Al-Rafidain Engineering Journal (AREJ) Vol. 31, No. 2, June 2026, pp. 23-37

Effect of Attention Mechanisms in a Fuzzified Nested U-Net for Medical Image Segmentation

Noor M. Basheer
noor.23enp115@student.uomosul.edu.iq

Ali Al-Saegh
ali.alsaegh@uomosul.edu.iq

Computer Engineering Department, College of Engineering, University of Mosul, Mosul, Iraq.

Received: July 9th, 2025 Revised: August 24th, 2025 Accepted: October 28th, 2025

ABSTRACT

Accurate segmentation of cardiac structures (right ventricle (RV), left ventricle (LV), and myocardium (Myo)) from cardiac MRI images plays an important role in the diagnosis and treatment of cardiovascular disease. However, despite all this, segmentation of these structures at the micro level remains a major challenge due to the complex anatomical diversity and noise inherent in MRI data. This paper examines the performance of various cardiac MRI segmentation techniques, with a primary focus on the overlapping U-Net structure, attention mechanisms, and fuzzy pooling strategies. This study comprehensively evaluates these methods, both independently and in combination, to determine their effectiveness in improving segmentation quality for the RV, LV, and myocardial regions. Additionally, the effect of thresholding strategies on segmentation accuracy is examined. The experimental results on the Automated Cardiac Diagnosis Challenge (ACDC) dataset show that the proposed model (combining nested U-Net, attention mechanisms, and fuzzy pooling) achieved a dice score of 98.20%, an accuracy of 96.83%, and a recall of 96.83%, superior to other basic methods. In comparison, the best-performing core model, ANU-Net, achieved a Dice score of 94.31%, accuracy of 95.19%, and recall of 93.44%. These findings underscore the superior performance of the hybrid model in terms of segmentation and boundary delineation accuracy. These results confirm the potential of hybrid deep learning models in developing cardiac image analysis. Future work will focus on improving these configurations across diverse datasets and also exploring real-time deployment strategies in clinical settings.

Keywords:
Attention Mechanism, Deep Learning for Medical Imaging Segmentation, Fuzzy Pooling, Nested U-Net, Medical Image Analysis.

https://rengj.uomosul.edu.iq
Email: alrafidain_engjournal3@uomosul.edu.iq

==========

1. INTRODUCTION

Medical image segmentation is an essential task in computer-aided diagnosis, treatment planning, and disease monitoring [1]. In cardiovascular medicine, cardiac magnetic resonance imaging (MRI) is the primary criterion for evaluating cardiac anatomy and function due to its high spatial resolution and excellent soft tissue contrast [2]. Accurate segmentation of major structures, LV, RV, and MYO, is important to diagnose conditions such as myocardial infarction, cardiomyopathy, and congenital malformations [3]. However, segmentation remains a challenge due to noise, low tissue contrast, and the complex geometry of cardiac structures [4]. Traditional image processing methods, including thresholding and edge detection, are often inadequate and lack robustness in clinical practice [5]. Deep learning methods, especially convolutional neural networks (CNNs), have proven superior performance in segmenting biomedical images [6]. Among these models, U-Net has become the most widely used model due to its encoding and decoding architecture with skip connections, which allows maintaining global context and fine details [7]. Despite its success, the U-Net standard has limitations when dealing with complex or irregular structures due to limited feature selectivity and insufficient contextual awareness [8].

In order to address this problem, attention mechanisms have been introduced, including channel attention (which focuses on key feature maps), spatial attention (which highlights critical areas) [9], and self-attention (that captures long-term dependencies) [10]. Hybrid forms further combine these elements in order to enhance multi-scale contextual learning [11]. Feature aggregation also plays a key role in segmentation performance.

