Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CHSEGNET: COMPOSITE HEART SEGMENTATION NETWORK FOR CARDIAC IMAGE SEGMENTATION

NANDHAGOPAL SUBARAMANI[1], E SASIKALA[2]

[1,2]Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu 603203, India

Email : cyberphd85@gmail.com

ABSTRACT

Cardiac MRI is vital in following the disease trajectory and treatment response. By providing detailed, quantitative data about the anatomy and function of the heart, cardiac MRI allows clinicians to accurately measure cardiac function is an invaluable resource for identifying disease progression and response to therapy. Segmentation employing DL allowed us to accurately contour the myocardial, right ventricle and left ventricle. The detailed cutting not only supports the diagnosis of heart failure, but also the understanding of the condition of the heart's operations following therapy and the response to therapies. Our objective is to create a Composite Heart Segmentation Network (CHSegNet) that leverages MRI imaging data to segment cardiac organs. The complex architecture of the heart with its many chambers, arteries, and tissues makes it difficult to achieve high segmentation effectiveness. Furthermore, the heart itself beats, inducing motion artifacts that degrade the quality and reliability of images. Consequently, a CHSegNet approach is investigated to address several issues in cardiac MRI image segmentation. A CHSegNet system is proposed. It's built upon the encoder-decoder pair that boosts cardiac semantic segmentation effectiveness with the dense layer. To maintain a spatial information, information from heavily convolutional regions of various sizes is combined by the dense level. The densely connected layer, enables the network to deal with variables at various spatial scales to record the worldwide as well as local context. This retains the cardiac structural detail and context required for high-fidelity segmentation of the larger cardiac architectures. To enable the deep connection to learn important semantic variables, the encoder's higher level receives the dense layer. With the roll of an intersection over union (IoU), and precision scores of 78.1%, 97.9% and accuracy of 99.5% correspondingly, the encoder layer generated the best effectiveness for segmentation according to an investigation of several situations and structural modifications. The assignment of a shallower stratum like more geographic information is preserved by encoder layer necessitating greater parallelism routes to capture the ideal set of variables since different configurations have varied effects on the network. Conversely, deeper layers, have fewer parallel routes needed to provide the best network because they have more useful characteristics but a lower spatial resolution.

Keywords- *Cardiac Segmentation, Deep Learning, Prediction, Dense Layer, Accuracy*

.

1. INTRODUCTION

In most industrialized countries and many creating countries, cardiovascular diseases such as cardiovascular disease, coronary artery disease, and cardiovascular diseases are the leading causes of death for people. Particularly in the elderly, heart disorders have a major impact on the rise in healthcare costs that lead to substantial disability and lost productivity [1]. To lower the death rate and enhance treatment results, early intervention and accurate diagnosis are essential. Healthcare professionals can greatly enhance treatment outcomes and lower healthcare costs by finding the risk variables and putting preventative measures like medication and lifestyle changes into practice [2]. The magnetic resonance imaging of the heart, or cardiac MRI, has become a useful technique for precisely diagnosing the condition. Cardiac MRI helps clinical professionals diagnose heart conditions early, evaluate the severity of the condition, and track the effectiveness of treatment by giving precise images of the anatomy and function of the heart [3]. To maximize the patient's treatment results, this information is essential for choosing the right therapy. Cardiomyopathies, infarction of the myocardium, and coronary artery disease can all be diagnosed with the help of cardiac magnetic resonance imaging (MRI) [4], which offers

information on the LV, RV, Myo heart MRI can offer comprehensive details regarding cardiovascular chamber size, blood circulation, and tissue in the heart characteristics by employing sophisticated imaging sequences [5].

Cardiac MRI's superior non-ionizing electromagnetic radiation, delicate tissue comparison, and advanced functional imaging characteristics are its primary benefits over other imaging modalities. Eliminating the need for CT scans or other ionizing-based diagnostics, cardiac MRI is excellent at visualizing the characteristics of myocardial tissue, allowing for in-depth evaluations of tissues that are scarred edema, and rigidity [6]. It is especially important for patients who are pregnant or young, as radiation exposure may have negative effects. Furthermore, high-resolution imagery is provided by cardiac MRI, which is less affected by individual characteristics and environmental elements including insufficient acoustic apertures and being overweight, which can lower the efficacy of echocardiography [7]. It is a trustworthy equipment for identifying and evaluating a range of cardiovascular issues because of these benefits. Echocardiography's intrinsic variability is outweighed by the reliability and impartiality of quantifiable data from cardiac MRI, such as ejection fraction and ventricular measures [8]. Cardiac MRI is the gold standard in cardiovascular diagnoses and research because it provides complete and safe imaging, in contrast to nuclear photography has limited spatial resolution and radioactive issues, yet it offers metabolic information. Additionally, cardiac MRI provides a complete tool for tracking the course of heart disease and the effectiveness of treatment. Additionally, it offers accurate and trustworthy information regarding both ischemia and non-ischemic structural damage related to cardiomyopathy [9]. Its significance in supporting considering healthcare decisions and improving treatment of patients is facilitated by its non-invasiveness, superior spatial accuracy and the ability to assess myocardial sustainability, perfusion, and physiological characteristics [10]. Additionally, cardiac MRI's non-invasiveness lowers the risk of repeated imaging, making it a dependable and safe method for tracking the course of a disease over time and assessing the efficacy of therapeutic therapies. Clinicians can evaluate the reaction to surgical procedures, rehabilitation for cardiac disease, or medication therapy and its capacity to monitor alterations in tissue properties and cardiac function throughout aging. This routine identifying aids in improving patient care and customizing treatment plans to meet the needs of each patient [11].

DL has made tremendous progress in recent years, especially in the area of clinical picture processing. The creation of autonomous semantics differentiation methods based on artificial neural networks is one noteworthy achievement that has completely changed the automatic analysis of clinical images [12]. Assigning the appropriate physiological architecture like the cardiovascular system, blood arteries, ligaments, lung capacity, and optic disc to every pixel in a picture is the main objective of the semantic segmentation [13]. The Fully Convolutional Network (FCN) is one of the early architectural pillars in this discipline. The FCN has a special architecture with two main parts: a pair of symmetric encoders and decoders. The task of obtaining pertinent information from the input image falls to the encoder pathway. This is accomplished employing a sequence of convolutional layers that increase the complexity of the retrieved variables while gradually reducing the image's spatial dimensions. The decoder circuit then gradually reconstructs the original image resolution by upsampling these variables [14]. The FCN generates precise pixel-wise segmentation maps by integrating high-level semantic data with low-level characteristics. The FCN's ability to automatically identify critical cardiac areas such as the LV, RV, and myocardial to interpret pictures of the detecting process is revolutionized by arbitrary size of heart illness utilizing cardiac MRI. The FCN has shown outstanding segmentation effectiveness in cardiac MRI imaging. This DL technique can identify irregularities in size, appearance, or form that could be early indicators of cardiac disease by segmenting cardiac architectures. The U-Net architecture was developed by [15], a distinctive design for networks that greatly expanded the classification of clinical photographs, building upon the fundamental work of FCNs. The usage of skip connections, which make it easier to combining low-level parameter information from the encoding path with higher-level semantic data from the receiver's path, is what makes the U-Net unique. Particularly in situations with little training data, this architectural innovation improves segmentation accuracy by allowing the system is designed to record both fine-grained and general context variables [16].

A range of heart chamber architectures in terms of size and form can be seen in cardiac MRI images. One-image scale or small compression kernel's receptive field is comparatively fixed, meaning that it can only effectively represent image information within a certain range of receptive fields. As a result, it cannot adequately represent the borders of heart chamber architectures of different sizes. By utilizing

data at various spatial resolutions and scales, the challenges of splitting cardiac MRI images can be addressed with the use of multi-scale DL algorithms [17]. This enables the segmentation technique to comprehend the general anatomy of the heart while also employing on more minute details. While multi-scale DL methods can greatly boost segmentation accuracy, they frequently result in higher computational demands and more complex patterns [18]. High-effectiveness hardware is needed because multi-scale DL patterns sometimes use numerous simultaneous network paths to display data at various scales, increasing computing needs. Hyper-parameters and subcomponent components must be carefully adjusted to create an efficient multi-scale DL pattern [19] – [20]. Extended development periods and greater difficulties in identifying the optimal design may result from this complexity. In order to accurately automatically segment cardiac MRI images, this research suggests a DL pattern that makes use of a tightly connected, multi-scale convolutional circuit. The learning architecture presented is employed by the suggested pattern Composite Heart Segmentation Network (CHSegNet) to extract different spatial dimensions and sensitivities from generative blocks with dense connections. The following are this research's primary contributions:

1. A new Composite Heart Segmentation Network (CHSegNet), an automated cardiac MRI classification system constructed around a fully convolutional and multi-scale densely linked neural network.
2. A thorough assessment of CHSegNet with various SPP topologies and multi-scale dense prevent placement sites employing the ACDC cardiac MRI database to look at the pattern's capacity for generalization.
3. A comparison of the effectiveness of the multi-scale architecture with that of the existing multi-scale architectures.
4. A comparison between CHSegNet and the most advanced semantic segmentation networks.

