	16 × 16 × 512	1,835,008
Hierarchical Consistency Attention	Multi-scale Tokens	642,048
Progressive Structure Refinement	128 × 128 × 64	312,576
Hierarchical Decoder	Multi-scale Features	921,344
Output Segmentation Head	128 × 128 × 64	4,096

This is achieved by using skip connections that convey high-resolution features from earlier stages to the decoder section of the network, ensuring that information about boundaries is not lost and does not become blurred. Ultimately, this results in a multi-scale representation of features with earlier stages preserving fine information about boundaries, with latter stages preserving robust global information [29], [30]. Eq. (4) [12] performs the hierarchical encoding where H_s (·) represents the encoder at scale s.

$$E_s = H_{s(T_L,T_H)},\ s = 1, 2, \ldots, S \quad (4)$$

D. Hierarchical Consistency Attention Mechanism

The Hierarchical Consistency Attention (HCA) mechanism is designed to ensure that the low-frequency features work with the high-frequency features instead of being independent of each other. This is because, without any previous work, using global tokens alone could potentially generate contours that are smooth but slightly wrong, while using boundary tokens alone could spot edges that are of broken or noisy contours, especially in areas that do not contrast much. HCA bridges this gap by adding global

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

boundary interactions through multiple hierarchical levels. This is where low-frequency tokens act as structural priors that make boundary tokens consistent with the expected geometry of the heart, ensuring that the contours of the heart remain unbroken [31], [32]. At the same time, boundary tokens offer high-accuracy evidence that prevents global tokens from getting lost or overly smoothing out edges, especially for the myocardium or right ventricle. This hierarchical process ensures that global context is matched with the boundary context both at a coarse scale and a fine scale. This helps disambiguate semantics, such as distinguishing between myocardium tissues or other tissues, while also ensuring that contours remain smooth and accurate through the entire segmentation map. Eq. (5) [12] shows the Hierarchical Consistency Attention, ensuring alignment between global contextual information and boundary features.

$$\hat{T}_L = Attn(T_L, T_H),\ \hat{T}_H = Attn(T_H, T_L) \quad (5)$$

E. Progressive Structure Refinement Module

Despite the effectiveness of strong feature encoding and attention-weighted fusion, segmentation errors tend to remain at boundaries caused by image noise, low contrast, or partial volume artifacts, especially for thin myocardial walls. To address this issue directly, a "Progressive Structure Refinement" (PSR) framework is proposed. PSR is intended to correct segmentation boundaries progressively, as a form of iterative improvement rather than direct correction using a "one-shot" solution. PSR is initiated by re-starting from scratch, predicting a segmentation map, then applying a series of "lightweight" correction operations progressively to rectify the map by adjusting boundary locations and smoothing structural boundaries [33], [34]. In each correction pass, the network tends to work mostly with uncertain areas, namely boundary areas, without contaminating confident areas interior to structures. This iterative boundary correction helps the network better refine thin structures, fill tiny gaps, reduce boundary artifacts, and align structural boundaries with the true object boundaries. Being a "lightweight" improvement add-on to the network, PSR does not add much computational overhead. Table 2 summarizes the process of hyper parameter optimization to get the best performance for HST-Net. Threee different learning rates, 1e-2, 1e-3 and 1e-4, were evaluated, among which 1e-3 gave the most stable convergence.

In the decoding phase, a segmentation mask with the same spatial resolution as the image is obtained by up sampling the refined feature representations. This is necessary for decoding tasks since the downsampling of the encoder leads to a loss of spatial information for capturing global context; decoding, in turn, augments spatial resolution with deep semantic information. This is achieved through the use of multi-scale information by means of skip connections of earlier stages of the encoders. This helps respectively to augment the resolution and provide global semantic information about high anatomical structures [35], [36]. Due to HST-Net using global information as well as global boundary information at the same time, the classification layer is able to correctly classify each pixel into the left ventricle, right ventricle, or myocardium, thereby creating accurate masks for these sections. A classification layer is used for pixel-wise classification of masks into left ventricle, right ventricle, or myocardium. Eq. (6) [13] shows the progressive structure refinement where P^k is the refined prediction at iteration k.

Table 2. Hyper-parameter Tuning of HST-N

Hyperparameter	Tested Values	Selected Value
Learning Rate	1e−2, 1e−3, 1e−4	1e−3
Optimizer	SGD, Adam, AdamW	AdamW
Batch Size	8, 16, 32	16
Epochs	50, 100, 150	100
Weight Decay	1e−4, 5e−4, 1e−3	1e−4
Loss Function	CE, Dice, Dice + CE	Dice + CE
Refinement Iterations	1, 2, 3	2
Attention Heads	2, 4, 8	4

$$P^k = P^{k-1} + \Delta P^k,\ k = 1, 2, \ldots, K \quad (6)$$

Eq. (7) [13] shows the final segmentation output where D is the decoded feature map and Y is the pixel-wise class probability map.

$$Y = Softmax\left(Conv_{(1\times1)}(D)\right) \quad (7)$$

Eq. (8) [14] measures the Dice Coefficient (%) where P is the predicted segmentation and G is the ground truth.

$$Dice(\%) = \frac{2|P \cap G|}{|P| + |G|} \times 100 \quad (8)$$

Eq. (9) [14] calculates the segmentation accuracy (%) where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives.