Traditional maximum pooling reduces dimensions but often ignores valuable spatial information [12]. In contrast, fuzzy pooling integrates fuzzy logic to assign membership values, which results in smoother and more robust feature reduction while maintaining boundaries, which is critical for medical imaging [13]. Thresholding is another important factor, as it directly affects sensitivity, quality, and metrics such as Dice similarity and Hausdorff distance [14]. Traditional maximum pooling reduces dimensions but often ignores valuable spatial information [15]. In contrast, fuzzy pooling integrates fuzzy logic to assign membership values, which results in smoother and more robust feature reduction while maintaining boundaries, which is critical for medical imaging [13]. Thresholding is another important factor, as it directly affects sensitivity, quality, and metrics such as Dice similarity and Hausdorff distance [14].

In this study, a hybrid framework combining nested U-Net, attention mechanisms, and fuzzy pooling is proposed for cardiac MRI segmentation [5]. The nested U-Net network provides dense skip connections to aggregate multi-scale features, reduce semantic gaps, and enhance gradient flow [7]. Attention units guide the selection of features at different stages, while fuzzy pooling replaces maximal pooling to preserve anatomical details [15]. The segmentation outputs are optimized using optimized thresholds, and quantitative features such as ventricular volume and myocardial thickness are extracted to classify the patient into subgroups: normal (NOR), myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal RV [3].

The main contributions of this paper are:

- A new hybrid model integrating nested U-Net, attention, and fuzzy pooling for robust cardiac MRI segmentation.
- Systematic evaluation of attention types, pooling strategies, and threshold effects on segmentation accuracy.
- Validation of the proposed model for both segmentation (RV, LV, MYO) and diagnostic classification of cardiac conditions.

The rest of the paper is organized as follows: Section 2 reviews relevant work. Section 3 describes methods and experimental setup. Section 4 presents the experimental results and comparisons. Section 5 discusses the results, and Section 6 concludes the paper.

2. RELATED WORKS

Recent advances in deep learning have greatly improved the segmentation and classification of medical images. The U-Net architecture [13] is a basic model due to its design, which symmetric encryption and decryption while skipping connections, maintaining spatial resolution while capturing hierarchical features. Initially designed for cell tracking, the U-Net has been widely applied in tumours, organs, and lesion segmentation. However, they face difficulties in dealing with complex spatial relationships and precise anatomical boundaries. The nested U-Net [14] enhances feature integration through dense and nested skip connections, thereby outperforming the standard U-Net by 2–3% in dice scores on datasets such as ACDC and ISIC [15]. However, in contrast, it lacks explicit attention mechanisms. Attention U-Net [16] integrates attention gates into skip connections in order to focus on relevant spatial areas while suppressing background noise, resulting in improvements in dice scores (e.g., +4.1% on pancreatic MRI) and sensitivity.

The U-Net [17] nested residual attention integrates the remaining blocks within the attention units in order to enhance the gradient flow, although it increases computational complexity. SCU-Net++ [18] combines sharp convolutional units with attention and residual trajectories to enhance boundary definition, achieving Dice scores above 0.92 in retinal and brain vascular segmentation; however, it may struggle with low-contrast images. Channel attention methods, such as pressure and excitation (SE) [19] and channel-based attention in U-Net [20], recalibrate feature responses but may neglect spatially important features. Hybrid attention modules, such as CBAM [21], which combine spatial attention and channel attention, improve performance but increase inference time. While spatial attention [22] enhances lesion localization, cross-attention (CCA) [23] efficiently captures long-term dependencies. Dual self-attention networks [24] and dual attention networks [25] effectively capture complex contextual information, but they require significant computational resources. In cardiac MRI, multiple attention models showed substantial improvements. Hybrid models that combine spatial and subjective attention [26] improve LV, RV, and MYO segmentation, albeit at the expense of increased training time. Temporal attention [7] improves CMRI segmentation across cardiac cycles but is sensitive to motion effects. Sectional attention [27] supports multi-organ segmentation, while light attention [20] balances accuracy and efficiency. Modern multi-scale attention strategies [28] fine-tune receptive fields while integrating features, but add complexity.