There are five primary parts in this research. The CHSegNet architecture is covered in the next section. The laboratory settings and the information sets used in this research are explained in depth in the third section. The results are discussed in depth in Section 4, and a succinct conclusion is given in the fifth subsection.

2. RELATED WORKS

Numerous techniques have been put out in the literature to enhance the diagnosis of LVH due to the significance of early care of this cardiac condition. For example, Hussain et al. [21] employed an AI technique to identify the clinical traits associated with genuine and ECG false positives. The objective of this research was to improve the technique's ability to identify HCM and to improve its use in clinical situations. In a different research [22], authors created multivariable logistic scores to diagnose increasing LVM and global wall thickness employing both traditional and sophisticated ECG measurements. Advanced ECG ratings demonstrated unique electrophysiological signatures for each disease, outperforming traditional criteria. Additionally, utilizing ECG and echocardiographic data, to detect LV dilation and LVH, Lin in [23] developed DL and AI algorithms (logistic regression and randomized forest). According to their findings, DL techniques significantly outperformed machine learning techniques in terms of the AUROC for LVH detection. The author combined a randomized decision forests and a neural network made of computers to create a diagnostic pattern for HCM. To forecast the disease and facilitate early identification and intervention, this pattern included forty genes that were found in both the training and verification sets. To identify LVH from CMR pictures captured in both long-axis and short-axis perspectives. Li in [24] created automated methods employing a three-dimensional version of ResNet. In an effort to improve diagnostic accuracy, the network outputs a decision about whether hypertrophy is present or not. In addition, the investigator in [25] presented a DL method that minimized weighted mean square error loss employing a DeepLabv3 architecture that has been tweaked and educated on parasternal long-axis images. Accurately measuring LVH and forecasting the primary reason for the thickening of the LV wall were the objectives. In order to improve the diagnostic accuracy of LVH, Zhao in [26] created a late averaging fusion neural network-based heterogeneous deep learning structure to identify the cause of this genetic cardiac condition. Furthermore, a CNN pattern under the TensorFlow method was proposed to help diagnose LVH by measuring posterior wall thickness. A CNN-LSTM was created for the differentiation of three typical causes of LVH (light chain cardiac amyloidosis, hypertensive heart disease, and HCM).

In a recent work [27], the authors evaluated the pretest likelihood of identifying the source of LVH using machine learning techniques including the use

of support vector machines, randomly generated forests, and decision tree methods. Predicting whether left ventricular hypertrophy was caused by hypertension was the specific focus of this work. A number of studies, including those in [27], employed variables R-wave and S-wave amplitudes, beats per minute, time intervals, and electrical axes from ECGs are used to train machine and deep learning patterns like logistic regression, random forest, LGBM, ResNet, and CNN, to improve LVH prediction. The collected results demonstrated that machine learning patterns surpassed CNN patterns that were directly learned on ECG data and surpassed conventional ECG standards for identifying and forecasting LVH. In related research, A DL pattern was developed by [28] to calculate LV mass from heart magnetic resonance imaging (CMR) employing ECG data in order to assess the technique's efficacy in identifying LVH. By extracting CMR morphological data, a DL technique was proposed to improve HCM mutation-risk assessment. This research employed an unimproved four-chamber perspective of movies to categorize the HCM genotypes. In an alternate setting, Parlati in [29] investigated left cardiovascular non-compaction (LVNC), an uncommon cardiomyopathy marked by an extremely thick and squishy left cardiovascular wall due to trabeculae enlargement. The objective was to use a DL-based technique (2D-U-Net) to calculate the proportion of the trabecular volume. The majority of the suggested research has demonstrated that DL patterns are highly accurate in diagnosing LVH, either by identifying its existence or by distinguishing the causes of this ventricular disease. For regular clinical integration, these patterns still need to be improved. Furthermore, it is still difficult to precisely determine the degree and severity of LVH in clinical practice [30]. To guide treatment choices and estimate prognosis which is imperative where the obstacle need to be overcome

Table 1: Comparative analysis on various existing approaches

Ref.	Authors	Data Modality	Methodology / Model Used	Objective	Key Findings
[21]	Maanja et al.	ECG	AI-based classification	Identify clinical traits of true vs. false-positive ECG cases; improve HCM detection	Improved identification of HCM and enhanced clinical applicability
[22]	Jamali et al.	ECG	Multivariable logistic regression (traditional & advanced ECG metrics)	Diagnose increased LVM and global wall thickness	Advanced ECG metrics outperformed traditional ECG criteria
[23]	Lin et al.	ECG & Echocardiography	DL models, Logistic Regression, Random Forest	Detect LV dilation and LVH	DL models significantly outperformed ML methods in AUROC
		Genetic data	Randomized decision forests + Neural networks	Predict HCM using genetic patterns	Forty genes identified for early diagnosis and intervention
[24]	Li et al.	CMR (Long- & Short-axis)	3D ResNet	Automated LVH detection	Improved diagnostic accuracy for hypertrophy presence
[25]	Hussain et al.	Echocardiography (PLAX images)	Modified DeepLabv3	Measure LVH and identify cause of LV wall thickening	Achieved accurate LVH quantification

Ref.	Authors	Data Modality	Methodology / Model Used	Objective	Key Findings
[26]	Zhao et al.	Multimodal cardiac data	Late-averaging fusion NN (heterogeneous DL)	Identify genetic cause of LVH	Improved diagnostic accuracy over standalone models
		Echocardiography	CNN (TensorFlow-based)	Measure posterior wall thickness	Enhanced LVH diagnosis
		Echocardiography	CNN–LSTM	Classify LVH causes (Amyloidosis, HHD, HCM)	Effective differentiation of LVH etiologies
[27]	Baptisa et al.	Clinical & ECG data	SVM, Random Forest, Decision Tree	Identify hypertension-induced LVH	Accurate prediction of LVH etiology
[19–21]	Multiple studies	ECG	Logistic Regression, RF, LGBM, ResNet, CNN	Improve LVH detection	ML models outperformed CNNs and conventional ECG criteria
[28]	Budai et al.	ECG & CMR	Deep Learning	Estimate LV mass and detect LVH	Improved LVH identification and mutation-risk assessment
[29]	Parlati et al.	Cardiac MRI	2D U-Net	Quantify trabecular volume in LVNC	Accurate assessment of LV non-compaction
[30]	Li et al.	Clinical practice	—	Assess LVH severity and prognosis	Precise grading of LVH remains challenging

3. METHODOLOGY

3.1. Dataset

To streamline the heart segmentation method and increase the accuracy of cardiac MRI interpretation, this research fitted DL patterns employing the ACDC databases. Clinical examinations obtained at the University Hospital of Dijon were employed to compile the ACDC database. It comprises 100 patients' cardiac MRI scans, which are split into five equally spaced diagnostic categories: heart attack, cardiomyopathy that is hypertrophic, dilated cardiomyopathy, abnormal right ventricular function, and healthy individuals. By ensuring a balanced database that represents a variety of cardiac diseases, this stratification enhances the design was trained on its generalizability thorough technique assessment for a variety of pathogenic and anatomical manifestations is supported by the inclusion of both pathological and normal instances. The objective of patient selection was to preserve good data quality while maximizing variety in heart function, geometry, and pathology. With a uniform imaging strategy that includes end-diastolic and MRI scans are anonymous, with end-systolic images for both short-axis perspectives. To ensure a reliable assessment, the database additionally contains ground truth subdivisions for the heart, the left ventricle and right ventricle with expert comments. The created cardiac-related automation technologies can be validated for therapeutic relevance employing these annotated databases. Secondary data from the ACDC database can be employed for controlled research without requiring direct ethical permissions for patient engagement. It is therefore a helpful instrument for enhancing automated cardiac MRI analysis. Several important characteristics of the ACDC cardiac MRI databases are as follows: Diverse imaging situations: Pictures of people with a variety of ages, physical types, and heart conditions are included in the ACDC databases,

which represent real-world clinical settings. Standardized imaging methods: Authors can more easily and reliably assess the effectiveness of their techniques because Standardized imaging methods were used to acquire the cardiac MRI information contained in ACDC databases.

- ✓ Variability in volume and timing based on the databases contain pictures of different stages of the cardiac cycle, recording the diastolic and systolic phases, which is essential for simulating the cardiovascular circulation is variable.
- ✓ The database offers more data kinds, such as clinical characteristics and patient demographics that can be utilized to enhance AI-driven systems' diagnostic capabilities.
- ✓ To ensure a high-quality labeling procedure, The ACDC database's fundamental comments follow a certain convention and protocol. Examples of the visuals and their corresponding ground truths are displayed.