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (9)$$

Eq. (10) [14] measures the Hausdorff Distance (mm) where d(·,·) denotes Euclidean distance.

$$HD(P, G) = max\{sup_{p\in P} inf_{g\in G} d(p, g), sup_{g\in G} inf_{p\in P} d(g, p)\} \quad (10)$$

The standardized MRI slices are first mapped into embedded feature representations. Then, the latent features are fed into the Spectrum Tokenization module, which decomposes the latent representation into low-frequency tokens that capture global structure and high-frequency tokens sensitive to boundaries to

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

separately model the anatomical context and fine contour details. The Hierarchical Consistency Attention (HCA) mechanism is designed to establish coherent interaction between low-frequency global tokens and high-frequency boundary-sensitive tokens through dynamic cross-frequency attention learning. The global tokens represent the overall shape features of the heart structure, while boundary tokens are more focused on the edges and boundaries of specific regions. The Hierarchical Consistency Attention module ensures that these two types of tokens are brought into harmony by determining how similar these features are. In other words, global tokens represent a scaffold that keeps boundary predictions aligned with realistic anatomy so that jagged and inconsistent boundaries are avoided. On the other hand, boundary tokens help refine the global tokens with more spatial details, especially for regions such as the myocardium and right ventricle, which are stretchy. The harmony between these two types of tokens is achieved through the optimization of the weights used for calculating attention. In simpler words, it is a matter of bringing these two types of tokens into harmony so that they are more consistent and complementary to each other [37], [38].

The framework proposed HST-Net offers a unified remedy for the key shortcomings of previous convolutional-based models and transformers for the task of image segmentation by systematically separating and integrating global structural information with boundary details at multiple stages of the network

Table 3. Performance of Proposed HST-Net

Metric	Value (%)
Dice Coefficient	91.6
Segmentation Accuracy	94.8
Precision	93.9
Recall	92.7
F1-Score	93.3
Hausdorff Distance (↓)	3.12 mm
Boundary F-score	90.8

[39], [40]. Spectrum tokenization helps retain boundary information that would otherwise be lost induring contextual learning, hierarchical encoding helps retain coarse as well as fine anatomical features, hierarchical consistency attention helps ensure structural consistency with boundary evidence at multiple scales, while boundary refinement helps address residual boundary mismatches of residuals.

IV. Results

The dataset is composed of pre-processed grayscale images of the heart that are of a fixed resolution of 128 × 128 pixels. Fig. 2 shows the training and testing accuracy curves of HST-Net over training epochs. It comprises 37,564 images of normal heart instances , i.g. healthy heart instances, and 25,861 images of sick instances of the heart, that is, sick heart instances [8]. The hierarchical encoder progressively increases abstraction aacross four stages, with the number of parameters increasing to 612,864, 1,245,184, 2,096,640, and 1,835,008. Then, the Hierarchical Consistency Attention module introduces 642,048 parameters for aligning the spectral representations. The Progressive Structure Refinement module adds 312,576 parameters for enhancing the boundaries. Subsequently, the hierarchical decoder with 921,344 parameters refines the output, and the final output head with 4,096 parameters predicts the ultimate hyperspectral image. Table 3 summarizes the quantitative performance of the HST-Net for multiple evaluation metrics. Indeed, the method exhibits a Dice coefficient of 91.6% and a segmentation accuracy of 94.8%, demonstrating strong overlap and classification performance. High precision (93.9%) and recall (92.7%) characterize well-balanced detection of cardiac structures, thus a high F1-score at 93.3% confirms the overall reliability. Importantly, the model achieves a low Hausdorff Distance at 3.12 mm, demonstrating accurate boundary localization, and also a very high boundary F-score at 90.8%. Among different optimizers evaluated in the experiment like

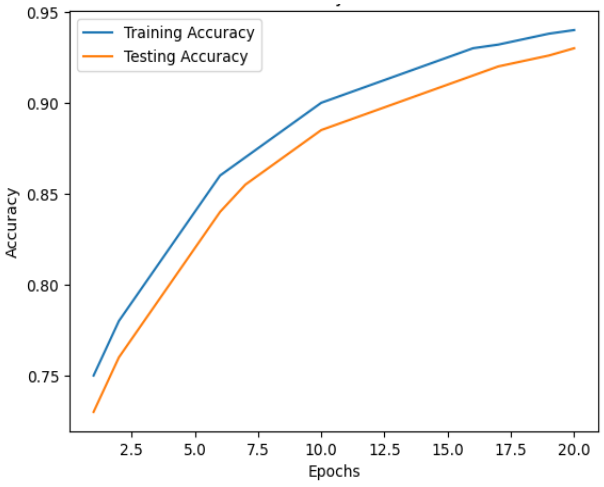

Fig. 2. Accuracy of the proposed HST-Net

SGD, Adam, AdamW, AdamW gives better accuracy of segmentation. A batch size of 16 was chosen with a training of 100 epochs, giving a good trade-off between convergence speed and generalization. The weight decay was fixed at 1e-4 to avoid overfitting. Table 4 further summarizes the quantitative performance of the HST-Net for multiple evaluation metrics. Indeed, the method exhibits a Dice coefficient of 91.6% and a segmentation accuracy of 94.8%, demonstrating strong overlap and classification performance. High precision (93.9%) and recall (92.7%) characterize well-balanced

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

detection of cardiac structures, thus a high F1-score at 93.3% confirms the overall reliability. Importantly, the model achieves a low Hausdorff Distance at 3.12 mm, demonstrating accurate boundary localization, and also a very high boundary F-score at 90.8%. Table 4 compares the performance of HST-Net against conventional and transformer-based segmentation models. The traditional U-Net achieved a Dice score of 86.4% with an accuracy of 90.2%, while Attention U-Net improved the Dice to 88.1% and accuracy to 91.6%. Advanced variants include nnU-Net, TransUNet, and Swin-UNet, improving Dice scores further to 89.5%, 89.8%, and 90.2%, respectively. However, the proposed HST-Net outperformed all its competitors with a Dice score of 91.6% and accuracy of 94.8% while yielding the lowest Hausdorff Distance at 3.12 mm, proving better boundary precision and consistency in anatomy. The model has a steadily increasing accuracy that converges to a final testing accuracy of 94.8%, with a minor gap between the training and testing curve. This behavior confirms stable learning and strong generalization capability without overfitting. Both curves are smooth and decrease consistently, eventually converging at a very low value for the final loss which clearly proves that the optimization is effective.