Assembling features is crucial for maintaining anatomical boundaries. Maximum pooling [29] is computationally efficient but may ignore fine structural details. Fuzzy pooling [30], employed in the proposed model, integrates fuzzy logic in order to maintain uncertainty, preserve soft boundaries, and enhance robustness in noisy medical images. Overall, attention mechanisms, overlapping architecture, and advanced pooling strategies collectively enhance segmentation performance, guiding the development of hybrid models that can analyze accurate, robust, and clinically important medical images.

3. METHODOLOGY

This section presents several key components, including the MRI dataset, the model used, as shown in Fig.1, as well as the threshold and the experimental setup.

3.1. MRI Datasets

The ACDC dataset consists of 150 patients and these are divided into five subgroups where each represents different diseases and these five totals include regular subjects (NOR) which is 30 healthy individuals and myocardial infarction (MINF) with 30 patients with previous myocardial infarction (part of ejaculation less than 40%) and expanding cardiomyopathy (DCM) includes 30 patients with left ventricular diastolic volume > 100 ml/m2 ejection fracture < 40% and hypertrophic cardiomyopathy (HCM) with 30 patients with high left ventricle mass (more than 110g/m2) and heart muscle thickness > 15mm with a natural part of ejaculation and abnormal right ventricle (RV) includes 30 patients with abnormal RV parameters (e.g., RV size > 110 ml/m2 or RV ejection fracture < 40%) [31]. The dataset used in this study is publicly available from the ACDC Challenge website at https://acdc.creatis.insa-lyon.fr/.

The Sunnybrook Cardiac Dataset (SCD) is the dataset used to validate our work. It is a collection of cardiac MRI images designed to evaluate automated segmentation algorithms. The dataset consists of cine-MRI scans obtained from patients representing four clinically distinct groups: healthy individuals, patients with left ventricular hypertrophy (LVH), patients with heart failure without infarction (HF-nonI), and patients with heart failure with infarction (HF-I). Each case is accompanied by endocardial and epicardial lines explained by experts, which provides the basic truth for segmentation tasks. In addition to imaging and contour data, the dataset also includes important clinical measures such as ejection fraction (EF), delayed gadolinium enhancement (LGE) information, and volumetric indices, making it suitable not only for assessing segmentation but also for functional and diagnostic analysis of heart disease. The dataset is available for research purposes through the Sunnybrook Research Institute's LV Segmentation Challenge website at [32]

3.2. U-Net and nested U-Net architecture.

U-Net is a deep learning model for biomedical image segmentation, featuring a symmetric encoder-decoder architecture. The encoder extracts spatial features using convolution, ReLU activation, and maximum pooling [15], while the decoder reconstructs the segmentation mask by upsampling. Skip connections connect the corresponding encryption and decryption layers. This maintains high-resolution spatial information and combines coarse and fine features to improve segmentation.

$$\boldsymbol{fenc(l) = \sigma(W(l) * f(l-1) + b(l))} \quad (1)$$

Where $f(l-1)$ is the input feature map from the previous layer, $W(l)$ is the convolution kernel, $b(l)$is the bias term, σ is the ReLU activation function, and $*$ denotes the convolution operation.

The decoder output at layer l is computed as [1]:

$$\boldsymbol{fdec(l) = \sigma(Wd(l) * Concat(fenc(l)\ Up(fdec(l+1))) + bd(l))} \quad (2)$$

Where Up $fdec(l+1)$is the unsampled feature from the deeper decoder layer, and $Concat$ denotes the concatenation of encoder and decoder features (skip connection).