Here, 30 cases of normal patients, thirty cases of DCM, thirty cases of HCM, thirty cases of heart failure with MINF, and thirty cases of RVA comprise the ACDC cardiac MRI database. The sample cardiac MRI scan and associated GT image are displayed. The database was divided between 50 testing cases (without ground truth) and 100 training examples (with ground truth). For reference segmentations of the LV, RV, and Myo, the ground truth is given in both ED and ES frames.

3.2. Prediction pattern

Composite Heart Segmentation Network (CHSegNet) is an integrated technique for segmenting clinical images that combines transformer-based and convolutional processes to increase segmentation accuracy. Employing an RGB image as input, the pattern generates a segmented output that divides the image into meaningful parts by classifying each pixel into a certain category. The MSA layer is crucial to capturing the original transformer-based pattern's global dependencies. Nevertheless, the depth-wise spatial correlations required for precise clinical picture segmentation are not captured by this method. Due to the pattern's inability to maintain fine-grained spatial hierarchies while utilizing global context for precise segmentation, this constraint may lead to less-than-ideal variable depiction. CNNs are employed by the encoder to initially extract local variables. A combination block that blends composites mechanisms for attention with depth-wise transformations, then processes these variables. The pattern's capacity to depict dependencies that are both local and intermediary is greatly enhanced by this combination. By more effectively handling global dependencies, the depth-wise attention block enhances variable integration and increases the pattern's overall performance in challenges involving categorization. This improves the resilience of the pattern, especially when it comes to segmenting intricate architectures in clinical images. Ultimately, the decoder ensures accurate segmentation results by gradually restoring the variable map's spatial resolution through upsampling and skip connections. The encoder, composite mechanism and depth-wise attention component, and decoder all of which are essential to improving segmentation accuracy are explained in more detail in the sections that follow.

3.3. Encoder

The encoder is a hybrid convolutional transformer architecture is seen in Fig 1. Convolutional and down-sampling layers are employed by the convolutional component to extract variables, gradually decreasing the changeable map's length. The variable enters the depth-wise attention after being processed by the composite mechanism. By combining depth-wise convolution with a composite mechanism which combines kernel convolution, PeerNet, and spatial convolution to collect multi-scale contextual information, composite mechanism enhances variable depiction. By integrating channel and spatial attention, the depth-wise attention enhances global variable depiction. By balancing local and global variable processing, this hybrid method maximizes encoder effectiveness. Two main drawbacks of hybrid patterns limiting local variable separation are addressed by the composite mechanism, as shown in Fig 1. The composite mechanism incorporates the following to solve these obstacles:

Depth-wise convolution: Improves local variable separation by more accurately capturing minute details.

Composite mechanism: Enhances the effectiveness of local variable separation and attention procedures.

Depth-wise convolution: This work use a depth-wise convolution procedure to improve computational efficiency while maintaining spatial

Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

information. In particular, the parameter maps' geographic dimensions are maintained by employing a 3 × 3 depth-wise convolution having a padding of 1 (padding = 1, stride = 1). Each channel can be treated independently since the total amount of input streams is matched by the total amount of groups in the inversion. Depth-wise convolution improves the pattern's overall computational efficiency by drastically lowering the computationally demanding nature and amount of parameters while maintaining the accuracy of spatial data as compared to ordinary convolution.

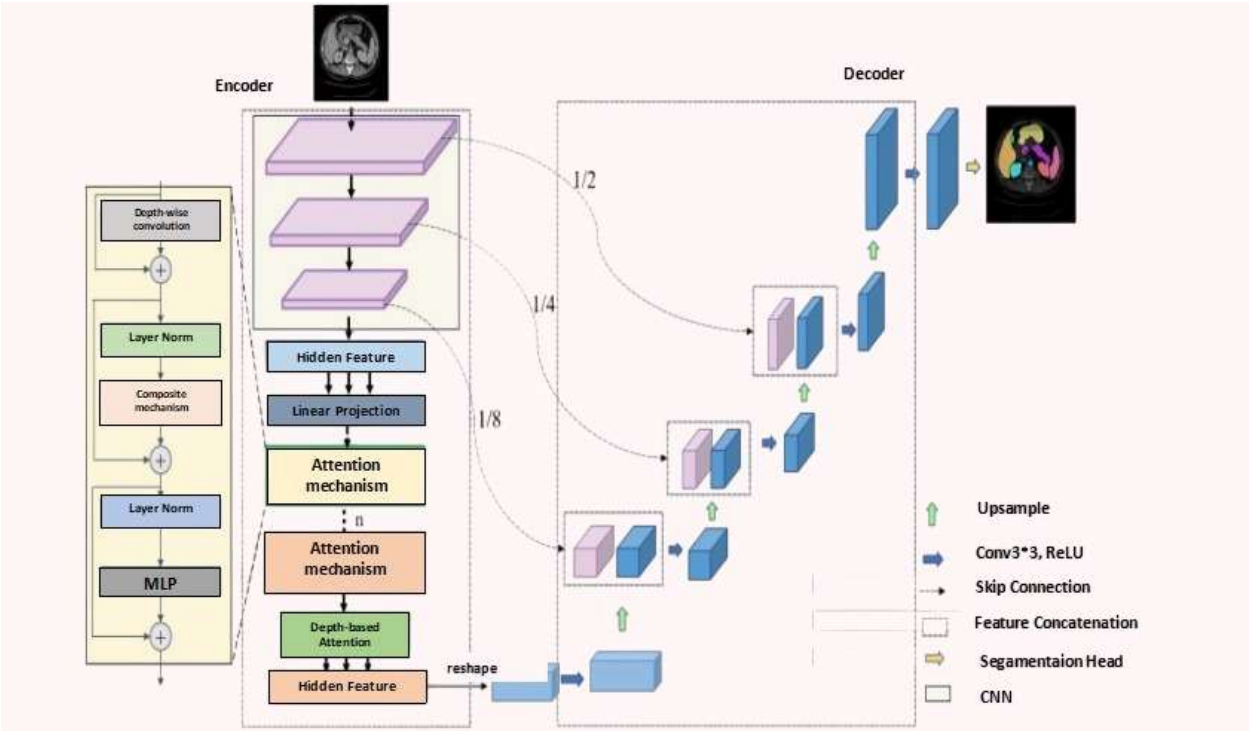

Fig 1: Composite Heart Segmentation Network (CHSegNet)

Composite mechanism (CM): It is an enhanced attention system that effectively combines channel and spatial data. The provided input variables are dynamically improved in multiple dimensions by the CM component by combining geometric changes, combined spatial-channel reconstruction and connections involving local and global parameters. When compared to traditional attention methods, CM exhibits greater adaptability and flexibility. It also patterns elements in complicated circumstances with ease while drastically lowering computing complexity, which enhances overall effectiveness. An overview of kernel convolution, PeerNet, and spatial convolution is given below. An offset matrix is first obtained by applying a 2D convolution to an input variable map $Y \in R^{N\times C\times H\times W}$, where H *and* W are the corresponding height and breadth of the parameter map, and N stands for the number of batches. An initial testing reference set $P_n \in R^{(1,2K,1,1)}$, where every single point is described by two reference values (x, y) *and* K is the total number of points for sampling (the kernel size) is created prior to modifying the sampling co-ordinates. The modified sample coordinates $(N, 2K, H, W)$ are derived by incorporating the acquired offsets into the initial sampling co-ordinates P_n. Bilinear interpolation is employed in a resampling procedure to calculate the variable values at the new locations because these modified sampling coordinates might not line up with integer grid points. To ensure compatibility with further network processing, 1×1 convolution is employed to translate the average level channel dimensionality (N, C, H, W) to the reconfigured constant map (N, K, H, W). Channel-wise information is extracted using a 1×1 convolution from the output variable map, whereas a geographic data is captured via $3 * 3$ convolution. Speed = 1 and padding = 0 are used in standard 1×1 convolutions, whereas speed = 1 and cushioning = 1 are typically used in $3 * 3$ convolutions.