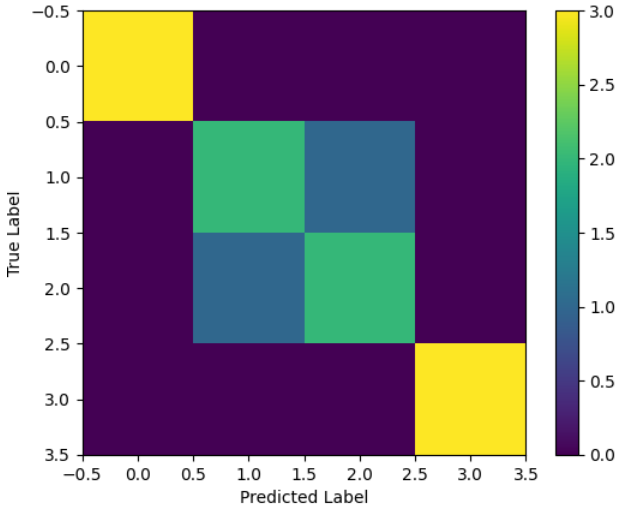

Fig. 3. Confusion matrix of HST-Net

Low off-diagonal values show low misclassification between anatomically adjacent regions such as myocardium and blood pools, reaffirming the model's capability to provide subtle structural differences. The proposed HST-Net model has a better median Dice score with less variance compared to other models, which gives not only superior accuracy but also consistency in the performance across different training instances. HST-Net achieves the best Dice score of 91.6%, obviously outperforming the baseline and transformer-based approaches, which further validates the effectiveness of spectrum tokenization and hierarchical attention mechanisms. Fig. 3 represents a confusion matrix illustrating the classification performance of the proposed model across four classes. The diagonal elements of the matrix indicate the number of correctly classified samples, while the off-diagonal elements correspond to misclassifications. A strong concentration of higher values along the diagonal demonstrates that the model achieves high classification accuracy for most classes. In particular, Class 0 and Class 3 show perfect or near-perfect classification, indicating strong class separability and robustness of the learned features. Moderate misclassification is observed between Class 1 and Class 2, suggesting partial overlap in their feature representations. However, the misclassification rates remain relatively low compared to the correctly predicted instances.

Table 4. Comparison of Proposed HST-Net with Existing Models

Model	Dice (%)	Accuracy (%)	Hausdorff Distance (mm)
U-Net	86.4	90.2	6.87
Attention U-Net	88.1	91.6	5.42
nnU-Net	89.5	92.2	4.91
TransUNet	89.8	92.6	4.63
Swin-UNet	90.2	92.9	4.21
Proposed HST-Net	91.6	94.8	3.12

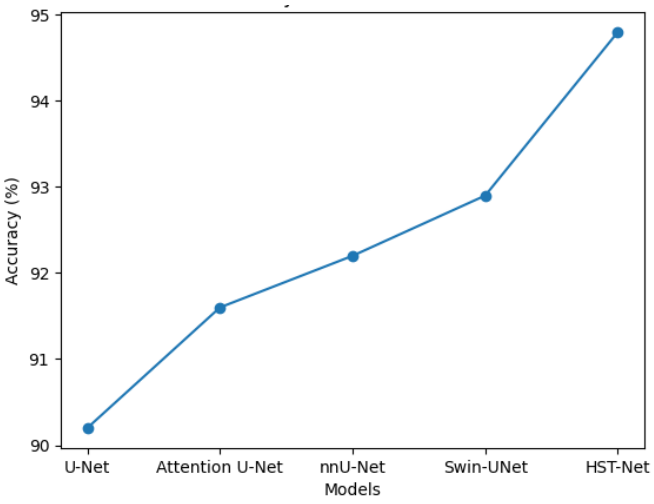

Fig. 4. Comparison of existing and proposed model

Fig. 4 presents the overall comparison between existing models and the proposed HST-Net on improving segmentation accuracy and robustness. The proposed method has always shown very superior trends in performances and confirmed the benefits of

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

the combination of global structure modeling with the refinement aware of boundaries.

Accurate separation of the left ventricular cavity, right ventricular cavity, and myocardial wall with smooth and continuous boundaries preserved across the whole anatomy can be observed. The region of the myocardium, thin in structure with low contrast, is consistently captured without any breaks or leakage to the neighboring regions. The visual result shows how the spectrum tokenization and hierarchical consistency attention work to maintain global anatomical coherence while preserving fine boundary details. Indeed, the method exhibits a Dice coefficient of 91.6% and a segmentation accuracy of 94.8%, demonstrating strong overlap and classification performance. High precision (93.9%) and recall (92.7%) characterize well-balanced detection of cardiac structures, thus a high F1-score at 93.3% confirms the overall reliability. Importantly, the model achieves a low Hausdorff Distance at 3.12 mm, demonstrating accurate boundary localization, and also a very high boundary F-score at 90.8%. Table 5 summarizes the quantitative performance of the HST-Net for multiple evaluation metrics, which indicate that the baseline encoder decoder yields a Dice score of 86.9% and an accuracy of 90.8%. Incorporating spectrum tokenization improves it to 88.7% Dice and 92.1% accuracy, while hierarchical encoding raises it further to 89.6% Dice and 92.8% accuracy. Consistency attention raises the performance to 90.8% Dice and 93.7% accuracy, and progressive refinement enhances it further to 91.2% Dice and 94.2% accuracy. The complete HST-Net model achieves the highest performance with 91.6% Dice and 94.8% accuracy, which proves the necessity of each proposed module.