To assess segmentation quality, the U-Net commonly uses the Dice Similarity Coefficient (DSC), given by [2]:

$$\boldsymbol{DSC = 2|X \cap Y| / (|X| + |Y|)} \quad (3)$$

In the Nested U-Net, dense, multi-level skip connections enhance segmentation by grouping features from different semantic levels, narrowing the gap between encoder and decoder outputs. Each stage of decoding combines the corresponding encoding outputs, deeper decoding outputs, and previous decoding features for improved predictions [17].

$$\boldsymbol{fused(l) = Concat\left(fenc(l), Up\left(fdec(l+1)\right), fnested(l)\right)} \quad (4)$$

where $fnested$ $(l)$the fused feature map at layer l, which integrates multiple sources: $fenc$ (l) (high-resolution details), $Up(fdec(l+1))$, upsampled decoder feature map from layer l+1, providing coarse-scale semantic context. $fnested$ (l), from earlier stages in the same layer, $Concat(.)$, representing previously combined multi-scale features from dense skip connections.

Key benefits include: dense connections for full feature aggregation, improved gradient flow for faster convergence, multi-scale feature reuse, and deep supervision with optional pruning to reduce computational requirements. This design enhances segmentation accuracy while maintaining effective inference. Features are combined by linking the sampled encoder and decoder and overlapping outputs (Fig. 1).

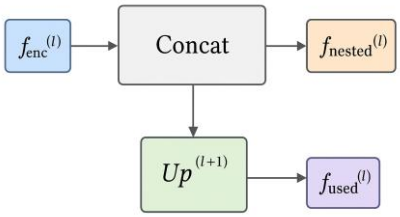

Fig. 1: The Feature fusion in the Nested U-Net.

3.3. Fuzzy Pooling

Fuzzy pooling integrates fuzzy logic into CNN pooling in order to improve segmentation, particularly for fine tissue boundaries [33]. In contrast, maximum aggregation weighs each feature based on its suitability using membership functions, such as Gaussian functions, to preserve spatial information and smooth activation. Ambiguous values are assigned to features (such as low, medium, and high), and ambiguous rules combine them into a weighted average, while also retaining details and reducing noise sensitivity. Fuzzy membership function (Gaussian) [3]:

$$\boldsymbol{\mu_{i(x)} = \frac{1}{1+\left(\frac{|x-c_i|}{\sigma_i}\right)}} \qquad (5)$$

- $\mu i(x)$ is the membership value for the i-th feature,
- ci is the fuzzy center (mean),
- σi is the spread (controls fuzziness),
- x is the input pixel or activation.

Fuzzy Output Aggregation [3]:

$$\boldsymbol{y = \sum i \mu i(x) \cdot xi / \sum i \mu i(x)} \qquad (6)$$

The weighted average is computed based on the membership strengths $\mu i(x)$. Fuzzy pooling enhances segmentation by preserving spatial detail, minimizing information loss, and refining boundary definition in medical images.

3.4. Attention Mechanism

Attention mechanisms [19] are crucial in deep learning, particularly for image segmentation, by highlighting relevant features and suppressing irrelevant ones. They convert input features into query vectors (Q), keys (K), and values (V), calculate the similarity between Q and K, and apply SoftMax to generate attention weights, which weigh the values to produce focused output. Self-attention (also known as gradient attention to the raster product) provides global context by comparing each element with all others, enabling models to examine all areas of the image equally. This is commonly used in transformer architectures [33]. Given an input feature matrix $X \in R\, n \times d$, here the interest is calculated as follows [4]:

$$\boldsymbol{attention\ (Q,K,V) = softmax\ (QK^T/\sqrt{\mathrm{d_k}})V} \qquad (7)$$

Where:

- $Q = XWQ, K = XWK, V = XWV$
- $WQ, WK, WV \in R\, d \times dk$ are learnable weight matrices.
- dk is the dimension of the key vectors.