A global average pool (GAP) is then employed to extract global variables, and 1×1 convolution is employed to reduce dimensionality. The worldwide

attentiveness computation's weighting of attention is then generated employing a sigmoid activation function. In order to improve important areas, the produced weights of attention and the input variables map are then multiplied element-wise to perform attention-weighted refining. Lastly, the results of attention-weighted, $1 \times 1, and\ 3 \times 3$ transformations variables are combined employing element-wise addition to create refined variables. By enhancing target region information, this fusion technique improves the variable depiction and its capacity for discrimination. Next, global variables are extracted employing a GAP, and dimensionality is reduced employing a 1×1 convolution. The attention components for the worldwide focus computation are then created using an activation function with a sigmoid. The resulting attention weights and the input variable map are then multiplied element-wise to perform attention-weighted refinement, which enhances key areas. Finally, improved variables are created by combining the results of attention-weighted parameters using element-wise additionally, $1 * 1$ a convolution, and 3 * 3 coupling. This method of fusion improves the parameter depiction and its ability to discriminate by improving target region information. The channel importance is denoted by W_γ in this process, is employed to distinguish between material and superfluous. The variable map is split into two parts according to the determined channel importance weights:

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^{C} \gamma_j}, \gamma = f_{GN}\left(Y_{in}\right) \quad (1)$$

$$Y_1 = W_1 \odot \mathrm{Y_{in}};\ Y_2 = W_2 \odot \mathrm{Y_{in}} \quad (2)$$

The principal information channels (high-weight variables) are represented by Y_1, while the paths of duplicated information (low-weight parameters) are represented by Y_2. To guarantee that the overall information doesn't change, the weight distribution satisfies $W_1 + W_2 = 1$. Weight masks $W_1\ and\ W_2$ are obtained from the significance of the canal vector W_γ. While repressing low-weight methods, high-weight circuits are able to maintain more traits thanks to element-wise multiplication (Ψ). To improve the information depiction capabilities, the spatial variables are finally rebuilt by concatenating data from several sources:

$$Y_w = \left(Y_1 + Y_2\right) \cup \left(Y_2 + Y_1\right) \quad (3)$$

In this case, $\cup$ stands for channel concatenation which allows the pattern to simultaneously hold data from many viewpoints. The CRU receives the spatially optimized variable map Y_w. Channel optimization is employed because redundant information is still present in the spatially optimized variable Y_w:

$$Y_{up},\ Y_{low} = f_{split}\left(Y_w\right) \quad (4)$$

Where, Y_{low} is the supplemental channel information (low-frequency variables) and Y_{up} is the principal channel information (high-frequency variables). The channels are divided employing a split ratio employing the function $fsplit(\cdot)$. After that, a transformation is carried out to extract variables employing point and group convolution:

$$Y_1 = f_{group}\left(Y_{up}\right) + f_{point}\left(Y_{up}\right) \quad (5)$$

$$Y_2 = point\left(Y_{low}\right) \cup Y_{low} \quad (6)$$

In this case, $f_{point}(\cdot)$ uses $1 * 1$ convolution to decrease the size of the transmission channel, whereas $f_{group}(\cdot)$ uses group-wise compression for variable extraction and increase computing efficiency. The optimal channel variables are represented by $Y_1\ and\ Y_2$. To fuse the two kinds of channel information, variable fusion is finally carried out by calculating softmax weights:

$$\beta_1 = \frac{e^{s_1}}{e^{s_1} + e^{s_2}};\ \beta_2 = \frac{e^{s_2}}{e^{s_1} + e^{s_2}} \quad (7)$$

Whereas, $\beta_1\ and\ \beta_2$ calculate from softmax, $s_1\ and\ s_2$ calculate using the global averaged pool, representing the channel variable importance scores.

$$y_{final} = \beta_1 Y_1 + \beta_2 Y_2 \quad (8)$$

In conclusion, the composite mechanism component improves the depictional capacity of spatial and channel data by efficiently learning input variables that provide local as well as international context information. The pattern's receptive field and information consumption efficiency are greatly enhanced by this multi-scale fusion technique. The component dynamically patterns the relative value of information by using the input content to adaptively create weight patterns across geographic and channel parameters. The composite mechanism component ensures thorough information fusion and significantly improves downstream job effectiveness by fully employing the possibility of input parameters in several dimensions by combining multichannel and spatial reconstructing.

3.4. Attention pattern

The attention component is intended to successfully use global context patterning to handle the difficulties of clinical image segmentation is shown in Fig 1. The component starts with a convolutional layer that maintains high-resolution details while enabling the separation of lightweight local variables. Prior to the final sigmoid activation, the channel attention branch (A^{ch}) applies layer normalization. After variable metamorphosis ($1 \times 1\ conv$) and potential amplification or rearranging, its objective is to equalize the variable statistics across the channel dimension. This improves channel-wise variable refinement by stabilizing training, preventing exploding or vanishing activations, and guaranteeing that the attention weights produced by the sigmoid function are in a well-behaved range. The attention component uses a dual-branch technique that concentrates on improving both spatial and channel information to further improve variable depiction. By dynamically determining the significance of various spatial and channel characteristics, this technique adaptively redistributes variable importance. The attention component balances local detail preservation with global context integration by integrating these improved variables to produce a thorough and reliable depiction of input variables. In the passageway dimension, the output parameters are expressed as $A^{ch}(X) \in R^{C\times1\times1.}$

$$A^{ch}(X) = F_{sigmoid}\left[W_z(\sigma_1)\left(W_v(X)\right)\right] * F_{softmax}\left(\sigma_2\left(W_q(X)\right)\right) \quad (9)$$

$$A^{sp}(X) = F_{sigmoid}\left[\sigma_3(F_{softmax}(\sigma_1)\left(W_{group}(W_q(X\right)\right) * \left(\sigma_2\left(W_v(X)\right)\right) \quad (10)$$

In this case, $\sigma_1, \sigma_2, and\ \sigma_3$ are reshape operations, and $W_q\ and\ W_v$ are both 1×1 convolution layers. The Global Pooling operation is represented by $F_{group}(\cdot)$. Matrix multiplication is shown by the symbol " $\times$." As a result, $Z^{sp} = A^{sp}(X) \odot^{sp} X \in R^{C\times H\times W}$ represents the output variables in the spatial dimension, where $\odot^{sp}$ represents the multiplication operation. This work add dropout layers after the channel partition Z^{ch} and the spatial branch Z^{sp}, respectively, to lessen the chance of over-fitting. To improve the pattern's ability to adapt to complicated clinical images, 10% of transistors are randomly eliminated during training when the percentage of neurons that quit is set at 0.1. This work conduct weighted fusion by introducing learnable weights $\lambda_1\ and\ \lambda_2$ to continuously alter the spatial and channel information contributions. The two aforementioned branches' output variables are combined in parallel:

$$Channel\ and\ spatial\ features\ (X) = \lambda_1 Z^{ch} + \lambda_2 Z^{sp} = \lambda_1 A^{ch}(X) \odot^{ch} X + \lambda_2\ A^{sp}(X) \odot^{sp} X \quad (11)$$

In this case, the network learns the trainable parameters $\lambda_1\ and\ \lambda_2$ with initial values of 0.5 and back-propagation optimization. To guarantee the stability of variable fusion, the weights meet the requirement $\lambda_1 + \lambda_2 = 1$ (executed employing softmax normalization). By incorporating multi-scale data and successfully boosting important traits while suppressing unnecessary ones, the depth-wise attention component's creative design increases segmentation accuracy. This method is a potent part of the suggested segmentation method since it guarantees accurate and trustworthy delineation of anatomical variables in clinical pictures. The decoder's architecture with skip connections are essential for recovering image resolution from the encoder's variable maps. To gradually recover spatial resolution, the decoder uses a sequence of up-sampling operations, lowering the total amount of variable channels while simultaneously doubling it at each step. Skip connections make it easier to fuse the up-sampled variable maps with matching encoder variable maps at each up-sampling stage. These fused variables are refined by a further convolutional layer, which improves the separation of segmentation-relevant data. A segmentation head processes the final refined variable map. By efficiently utilizing high-dimensional encoder characteristics, this architecture improves segmentation effectiveness by allowing the decoding device to retrieve intricate limits and information from the image. Algorithm 1 explains the functionality of the proposed model.

Algorithm 1:

Input: RGB clinical image $I \in R^{H\times W\times3}$

Output: Pixel-wise segmented mask $S \in R^{H \times W}$

Step 1: Input Preprocessing

1. Resize the input RGB image I to a fixed resolution.
2. Normalize pixel intensities.
3. Convert image to tensor format for network processing.

Step 2: Encoder Feature Extraction

1. Pass I through convolutional layers with down-sampling to extract low-level local features.
2. Gradually reduce spatial dimensions while increasing channel depth.
3. Preserve spatial hierarchies using convolutional kernels.

Step 3: Depth-wise Convolution

1. Apply a 3×3 depth-wise convolution with: Padding = 1, Stride = 1, Groups = number of input channels
2. Maintain spatial resolution while reducing computational cost.
3. Generate depth-wise feature map Y_{dw}.

Step 4: Composite Mechanism (CM) for Multi-scale Feature Learning

1. Offset Generation: Apply 2D convolution on feature map $Y \in R^{N \times C \times H \times W}$ to compute offset matrix.

2. Sampling Coordinate Adjustment:

- ✓ Initialize reference sampling grid $P_n \in R^{1 \times 2K \times 1 \times 1}$.
- ✓ Add offsets to obtain modified sampling coordinates.

3. Bilinear Interpolation: Resample feature values at new coordinates using bilinear interpolation.

4. Channel Projection: Apply 1×1 convolution to map features to $R^{N \times K \times H \times W}$.

5. Spatial and Channel Feature Extraction

- ✓ Channel features: 1×1 convolution.
- ✓ Spatial features: 3×3 convolution.

6. Global Attention Generation

- ✓ Apply Global Average Pooling (GAP).
- ✓ Reduce dimensionality using 1×1 convolution.
- ✓ Generate attention weights using sigmoid activation.