V. Discussion

The proposed HST-Net achieves an average Dice coefficient of 91.6%, which indicates a strong spatial agreement between the predicted segmentation and the ground truth annotations. A comparative statistical analysis of cardiac MRI segmentation methods is given in Table 6, which indicates a consistent increase in performance over time. The previous works like [1] based on classical and machine learning methods to perform right ventricle segmentation, showed Dice coefficients between 75.0 and 85.0, and a segmentation accuracy between 88.0 and 90.0, meaning moderate results. As deep learning is adopted, it is observed to improve a lot. As an example, [3] Dice and accuracy with CNN and shape priors were improved to 88.090.2 percent and 91.092.5 percent respectively with LV, RV and myocardium, which show better structural learning. On the same note, the hybrid CNN-based methods in [9] showed 86.0-90.0% Dice and 90.0-92.0% accuracy with multi-structure segmentation. However, U-Net and GAN-based architectures and [12] reached 90.0 -93.0% Dice and 92.0 -94.0% accuracy in ventricular segmentation, and [11] reached 90.5-92.3% Dice and 93.0-94.2% accuracy in left atrium segmentation. In [15], the Residual CNN using ASPP was also shown to compete with the 89.0 -91.2% Dice and 92.0 -93.5% accuracy of LV myocardium. More recent, [7] with M&Ms Challenge deep learning models obtained 87.091.0% Dice and 91.593.0% accuracy in right ventricle segmentation. Comparatively, the suggested HST-Net model that combines Hierarchical Spectrum Tokenization with PSR has a Dice coefficient of 91.6% and segmentation accuracy of 94.8% on LV, RV and myocardium. This shows that the suggested method not only performs better than the previous classical and hybrid, but also offers slight but significant gains over the current deep learning models., These results indicate that the model exhibits strong robustness, improved feature representation, and enhanced generalization capability for multi-structure cardiac MRI segmentationA Dice score above 90% in cardiac MRI segmentation is generally considered indicative of high clinical reliability, as it reflects minimal overlap error in cardiac structures such as the left ventricle, right ventricle, and myocardium. The achieved pixel-wise segmentation accuracy of 94.8% further confirms that the model correctly classifies the majority of image pixels, demonstrating robustness in both homogeneous regions and boundary areas. The observed 2.1–3.4% improvement in Dice score over existing models is statistically and clinically meaningful. In cardiac MRI analysis, even a 1% increase in Dice score can significantly reduce volumetric estimation errors, thereby improving the accuracy of derived clinical indices such as ejection fraction and myocardial mass [1], [2]. Similarly, the 1.9–2.6% gain in segmentation accuracy suggests enhanced consistency across slices and patients, reducing inter-subject variability. Specifically, the spectrum tokenization mechanism separates global structural

Table 5. Ablation study on proposed HST-Net

Configuration	Dice (%)	Accuracy (%)
Baseline Encoder–Decoder	86.9	90.8
+ Spectrum Tokenization	88.7	92.1
+ Hierarchical Encoding	89.6	92.8
+ Consistency Attention	90.8	93.7
+ Progressive Refinement	91.2	94.2
Full HST-Net (All Modules)	91.6	94.8

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

information from high-frequency boundary components, allowing the network to preserve anatomical consistency even when local intensity variations or noise are present. In addition, hierarchical multi-scale encoding and progressive refinement modules help suppress localized distortions introduced by motion artifacts.

As shown in Table 6 summarizes comparative statistical analysis with existing cardiac MRI segmentation studies. Early cardiac MRI segmentation approaches summarized by Petitjean et al. [1] reported Dice scores ranging from 75% to 85% for right ventricle segmentation, highlighting significant performance variability. HST-Net substantially outperforms these methods, demonstrating improved robustness to anatomical differences. CNN-based methods with shape priors, such as those proposed by Zotti et al. [3], improved Dice scores to approximately 88–90%; however, they relied on explicit constraints to preserve shape. In contrast, HST-Net achieves higher Dice scores through hierarchical spectrum tokenization without predefined shape assumptions. Deep learning models reviewed by Chen et al. [7] emphasized that conventional CNNs often struggle to capture global cardiac geometry. This limitation is reflected in their reported Dice scores below 90%. Recent GAN-based methods by Zhang et al. [9] improved segmentation performance to the 90–93% range but introduced training instability and higher computational cost. HST-Net achieves comparable or superior performance without adversarial learning. The GCW-UNet model by Wong et al. [10] achieved Dice scores up to 92.3% but was structure-specific, whereas HST-Net maintains consistent performance across multiple cardiac regions. Furthermore, challenge-based evaluations reported by Martín-Isla et al. [7] show that right ventricle segmentation remains difficult, with Dice values often below 91%, a limitation that HST-Net effectively overcomes.