Channel attention (e.g., SENet) assigns weights to feature map channels, thereby enhancing important features by recalibrating channel responses without altering spatial information [12][34]. It improves feature representation but does not directly account for spatial locations, which may limit the accuracy of boundary segmentation. Squeeze-and-Excitation (SE) Block performs [5]:

Squeeze: Global average pooling:

$$\boldsymbol{z_c = \frac{1}{Hxw} \sum_{i=1}^{H} \sum_{j=1}^{w} x_c\ (i,j)} \qquad (8)$$

Excitation:

$$\boldsymbol{s_c = \sigma(w_2 . \delta(w_1 . z))} \qquad (9)$$

Where δ is ReLU, σ is sigmoid, and $W1, W2$ are FC layers. Recalibration:

$$\boldsymbol{\hat{x}_c = s_c \cdot x_c} \qquad (10)$$

While spatial attention (e.g., CBAM, BAM) focuses on important image regions by learning spatial weights at the pixel level, and promoting segmentation of small or irregular structures such as tumours [22][23]. Its limitation is low efficiency when a global context is required, as it emphasizes local information.

$$\boldsymbol{M_s} = \boldsymbol{\sigma Cf^{7\cdot7}([Avgpool(x); maxpool(x)])} \quad (11)$$

Where $f7 \times 7$ is a convolution layer with a 7×7 kernel, and the pooled features are concatenated.

$$\hat{\boldsymbol{x}} = \boldsymbol{M_s} \odot \boldsymbol{x} \quad (12)$$

Table 1 summarizes the key attention mechanisms, highlighting their primary strengths, weaknesses, and typical applications in segmentation.

Table 1: Comparison between types of attention mechanisms.

Attention Mechanism	Type	Strengths	Weaknesses	Common Use Cases
Self-Attention	Global Attention	Focuses on all parts of the image equally. Captures global relationships across the entire image.	Computationally expensive. May overlook local spatial details, which are crucial for precise segmentation.	Image classification, object detection, and large-scale segmentation tasks. Tumor segmentation (MRI scans).
Channel Attention	Global Attention	Focuses on important channels (feature maps). Enhances feature representation across channels.	Does not capture spatial context (pixel-level information). Can miss fine spatial details.	General segmentation tasks. Medical image segmentation (e.g., brain tumor segmentation[35]).
Spatial Attention	Local Attention	Focuses on important spatial regions (pixels). Good for small, irregular features (e.g., tumors).	Struggles with capturing global context or dependencies across the image.	Small object or tumor segmentation (e.g., CBAM for liver tumor). Segmentation of small structures (e.g., cell segmentation[36,37]

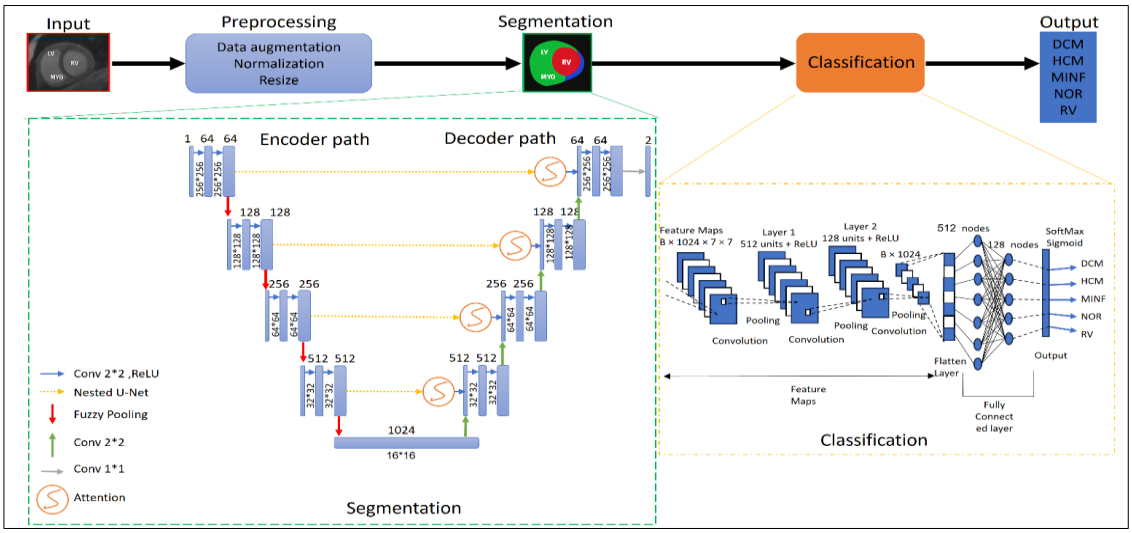

Fig. 2: The structure of the proposed hybrid model.