7. Attention-weighted Refinement

- ✓ Multiply attention weights element-wise with input features.

8. Feature Fusion

- ✓ Fuse attention-weighted, spatial, and channel features via element-wise addition.

Step 5: Channel Re-weighting and Reconstruction

1. Compute channel importance: $W_\gamma = \frac{\gamma_i}{\sum_{j=1}^{C}\gamma_j}, \gamma = f_{GN}\,(Y_{in})$
2. Split features: $Y_1 = W_1 \odot \mathrm{Y_{in}};\ Y_2 = W_2 \odot \mathrm{Y_{in}}$
3. Ensure: $W_1 + W_2 = 1$
4. Spatial reconstruction: $Y_w = (Y_1 + Y_2) \cup (Y_2 + Y_1)$

Step 6: Channel Refinement Unit (CRU)

1. Split channels into high- and low-frequency components: $Y_{up},\ Y_{low} = f_{split}\,(Y_w)$;
2. Feature transformation:
3. High-frequency: $Y_1 = f_{group}\,(Y_{up}) + f_{point}\,(Y_{up})$
4. Low-frequency: $Y_2 = point\,(Y_{low}) \cup Y_{low}$
5. Compute fusion weights using softmax: $\beta_1 = \frac{e^{s_1}}{e^{s_1}+e^{s_2}};\ \beta_2 = \frac{e^{s_2}}{e^{s_1}+e^{s_2}}$
6. Fuse refined features: $y_{final} = \beta_1 Y_1 + \beta_2 Y_2$

Step 7: Depth-wise Attention Module

- ✓ Apply initial convolution to preserve high-resolution features.
- ✓ Channel Attention Branch: Apply 1×1 convolution, Normalize using Layer Normalization, Generate attention map: $A^{ch}(X)$
- ✓ Spatial Attention Branch: Apply group convolution and softmax, Generate spatial attention map $A^{sp}(X)$, apply attention: $Z^{ch} = A^{ch}(X) \odot \mathrm{X};\ Z^{sp} = A^{sp}(X) \odot \mathrm{X}$
- ✓ Apply dropout (rate = 0.1) to both branches.
- ✓ Weighted fusion: $X_{att} = \lambda_1 Z^{ch} + \lambda_2 Z^{sp}$ Initialize $\lambda_1 = \lambda_2 = 0.5$, Enforce $\lambda_1 + \lambda_2 = 1$ using softmax.

Step 8: Decoder and Segmentation Output

1. Upsample feature maps step-by-step to restore spatial resolution.
2. Fuse upsampled features with corresponding encoder features using skip connections.
3. Apply convolutional refinement at each decoding stage.
4. Pass final feature map through segmentation head.
5. Generate pixel-wise segmentation mask S.

End

4. PERFORMANCE METRICS

The segmentation effectiveness of cardiac MRI was evaluated in this work employing three assessment metrics: The evaluation metrics comprise the Dice, IoU, and Acc. These metrics are particularly relevant in the context of multi-class segmentation problems, the corresponding metrics can be calculated as follows where A *and* B denote the sets of pixels for class c in the ground truth and predicted mask, respectively, and |S| represents the cardinality of set S.

Dice coefficient (dice): a similarity metric between two collections that is typically applied to semantic partitioning issues. Eq. (12) illustrates how the set is generated from the cardiac image segmented by the network architecture, assessing the similarity between the two segmentation outputs and the reference image.

$$Dice\,(A,B) = 2 * \frac{|A \cap B|}{|A| + |B|} \qquad (12)$$

The full spatial resemblance representing the overlap between two databases from binary segmentation, indicating the extent of spatial agreement, is represented by dice values between 0 and 1. An precise segmentation of the heart image with a Dice value close to 1 signifies strong overlap with the

cardiac MRI image. On the other hand, a dice value near zero indicates inadequate segmentation and little overlap with the cardiac MRI image.

The overlap between the ground truth and the anticipated segmentation is measured by intersection over union, or IoU.

$$IOU\ (A,B) = 2 * \frac{|A \cap B|}{|A \cap B|} \quad (13)$$

It shows the extent to which the projected region matches it measures alignment with the actual region, penalizing both under-segmentation (missing pixels in B) and over-segmentation (too many pixels in A).

Accuracy: A ratio of successfully categorized pixels to the image's total pixel count;

$$Acc\ (A,B) = 2 * \frac{|A \cap B| + |A^c \cap B^c|}{|AUB|} \quad (14)$$

Where, $|A \cap B|$ refers to the total number of pixels in the union of the ground truth and predicted sets, $|A^c \cap B^c|$is the number of pixels correctly predicted as belonging to the target class, and $|AUB|$ is the number of pixels correctly predicted as not belonging to the target class.

4.1. Ablation research

Our suggested network's initial learning rate (Lr) is set at 0.0001, and parameter updates are made employing the Adam optimization technique. The regularization method is batch normalization and the dropout rate is set at 0.5. To offset the impact of the smaller number of batches is used with the model trained for up to 100 epochs and a number of batches of 4. Each pixel in the ground truth maps is normalized into $0-3$ for labeling purposes with 0 denoting background, 1 denoting RV, 2 denoting Myo, and 3 denoting LV. The segmentation effectiveness will be determined by averaging all assessment indicators since the objective of the CHSegNet development is to generate an accurate segmentation of all three heart chambers. In the framework of MRI cardiac segmentation, attaining high accuracy in identifying heart layouts is essential for supporting clinical evaluation and treatment planning. Examining the solo effectiveness of a unique network design, like the suggested CHSegNet, is only one aspect of its development and validation. A methodical comparison with alternative segmentation networks centered around CNN is necessary for a thorough assessment. While verifying that the noted benefits are the result of differences in this contrast strategy helps highlight the benefits and potential disadvantages of the proposed paradigm rather than differences in training sets.

Table 2. Performance comparison

Model	Performance metrics		
	Dice	**IoU**	**Accuracy**
DenseNet	54.5	59.1	89.3
CNet	58.9	58.2	89.7
FCN	63.2	57.6	68.9
SegNet	63.9	54.7	70.9
RNN	64.1	54.7	75.9
CHSegNet	**85.2**	**78.1**	**99.5**

Table 3. Performance comparison with segmentation outcomes

Model	Performance metrics			
	Precision	**Recall**	**Accuracy**	**F1-score**
DenseNet	94.8	95.9	95.4	91.4
CNet	90.7	87.5	91.2	89.6
FCN	89.4	85.4	90.3	94.6
SegNet	94.1	94.1	94.5	92.9
RNN	92.8	92.8	92.9	96.1
CHSegNet	**97.9**	**97.7**	**99.5**	**98.5**

Table 4. Comparison based on ablation study

Mechanism	Training DSC	Validation DSC	Test DSC	Observation
Without attention mechanism	96.8	95.5	95.7	Over-fitting
Without channel information	98.1	96.5	96.4	Under-fitting
CHSegNet	**100**	**97.9**	**97.8**	**Best performance**

Table 5. P-value computation

Model	p-value	significance
DenseNet	0.23	No
CNet	0.019	No
FCN	0.35	No
SegNet	0.46	No
RNN	0.50	No
CHSegNet	0.0020	Yes

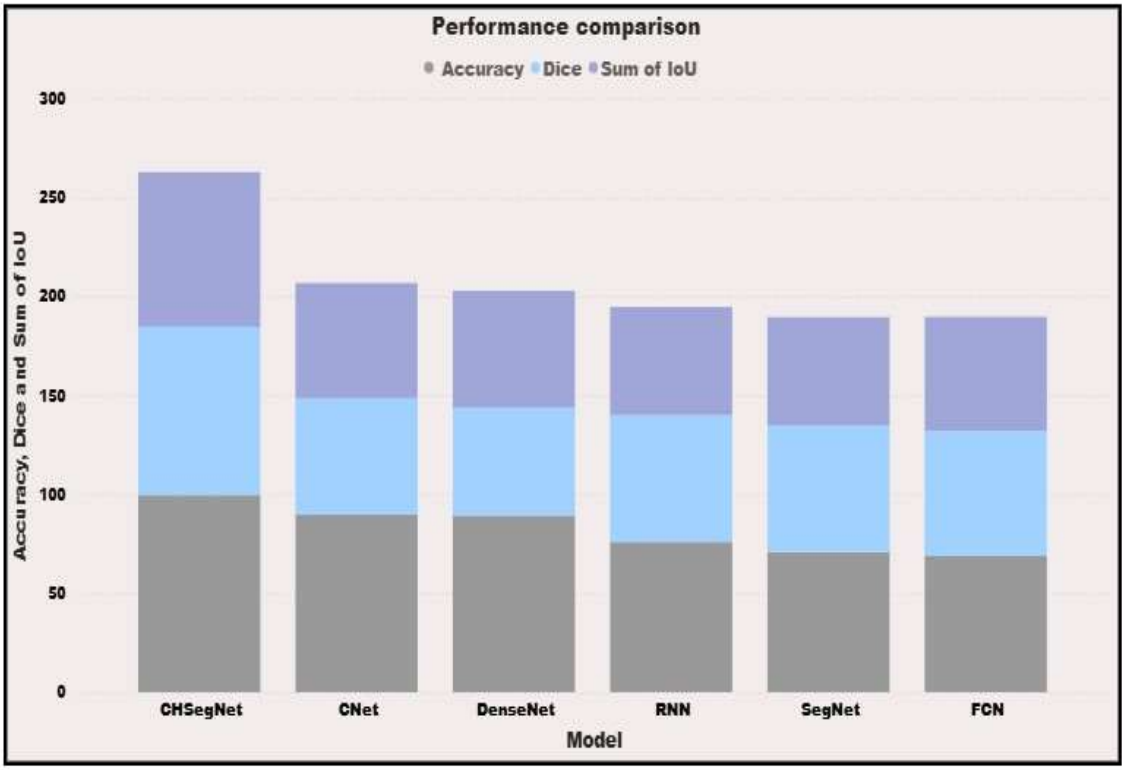

Fig 2. Performance comparison w.r.t. accuracy, dice, IoU

Journal of Theoretical and Applied Information Technology
30th April 2026. Vol.104. No.8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