In addition to overlap-based metrics such as Dice coefficient, boundary-based evaluation using Hausdorff Distance (HD) provides important clinical insight into segmentation reliability. The HST-Net model achieves a Hausdorff Distance of 3.12 mm. This implies a small maximum distance between the predicted boundary and the expert-provided ground truth. In the context of cardiac MRI segmentation, an HD of less than 5 mm is deemed acceptable from a clinical point of view because the deviation is negligible and does not influence the volume of the ventricles or the myocardium's thickness. The low Hausdorff Distance also implies the model's ability to precisely locate the boundary in the most challenging regions of the heart, such as the right ventricle and the myocardium. Precise boundary identification is a prerequisite in the subsequent processes of calculating the ejection fraction, myocardium volume, and other functional analyses of the heart. Compared with other models that achieve a Hausdorff Distance of more than 4-6 mm, the low HD of the HST-Net model implies a high level of anatomical consistency and robustness. The Hausdorff Distance of 3.12 mm implies the model's ability to

Table 6. Comparative Statistical Analysis with Existing Cardiac MRI Segmentation Studies

Author (Ref. No.)	Year	Method / Architecture	Target Structures	Dice Coefficient (%)	Segmentation Accuracy (%)
Petitjean et al. [1]	2015	Classical + ML methods (collation study)	Right ventricle	75.0–85.0	88.0–90.0
Zotti et al. [3]	2018	CNN with shape prior	LV, RV, Myocardium	88.0–90.2	91.0–92.5
Chen et al. [7]	2020	CNN / hybrid DL (reviewed models)	Multi-structure	86.0–90.0	90.0–92.0
Zhang et al. [9]	2022	U-Net / GAN (TLMDB-GAN)	Ventricles	90.0–93.0	92.0–94.0
Wong et al. [10]	2022	GCW-UNet	Left atrium	90.5–92.3	93.0–94.2
Martín-Isla et al. [6]	2023	M&Ms Challenge DL models	Right ventricle	87.0–91.0	91.5–93.0
Ahmad et al. [14]	2022	Residual CNN + ASPP	LV myocardium	89.0–91.2	92.0–93.5
HST-Net (Proposed)	—	Hierarchical Spectrum Tokenization + PSR	LV, RV, Myocardium	**91.6**	**94.8**

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

produce segmentation results well within the acceptable range from a clinical point of view.

Despite its strong segmentation performance, the proposed HST-Net exhibits certain limitations. The incorporation of hierarchical spectrum-tokenization and progressive refinement modules increases computational complexity compared to conventional CNN-based architectures such as U-Net. In our experimental setup (NVIDIA RTX 3090 GPU, batch size = 8), HST-Net required approximately 1.8× longer training time per epoch (≈ 145 seconds) compared to the baseline U-Net (≈ 80 seconds). The total training time for 200 epochs was approximately 8.1 hours, whereas the baseline converged in 4.5 hours under identical conditions. Additionally, HST-Net consumed ~11.2 GB of GPU memory, compared to ~6.7 GB for the baseline model, reflecting a 67% increase in memory usage. These results highlight a clear accuracy–efficiency trade-off introduced by the hierarchical design

The improved segmentation accuracy achieved by HST-Net has important clinical and research implications. Accurate and reliable cardiac structure delineation is essential for quantitative ventricular analysis, disease diagnosis, and treatment planning [2], [4]. Automated segmentation frameworks with Dice scores exceeding 90% can significantly reduce manual annotation time and inter-observer variability, enabling large-scale clinical studies and routine clinical deployment [6], [8]. Moreover, the demonstrated robustness in right ventricle and myocardial segmentation directly addresses key challenges identified in prior studies [1], [7]. The hierarchical and anatomically consistent design of HST-Net provides a strong foundation for future extensions to multi-modal cardiac imaging and longitudinal studies, supporting more precise and personalized cardiovascular assessment.

VI. Conclusion

This study addresses the persistent challenge of providing consistent cardiac structure segmentation in MRI, which includes anatomical complexity and variability of imaging parameters. Traditional deep learning methods tend to lose sight of local structural continuity while delineating finer boundaries, in thin tissue with irregular morphology. To address these challenging issues, this work proposed a spectrum-aware and hierarchy-driven segmentation framework that explicitly separated and coordinated structure and boundary information through multi-scale learning and refinement. Experimental results show that the proposed algorithm gives strong quantitative performance. It has 91.6% Dice coefficient, 94.8% segmentation accuracy, and Hausdorff Distance of 3.12 mm. These results confirm the integration of frequency-aware feature decomposition, hierarchical consistency and progressive refinement. It boosts the precision and anatomical coherence of the cardiac segmentation results. Future validation of HST-Net will therefore involve cross-institutional evaluation using multi-center datasets containing heterogeneous patient populations and acquisition conditions. To improve robustness under these variations, domain generalization strategies such as intensity normalization, scanner-invariant feature learning, and advanced data augmentation incorporating simulated noise, motion artifacts, and resolution variations will be investigated. Additionally, transfer learning and domain adaptation techniques will be employed to fine-tune the model using limited labeled data from new clinical centers while preserving previously learned anatomical representations.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The data will be available based on request.

Author Contribution

Naga Chandrika G. contributed to the conceptualization of the study, development of the methodology, and drafting of the initial manuscript. Shamia D. was responsible for data collection, preprocessing, and implementation of the experimental work. V. Kavithamani contributed to model development, validation, and performance analysis. Amitha Ida Chandran assisted in data interpretation, result analysis, and manuscript editing. K. Venu contributed to visualization, preparation of figures, and technical refinement of the manuscript. Kunchanapalli Rama Krishna provided overall supervision, critical review of the manuscript, and guidance throughout the research. All authors read and approved the final version of the manuscript.