3.5. The Proposed Model

As shown in Fig. 2, upon input, cardiac MRI images (356×356×3) are pre-processed by resizing to 256×256, their intensity is normalized, and then systematically increased to improve the robustness of the model. The reinforcement pipeline included geometric transformations (rotation, flipping, translation, scaling, cropping, and shearing), color and statistical adjustments (brightness, contrast, gamma adjustments, noise injection, and elastic deformation), and advanced mixing-based strategies (Mixup and CutMix). Through these steps, the dataset was gradually expanded to approximately 662472 images, providing high contrast and ensuring better generalization of the proposed model.

It is then fed into the core segmentation network, which integrates Nested U-Net mechanisms, fuzzy pooling and attention for robust heart image analysis. The encoder consists of two Conv2D layers followed by fuzzy pooling at each level, which helps maintain fuzzy or

ambiguous boundaries. In the decoder, the attention gates focus on the LV, RV, and MYO regions, where each decoder block applies three sequential attention units - channel attention (CAM) to emphasize important feature channels, spatial attention (SAM) to focus on key anatomical regions, and self-attention (SA) to capture long-term dependencies. The CAM → SAM → SA sequence improves both local detail and global context (Fig. 3), improving segmentation and providing cleaner information and features are combined with skip and sampling connections in order to restore spatial detail. The network produces the final hash map through a 1×1 convolution followed by a SoftMax operation, which generates an output of 256×256. This design enables the integration of features on multiple scales, enhances boundary maintenance, and leverages focused attention to accurately segment the heart.

The classification process then takes place, where the classification module is designed as a convolutional neural network (CNN) classifier that acts on deep feature representations extracted from the backbone of the previous segmentation. The input consists of feature maps with a size of B × 1024 × 7 × 7 (with a batch size of B = 8), where 1024 denotes the number of channels and 7 × 7 represents the spatial resolution. To reduce dimensions and create a compact representation, the Global Average Aggregation Layer (GAP) aggregates spatial information, producing a vector of size B × 1024 while preserving the most prominent features. This feature vector is then passed through two fully connected layers: the first contains 512 units activated by ReLU, including leakage at a rate of 0.5 and optional batch normalization to improve stability; the second contains 128 units also activated by ReLU with leakage at a rate of 0.5 to enhance discriminative ability. Finally, the output layer consists of five units, each corresponding to one of the target heart conditions (dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), myocardial infarction (MINF), normal state (NOR), and right ventricular anomaly (RV)). The Softmax activation function is applied to this layer.

The model training process is done using Adam Optimizer (LR = 1e-4), batch size 8, for 100 epochs, using early stop and leak (0.5) to prevent over-processing, and save the best checkpoint.

The image segmentation threshold, which is a value that separates important objects in the foreground from the background, divides the image into two categories: foreground and background. In grayscale images, pixel values typically range from 0 (black) to 255 (white), and the threshold determines which pixels are classified as foreground or background. For example, when choosing a threshold of T=128, each pixel P≥T is classified as a foreground, and each pixel P< T is classified as background (0 or black).

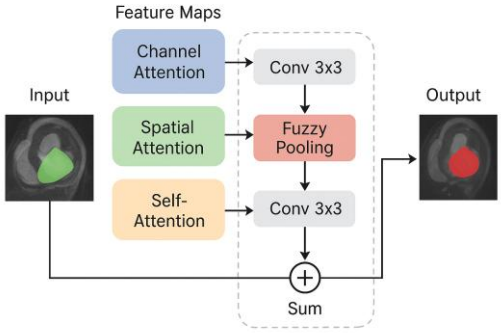

Fig. 3: Integration strategy.