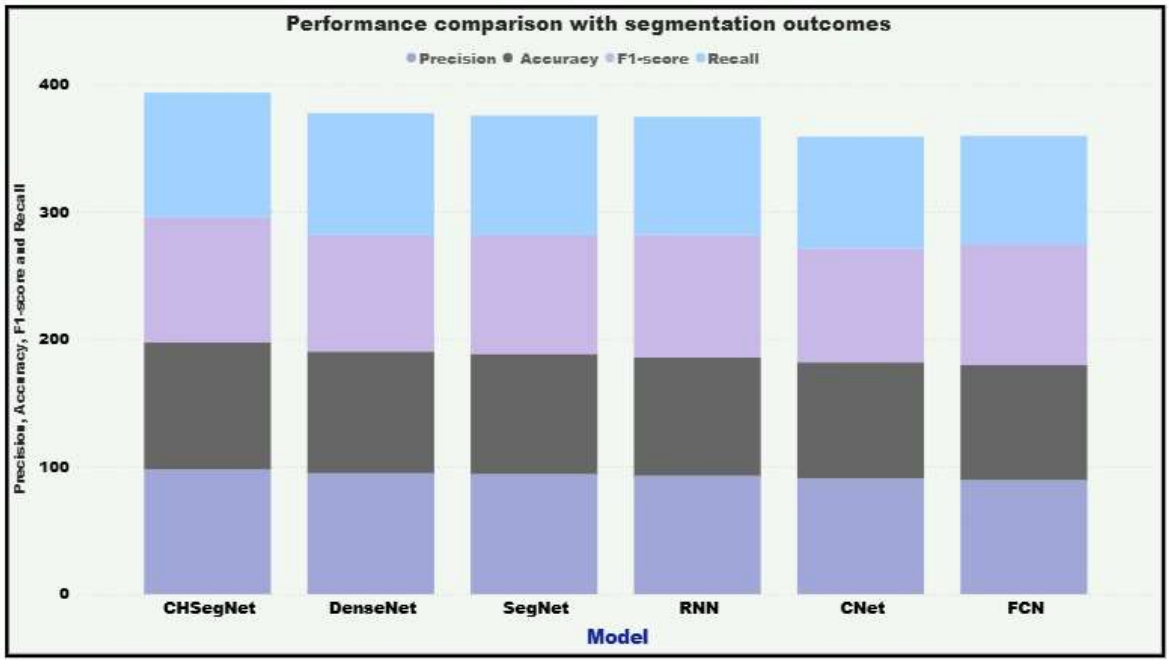

Fig 3. Segmented outcomes w.r.t. precision, accuracy, F1-score and recall

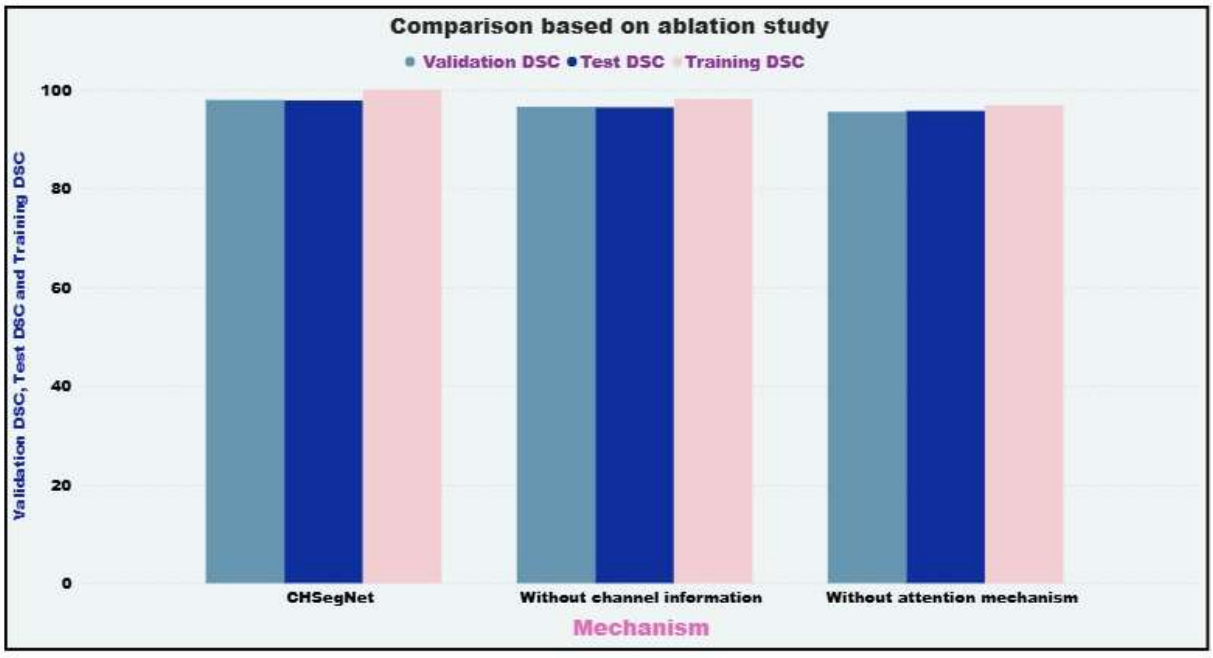

Fig 4. Comparison based on ablation study

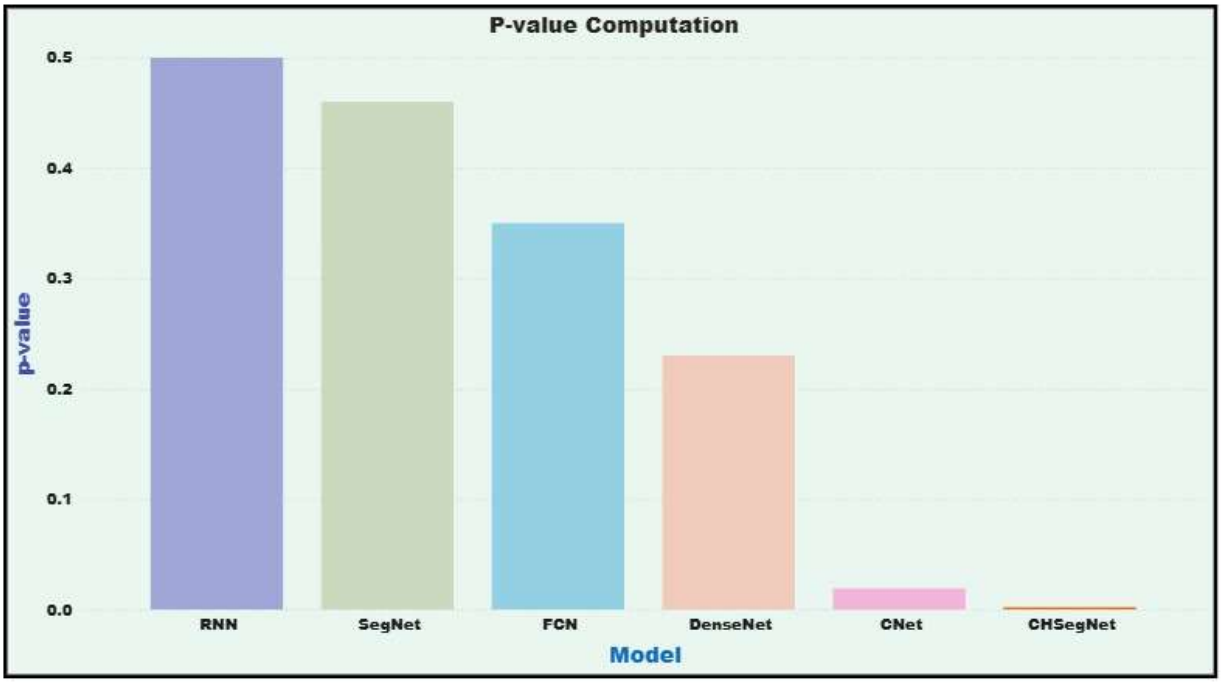

Fig 5. P-value computation

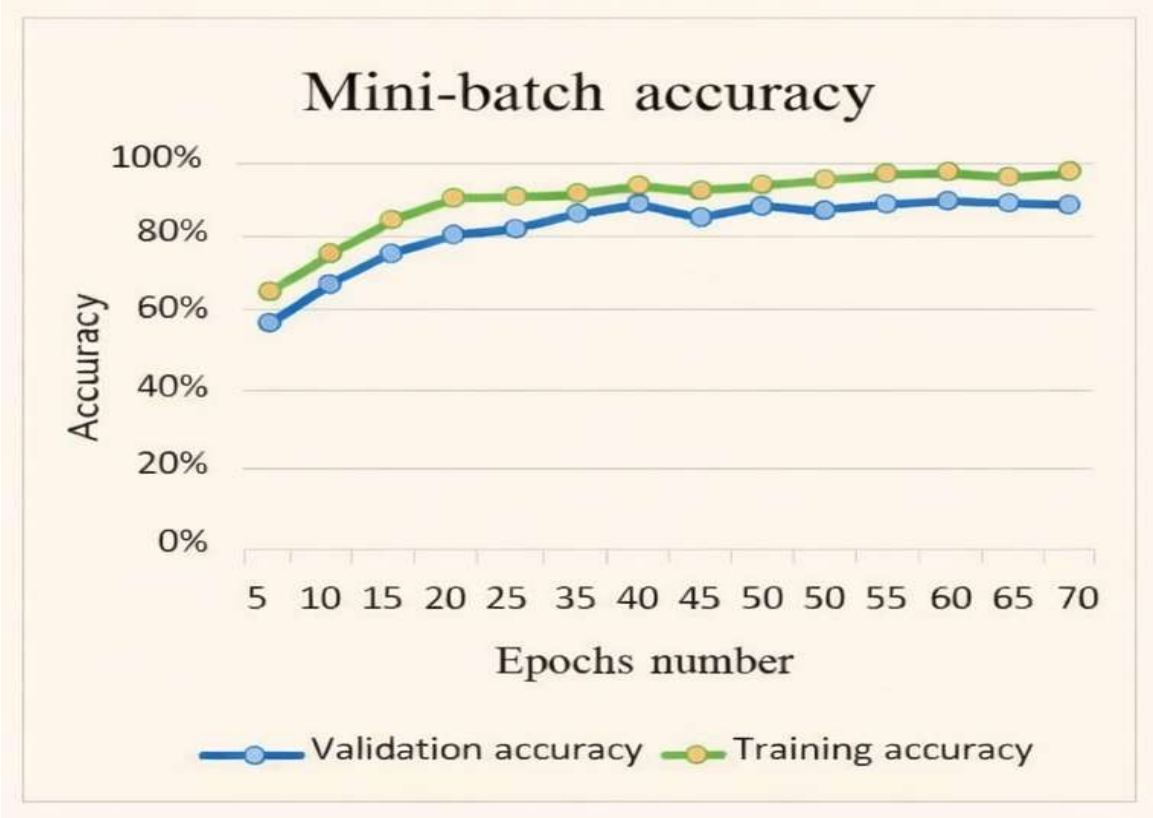

Fig 6. Mini-batch accuracy

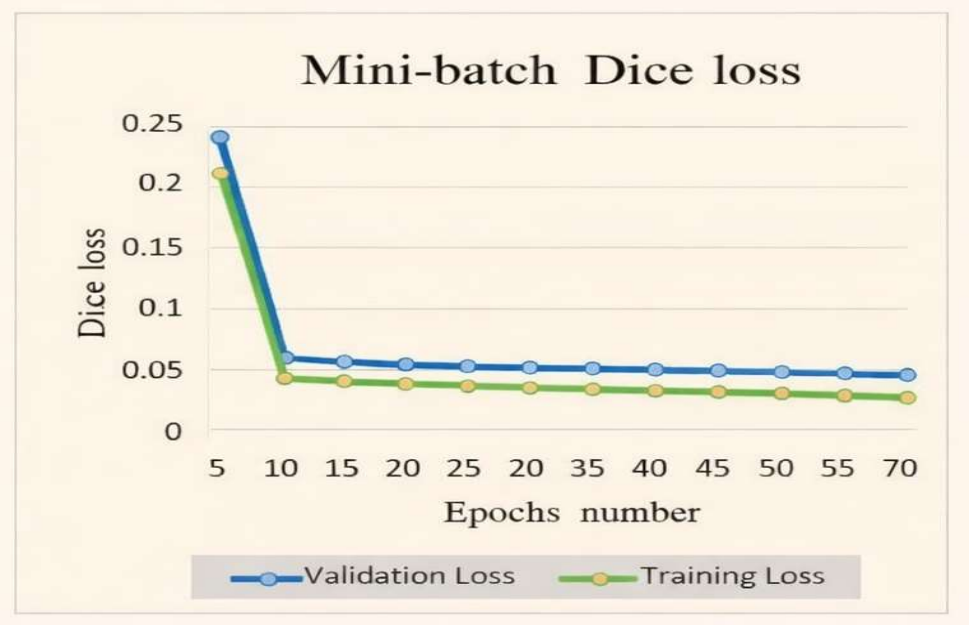

Fig 7. Mini-batch dice loss

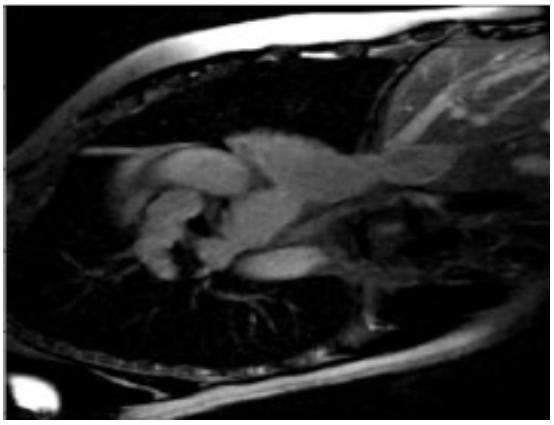

Fig 8a. Input sample

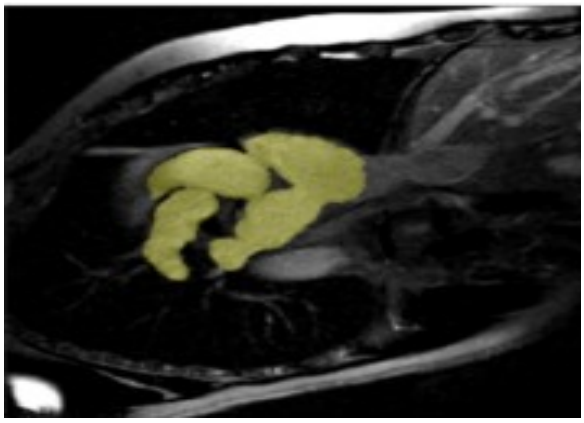

Fig 8b. Segmented outcome

The suggested CHSegNet can be directly and impartially compared to other networks, such as FC-DenseNet, CNet, U-Net, FCN-8, and SegNet, thanks to this uniform configuration. By comparing SDNet to various architectures, we can better understand how its multi-scale dense connections impact its capacity to extract contextual information and intricate spatial details from CMRI pictures. Additionally, this study shows which architectural components like missing links, dense interactions, and multi-scale pooling work best for addressing cardiac MRI delineation issues. Table 2 demonstrates that because the suggested CHSegNet successfully integrates strong connections and multi-scale pooling, it performs better than existing networks, which enables the capture of both contextual and detailed information.