Declarations

Ethical Approval

Not Applicable.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] Petitjean, C., Zuluaga, M. A., Bai, W., Dacher, J. N., Grosgeorge, D., Caudron, J., ... & Yuan, J. (2015). Right ventricle segmentation from cardiac MRI: a collation study. Medical image

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

analysis, 19(1), 187-202. https://doi.org/10.1016/j.media.2014.10.004

[2] Wei, D., Li, C., & Sun, Y. (2015). Medical image segmentation and its application in cardiac MRI. Biomedical Image Understanding, 47-89. https://doi.org/10.1002/9781118715321.ch2

[3] Zotti, C., Luo, Z., Lalande, A., & Jodoin, P. M. (2018). Convolutional neural network with shape prior applied to cardiac MRI segmentation. IEEE journal of biomedical and health informatics, 23(3), 1119-1128. https://doi.org/10.1109/JBHI.2018.2865450

[4] Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E., & Frangi, A. F. (2016). A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine, 29(2), 155-195. https://doi.org/10.1007/s10334-015-0521-4

[5] Earls, J. P., Ho, V. B., Foo, T. K., Castillo, E., & Flamm, S. D. (2002). Cardiac MRI: recent progress and continued challenges. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 16(2), 111-127. https://doi.org/10.1002/jmri.10154

[6] Martín-Isla, C., Campello, V. M., Izquierdo, C., Kushibar, K., Sendra-Balcells, C., Gkontra, P., ... & Lekadir, K. (2023). Deep learning segmentation of the right ventricle in cardiac MRI: the M&Ms challenge. IEEE Journal of Biomedical and Health Informatics, 27(7), 3302-3313. https://doi.org/10.1109/JBHI.2023.3267857

[7] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., & Rueckert, D. (2020). Deep learning for cardiac image segmentation: a review. Frontiers in cardiovascular medicine, 7, 25. https://doi.org/10.3389/fcvm.2020.00025

[8] Dataset collection: https://www.kaggle.com/datasets/danialsharifrazi/cad-cardiac-mri-dataset

[9] Zhang, Y., Feng, J., Guo, X., & Ren, Y. (2022). Comparative analysis of U-Net and TLMDB GAN for the cardiovascular segmentation of the ventricles in the heart. Computer Methods and Programs in Biomedicine, 215, 106614. https://doi.org/10.1016/j.cmpb.2021.106614

[10] Wong, K. K., Zhang, A., Yang, K., Wu, S., & Ghista, D. N. (2022). GCW-UNet segmentation of cardiac magnetic resonance images for evaluation of left atrial enlargement. Computer Methods and Programs in Biomedicine, 221, 106915. https://doi.org/10.1016/j.cmpb.2022.106915

[11] Morris, E. D., Ghanem, A. I., Dong, M., Pantelic, M. V., Walker, E. M., & Glide-Hurst, C. K. (2020). Cardiac substructure segmentation with deep learning for improved cardiac sparing. Medical physics, 47(2), 576-586. https://doi.org/10.1002/mp.13940

[12] Mortazi, A., Karim, R., Rhode, K., Burt, J., & Bagci, U. (2017, September). CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 377-385). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-66185-8_43

[13] Zhang, X., Noga, M., & Punithakumar, K. (2020). Fully automated deep learning based segmentation of normal, infarcted and edema regions from multiple cardiac MRI sequences. In Myocardial pathology segmentation combining multi-sequence CMR challenge (pp. 82-91). Cham: Springer International Publishing. https://doi.org/10.48550/arXiv.2008.07770

[14] Ahmad, I., Qayyum, A., Gupta, B. B., Alassafi, M. O., & AlGhamdi, R. A. (2022). Ensemble of 2D residual neural networks integrated with atrous spatial pyramid pooling module for myocardium segmentation of left ventricle cardiac MRI. Mathematics, 10(4), 627. https://doi.org/10.3390/math10040627

[15] Romaguera, L. V., Costa, M. G. F., Romero, F. P., & Costa Filho, C. F. F. (2017, March). Left ventricle segmentation in cardiac MRI images using fully convolutional neural networks. In Medical Imaging 2017: Computer-Aided Diagnosis (Vol. 10134, pp. 760-770). SPIE. https://doi.org/10.1117/12.2253901

[16] Ammari, A., Mahmoudi, R., Hmida, B., Saouli, R., & Bedoui, M. H. (2021). A review of approaches investigated for right ventricular segmentation using short-axis cardiac MRI. IET Image Processing, 15(9), 1845-1868. https://doi.org/10.1049/ipr2.12165

[17] Zhang, T., Li, A., Wang, M., Wu, X., & Qiu, B. (2019). Multiple attention fully convolutional network for automated ventricle segmentation in cardiac magnetic resonance imaging. Journal of Medical Imaging and Health Informatics, 9(5), 1037-1045. https://doi.org/10.1166/jmihi.2019.2685

[18] Kravchenko, D., Isaak, A., Mesropyan, N., Peeters, J. M., Kuetting, D., Pieper, C. C., ... & Luetkens, J. A. (2025). Deep learning super-

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

resolution reconstruction for fast and high-quality cine cardiovascular magnetic resonance. European Radiology, 35(5), 2877-2887. https://doi.org/10.1007/s00330-024-11145-0

[19] Jiang, C., Wang, Y., Yuan, Q., Qu, P., & Li, H. (2025). A 3D medical image segmentation network based on gated attention blocks and dual-scale cross-attention mechanism. Scientific Reports, 15(1), 6159. https://doi.org/10.1038/s41598-025-90339-y

[20] Lu, Y., Zhao, Y., Chen, X., & Guo, X. (2022). A Novel U-Net Based Deep Learning Method for 3D Cardiovascular MRI Segmentation. Computational Intelligence and Neuroscience, 2022(1), 4103524. https://doi.org/10.1155/2022/4103524

[21] Suganyadevi, S., Pershiya, A. S., Balasamy, K., et al. "Deep learning based alzheimer disease diagnosis: A comprehensive review". *SN Computer Science*, Vol.5 no.4, pp.391, 2024, https://doi.org/10.1007/s42979-024-02743-2.

[22] Balasamy, K., Krishnaraj, N., & Vijayalakshmi, K. "An adaptive neuro-fuzzy based region selection and authenticating medical image through watermarking for secure communication", *Wireless Personal Communications*, Vol.122, no.3, pp. 2817–2837, 2021. https://doi.org/10.1007/s11277-021-09031-9.

[23] Suganyadevi, S., & Seethalakshmi, V. "CVD-HNet: Classifying Pneumonia and COVID-19 in Chest X-ray Images Using Deep Network". *Wireless Personal Communications*, Vol.126, no. 4, pp.3279–3303, 2022. https://doi.org/10.1007/s11277-022-09864-y.

[24] Balasamy, K., & Suganyadevi, S. "Multi-dimensional fuzzy based diabetic retinopathy detection in retinal images through deep CNN method". *Multimedia Tools and Applications*, Vol 83, no. 5, pp.1–23. 2024. https://doi.org/10.1007/s11042-024-19798-1.

[25] Shamia, D., Balasamy, K., and Suganyadevi, S. "A secure framework for medical image by integrating watermarking and encryption through fuzzy based roi selection", *Journal of Intelligent & Fuzzy systems*, 2023, Vol. 44, no.5, pp.7449-7457. https://doi.org/10.3233/JIFS-222618.

[26] Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey. *Wireless Pers Commun* 137, 1685–1714 (2024). https://doi.org/10.1007/s11277-024-11428-1.

[27] Suganyadevi, S., Seethalakshmi, V. Deep recurrent learning based qualified sequence segment analytical model (QS2AM) for infectious disease detection using CT images. *Evolving Systems* 15, 505–521 (2024). https://doi.org/10.1007/s12530-023-09554-5.

[28] T. Gopalakrishnan, S. Ramakrishnan, K. Balasamy and A. S. Muthananda Murugavel, "Semi fragile watermarking using Gaussian mixture model for malicious image attacks," 2011 World Congress on Information and Communication Technologies, Mumbai, India, 2011, pp. 120-125. https://doi.org/10.1109/WICT.2011.6141229.

[29] Renuka Devi, K., Suganyadevi, S and Balasamy, K.. "Healthcare Data Analysis U sing Deep Learning Paradigm ". *Deep Learning for Cognitive Computing Systems: Technological Advancements and Applications*, edited by M.G. Sumithra, Rajesh Kumar Dhanaraj, Celestine Iwendi and Anto Merline Manoharan, Berlin, Boston:De Gruyter, 2023, pp. 129–148. https://doi.org/10.1515/9783110750584-008.

[30] M. Hossin, M.N. Sulaiman, A review on evaluation metrics for data classification evaluations, Int. J. Data Min. Knowl. Manag. Process. 5 (2) (2015) 1. https://doi.org/10.5121/ijdkp.2015.5201.

[31] Liu, Yu, Gabriella Captur, James C. Moon, Shuxu Guo, Xiaoping Yang, Shaoxiang Zhang, and Chunming Li. "Distance regularized two level sets for segmentation of left and right ventricles from cine-MRI." Magnetic resonance imaging 34, no. 5 (2016): 699-706, https://doi.org/10.1016/j.mri.2015.12.027

[32] Queirós, Sandro, Daniel Barbosa, Brecht Heyde, Pedro Morais, João L. Vilaça, Denis Friboulet, Olivier Bernard, and Jan D'hooge. "Fast automatic myocardial segmentation in 4D cine CMR datasets." Medical image analysis 18, no. 7 (2014): 1115-1131, https://doi.org/10.1016/j.media.2014.06.001

[33] Hu, Huaifei, Haihua Liu, Zhiyong Gao, and Lu Huang. "Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming." Magnetic resonance imaging 31, no. 4 (2013): 575-584, https://doi.org/10.1016/j.mri.2012.10.004

[34] Ringenberg, Jordan, Makarand Deo, Vijay Devabhaktuni, Omer Berenfeld, Pamela Boyers, and Jeffrey Gold. "Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI." Computerized Medical Imaging and Graphics 38, no. 3 (2014): 190-201, https://doi.org/10.1016/j.compmedimag.2013.12.011

[35] Rosen, Boaz D., Thor Edvardsen, Shenghan Lai, Ernesto Castillo, Li Pan, Michael Jerosch-Herold, Shantanu Sinha et al. "Left ventricular

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

concentric remodeling is associated with decreased global and regional systolic function: the Multi-Ethnic Study of Atherosclerosis." Circulation 112, no. 7 (2005): 984-991, https://doi.org/10.1161/CIRCULATIONAHA.104.500488