Despite not having explicit multi-scale pooling, FCNet performs comparatively well because of its dense connections, which improve variable repurposing and the preservation of spatial details. Because they have already trained backbones or skip connections, networks like DNet and U-Net function well; yet, they absence a dense and multi-scale structures necessary to improve effective handling of fine details in complex cardiac MRI segmentation tasks as in Table 3. Conversely, a typical encoder-decoder network is employed in architectures like SegNet, but the decoder portion is not supported by

skip connections. When compared to U-Net or FC-DenseNet, this result demonstrates lower detail retention. Lower segmentation accuracy results from the network's inability to reconstruct exact boundary details in the absence of skip connections, especially for cardiac MRI. This investigation emphasizes the significance of multi-scale and strong connections in deep network design for cardiac MRI segmentation, where precise detection of local as well as worldwide anatomical details is essential. Complex cardiac organs are difficult for DNet, U-Net, and FCN8 to accurately category, which causes significant departures from the actual truth; SegNet performs the worst because it has trouble correctly segmenting myocardial structures, probably because of lost finely detailed information as in Table 4, which are crucial for accurately outlining each cardiac chamber's boundaries and FC-DenseNet also demonstrates strong performance as in Fig 2 to Fig 8, however, it encounters difficulties in cases characterized by substantial variations in size and scale based on ablation study in Table 5.

5. CONCLUSION

To create the best cardiac segmentation maps, this work determines the optimal segmentation configuration, which depends heavily on the encoder's variable size and layer depth. The incorporation of more parallel routes in the Composite Heart Segmentation Network (CHSegNet) facilitates a more comprehensive multi-scale representation, thereby enhancing the capture of fine-grained information, generally help earlier encoder layers (with greater resolutions). On the other hand, the pattern has fewer paths to effectively balance multi-scale context and computational efficiency when the Composite Heart Segmentation Network (CHSegNet) is positioned deeper in the encoder, where variables are more compressed. Therefore, obtaining with cardiac magnetic resonance imaging data, this flexible usage of CHSegNet pathways across layers greatly improves segmentation efficacy. High-quality segmented images are produced by the created multi-scale DL pattern, improving organ visualization and assisting clinicians in better informing patients of the diagnosis. The dice score of CHSegNet is 85.2, IoU is 78.1 and accuracy is 99.5% respectively. Additionally, the proposed CHSegNet technique lessens subjective biases, guaranteeing consistent diagnosis outcomes for various scenarios and clinical professionals. Additionally, the amount of time needed to manually draw heart structures will be greatly reduced by automating the segmentation process, allowing for facilitating quicker assessment and scheduling for therapies. Lastly, the suggested clinical workflows can incorporate automated CHSegNet segmentation, freeing up physicians to concentrate on activities related to interpretation instead of the laborious manual segmentation of organ sections.

REFERENCES

[1] Teo, K.K.; Rafiq, T. Cardiovascular Risk Factors and Prevention: A Perspective from Developing Countries. Can. J. Cardiol. 2021, 37, 733–743.

[2] Bernard, O.; Lalande, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.A.; Cetin, I.; Lekadir, K.; Camara, O.; Gonzalez Ballester, M.A.; et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Trans. Med. Imaging 2018, 37, 2514–2525.

[3] Awan, M.J.; Rahim; Salim, N.; Rehman, A.; Garcia-Zapirain, B. Automated Knee MR Images Segmentation of Anterior Cruciate Ligament Tears. Sensors 2022, 22, 1552.

[4] Jalali, Y.; Fateh, M.; Rezvani, M.; Abolghasemi, V.; Anisi, M.H. ResBCDU-Net: A Deep Learning Framework for Lung CT Image Segmentation. Sensors 2021, 21, 268.

[5] Rayed, M.E.; Islam, S.M.S.; Niha, S.I.; Jim, J.R.; Kabir, M.M.; Mridha, M.F. Deep Learning for Medical Image Segmentation: State-of-the-Art Advancements and Challenges. Inf. Med. Unlocked 2024, 47, 101504.

[6] Zedan, M.J.M.; Zulkifley, S.R. Automated Glaucoma Screening and Diagnosis Based on Retinal Fundus Images Using Deep Learning Approaches: A Comprehensive Review. Diagnostics 2023, 13, 2180.

[7] Wang,Z.;Peng, Y.; Li, D.; Guo, Y.; Zhang, B. MMNet: A Multi-Scale Deep Learning Network for the Left Ventricular Segmentation of Cardiac MRI Images. Appl. Intell. 2022, 52, 5225–5240.

[8] Abdani, S.R.; Zulkifley, M.A.; Zulkifley, N.H. Group and Shuffle Convolutional Neural Networks with Pyramid Pooling Module for Automated Pterygium Segmentation. Diagnostics 2021, 11, 1104.

[9] Zulkifley, M.A.; Abdani, S.R.; Zulkifley, N.H.; Shahrimin, M.I. Residual-Shuffle Network with Spatial Pyramid Pooling Module for COVID-19 Screening. Diagnostics 2021, 11, 1497.

[10] Zhang, J.; Zhang, Y.; Zhang, H.; Zhang, Q.; Su, W.; Guo, S.; Wang, Y. Segmentation of

Biventricle in Cardiac Cine MRI via Nested Capsule Dense Network. PeerJ Comput. Sci. 2022, 8, e1146.
[11] Shen, D.; Pathrose, A.; Sarnari, R.; Blake, A.; Berhane, H.; Baraboo, J.J.; Carr, J.C.; Markl, M.; Kim, D. Automated Segmentation of Biventricular Contours in Tissue Phase Mapping Using Deep Learning. NMR Biomed. 2021, 34, e4606.
[12] Yan, Z.; Su, Y.; Sun, H.; Yu, H.; Ma, W.; Chi, H.; Cao, H.; Chang, Q. SegNet-Based Left Ventricular MRI Segmentation for the Diagnosis of Cardiac Hypertrophy and Myocardial Infarction. Comput. Methods Programs Biomed. 2022, 227, 107197.
[13] Stofa, M.M.; Zulkifley, A.; Atiqi, M.A.; Zainuri, M. Micro-Expression-Based Emotion Recognition Using Waterfall Atrous Spatial Pyramid Pooling Networks. Sensors 2022, 22, 4634.
[14] Lell, M.M.; Kachelrieß, M. Recent and upcoming technological developments in computed tomography: High speed, low dose, deep learning, multienergy. Investig. Radiol. 2020, 55, 8–19.
[15] Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. arXiv 2021, arXiv:2102.04306.
[16] Cai, S.; Tian, Y.; Lui, H.; Zeng, H.; Wu, Y.; Chen, G. Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quant. Imaging Med. Surg. 2020, 10, 1275.
[17] Ma, S.; Tang, J.; Guo, F. Multi-task deep supervision on attention R2U-net for brain tumor segmentation. Front. Oncol. 2021, 11, 704850.
[18] Hussain, T.; Shouno, H. MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi attention gated residual U-Net. IEEE Access 2024, 12, 40290–40310.
[19] Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote. Sens. 2020, 162, 94–114.
[20] Hussain, I.; Nataliani, Y.; Ali, M.; Hussain, A.; Mujlid, H.M.; Almaliki, F.A.; Rahimi, N.M. Weighted Multiview K-Means Clustering with L2 Regularization. Symmetry 2024, 16, 1646.
[21] Hussain, T.; Shouno, H.; Mohammed, M.A.; Marhoon, H.A.; Alam, T. DCSSGA-UNet: Biomedical image segmentation with DenseNet channel spatial and semantic guidance attention. Knowl.-Based Syst. 2025, 314, 113233.
[22] Jamali, A.; Roy, S.K.; Li, J.; Ghamisi, P. TransU-Net++: Rethinking attention gated TransU-Net for deforestation mapping. Int. J. Appl. Earth Obs. Geoinf. 2023, 120, 103332.
[23] Lin, A.; Chen, B.; Xu, J.; Zhang, Z.; Lu, G.; Zhang, D. Ds-transunet: Dual swin transformer u-net for medical image segmentation. IEEE Trans. Instrum. Meas. 2022, 71, 4005615.
[24] Li, Z.; Li, Y.; Li, Q.; Wang, P.; Guo, D.; Lu, L.; Jin, D.; Zhang, Y.; Hong, Q. Lvit: Language meets vision transformer in medical image segmentation. IEEE Trans. Med. Imaging 2023, 43, 96–107.
[25] Hussain, T.; Shouno, H.; Hussain, A.; Hussain, D.; Ismail, M.; Mir, T.H.; Hsu, F.R.; Alam, T.; Akhy, S.A. EFFResNet-ViT: A fusion-based convolutional and vision transformer model for explainable medical image classification. IEEE Access 2025, 13, 54040–54068.
[26] Zhao, X.; Jia, H.; Pang, Y.; Lv, L.; Tian, F.; Zhang, L.; Sun, W.; Lu, H. M2SNet: Multi-scale in multi-scale subtraction network for medical image segmentation. arXiv 2023, arXiv:2303.10894.
[27] Baptista, E.A.; Queiroz, B.L. Spatial analysis of cardiovascular mortality and associated factors around the world. BMC Public Health 2022, 22, 1556.
[28] Budai, A.; Suhai, F.I.; Csorba, K.; Dohy, Z.; Szabo, L.; Merkely, B.; Vago, H. Automated classification of left ventricular hypertrophy on cardiac mri. Appl. Sci. 2022, 12, 4151.
[29] Parlati, A.L.M.; Nardi, E.; Marzano, F. Advancing Cardiovascular Diagnostics: The Expanding Role of CMR in Heart Failure and Cardiomyopathies. J. Clin. Med. 2025, 14, 865.
[30] Li,S.; Feng, Z.; Xiao, C.; Wu, Y.; Ye, W.; Zhang, F. The establishment of hypertrophic cardiomyopathy diagnosis model via artificial neural network and random decision forest method. Mediat. Inflamm. 2022, 1–15.