[36] Qian, Xiaohua, Yuan Lin, Yue Zhao, Jing Wang, Jing Liu, and Xiahai Zhuang. "Segmentation of myocardium from cardiac MR images using a novel dynamic programming based segmentation method." Medical physics 42, no. 3 (2015): 1424-1435, https://doi.org/10.1118/1.4907993

[37] Zhang, Hongyang, Wenxue Zhang, Weihao Shen, Nana Li, Yunjie Chen, Shuo Li, Bo Chen, Shijie Guo, and Yuanquan Wang. "Automatic segmentation of the cardiac MR images based on nested fully convolutional dense network with dilated convolution." Biomedical signal processing and control 68 (2021): 102684, https://doi.org/10.1016/j.bspc.2021.102684

[38] Abdeltawab, Hisham, Fahmi Khalifa, Fatma Taher, Norah Saleh Alghamdi, Mohammed Ghazal, Garth Beache, Tamer Mohamed, Robert Keynton, and Ayman El-Baz. "A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images." Computerized medical imaging and graphics 81 (2020): 101717, https://doi.org/10.1016/j.compmedimag.2020.101717

[39] Fu, Fan, Jianyong Wei, Miao Zhang, Fan Yu, Yueting Xiao, Dongdong Rong, Yi Shan et al. "Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network." Nature communications 11, no. 1 (2020): 4829, https://doi.org/10.1038/s41467-020-18606-2

[40] Fu, F., J. Wei, M. Zhang, F. Yu, Y. Xiao, D. Rong, Y. Shan et al. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. Nat Commun 11, 4829. 2020. https://doi.org/10.1038/s41467-020-18606-2

Author Biography

Naga Chandrika Gogulamudi currently working as Senior Assistant Professor in the department of Information Technology at VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad-500090, Telangana, India. She received B.Tech degree in Computer Science and Information Technology from JNTU, Hyderabad and M.Tech in Software Engineering from JNTU, Anantapur and Ph.D in Computer Science and Engineering from ANU, Guntur in 2022. She has published around 25 papers in international journals and conferences. She is serving as a reviewer for reputed journals like Computers, Materials and Continua, Computer Systems Science and Engineering. Her areas of interest include Social Network Analysis, Machine Learning, Medical Imaging, and Image Processing.

Dr. D. Shamia is an Associate Professor in the Department of Electronics and Communication Engineering at V.S.B. College of Engineering Technical Campus, Ealur Pirivu, Solavampalayam (PO), Kinathukadavu, Coimbatore, Tamil Nadu, India. She completed her B.E. in Electronics and Instrumentation Engineering and M.Tech. in Applied Electronics from Karunya University, Coimbatore, and was awarded a Ph.D. in Electronics and Communication Engineering in 2019. With over 15 years of teaching and research experience, she has contributed significantly to academics and technical mentoring. She is a Life Member of ISTE. Her research interests include image and face recognition, medical image analysis, embedded systems, Internet of Things (IoT), and artificial intelligence, with a focus on practical and interdisciplinary applications.

Ms. V. Kavithamani is currently working as an Assistant Professor in the Department of Electronics and Communication Engineering at Jai Shriram Engineering College. She holds a Master of Engineering degree in Power Systems from SNS College of Technology, Coimbatore, and a Bachelor of Engineering in Electrical and Electronics Engineering from Velalar College of Engineering and Technology, Erode. With a total of 14 years of teaching experience, she has established strong expertise across multiple domains. Her academic and research interests include Digital Image Processing, Signal Processing, Machine Learning, and Artificial Intelligence, through which she continues to guide students and contribute to technological advancements.

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485

Dr. Amitha Ida Chandran is an Associate Professor in the Department of Computer Science and Engineering at KPR Institute of Engineering and Technology, located at Avinashi Road, Arasur, Coimbatore, Tamil Nadu, India. She holds a Ph.D. in Information Technology with a specialization in Image Processing and brings over 16 years of rich teaching and research experience. Her research interests encompass computer vision, deep learning, image analysis, and intelligent systems. Dr. Amitha has published several research papers in reputed international journals and conferences and actively contributes to the academic community through professional body memberships. She is deeply committed to academic excellence, innovative and student-centered teaching methodologies, and conducting impactful research that addresses real-world technological challenges

Dr. K. Venu is currently serving as an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering at Kongu Engineering College, Tamil Nadu, India. She completed her Ph.D. in Deep Learning under Anna University, where her research focused on advanced learning models and intelligent computing techniques. With over 9 years of teaching experience, she has actively contributed to both academic instruction and research mentoring. She has published 7 research articles in reputed international journals and presented 11 papers at international and national conferences. In addition, she has authored 3 book chapters with well-known publishers. Her research interests include Machine Learning, Deep Learning, and Nature-Inspired Computing, with applications in intelligent and data-driven systems.

Dr. Kunchanapalli Rama Krishna is a Professor in the CSIT Department at KL University with over 33 years of academic and research experience in Computer Science & Engineering. He holds a Ph.D. in CSE and has served at reputed institutions including Galgotias University, C-DAC, and IIMT College of Engineering. As an AICTE Margdarshak and NAAC e-Assessor, he has guided institutions toward NBA accreditation and academic excellence. He has published over 30 research papers in reputed national and international journals and conferences. Dr. Rama Krishna has also led funded projects and national conferences on cryptography and modernizing education. He actively participates in FDPs and holds certifications from IITs, IIMs, and NPTEL/SWAYAM. His guidance has shaped numerous UG, PG, and Ph.D. scholars, contributing to quality education and innovation.

Manuscript received 19 January 2026; Revised 26 March 2026; Accepted 2 April 2026; Available online 11 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1485