3.6. System and Experimental Setup

Experiments have been conducted using Anaconda Navigator to manage environments and packages. The deep learning model was developed in PyTorch with GPU acceleration for medical image segmentation. Major libraries included NumPy for calculations, OpenCV for image preprocessing, and Matplotlib for visualizing results. The model's performance for LV, RV, and MYO was also evaluated, with the metrics summarized in Table 2.

Table 2: The evaluation metrics.

Metric	Description	Formula	Key Components
Dice Coefficient (DSC)	Measures overlap between predicted and ground-truth segmentations. Higher = better.	$DSC = 2\|X \cap Y\| / (\|X\| + \|Y\|)$	X: Prediction, Y: Ground truth
Hausdorff Distance (HD)	Measures the largest boundary distance between the prediction and ground truth. Lower = better.	$HD(X,Y) = max\{ max_x \in X\ min_y \in Y\ d(x,y), max_y \in Y\ min_x \in X\ d(x,y) \}$	x, y: Boundary points in prediction and ground truth
Accuracy	Percentage of correctly classified pixels (both foreground and background).	$Accuracy = (TP + TN) / (TP + TN + FP + FN)$	TP: True Positive TN: True Negative FP: False Positive FN: False Negative
Precision	Proportion of true positives among all predicted positives.	$Precision = TP / (TP + FP)$	
Recall (Sensitivity)	Proportion of true positives among all actual positives.	$Recall = TP / (TP + FN)$	

4. Experimental Results

In this section, the results of the experimental evaluation will be presented utilizing different attention mechanisms integrated featuring the nested U-Net model with fuzzy pooling. These experiments aim to evaluate the effect of varying attention mechanisms and threshold values on the segmentation performance of three major cardiac structures: the RV, LV, and MYO.

4.1. Case 1: using one type of attention mechanism with multiple threshold values

The first experiment used a single attention mechanism and tested the model with varying probability thresholds (0.3, 0.5, 0.7, and 0.9). The performance of the model was evaluated using the mean Dice scores, accuracy, and recall across the three cardiac structures, as summarized in Table 3. The relationship between the metrics and the change in threshold for the model with one type of attention is illustrated in Fig. 4.

Table 3: Gradual Threshold Adjustment Impact on Segmentation and Classification (Case 1: Single Attention).

Metric	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7	Threshold = 0.9
Dice – RV	0.9819	0.8755	0.8755	0.8755
Dice – LV	0.9800	0.8810	0.8810	0.8810
Dice – MYO	0.9843	0.8606	0.8606	0.8606
HD – RV	0.8016	0.8640	0.8640	0.8640
HD – LV	0.7667	0.8806	0.8806	0.8806
HD – MYO	0.7850	0.9210	0.9210	0.9210
Accuracy	0.9329	0.9671	0.9671	0.9671
Precision	0.9683	0.9535	0.9635	0.9835
Recall	0.9683	0.9534	0.9683	0.9535

The model achieves the highest performance at a threshold of 0.3, where the dice scores for RV, LV, and MYO exceed 0.98. In addition, as the threshold increases, a significant decrease in performance is observed, especially in the recall measure, indicating reduced sensitivity of the model.

In addition, it is essential to note that using different types of attention mechanisms (self-attention, spatial attention, or channel attention) under the same experimental conditions yields almost indistinguishable performance measures. Where the variance of the segmentation results between these types of interest does not exceed 0.0001 across all thresholds. This consistency suggests that the choice of attention mechanism has little effect on the overall model performance. Therefore, the probability threshold has a greater impact on segmentation accuracy and sensitivity than the specific type of attention applied.

4.2. Case 2: using two types of attention mechanisms with multiple threshold values

This experiment investigated the interaction between two attentional mechanisms, self-attention and spatial attention. This setup enables us to investigate how the combination of different types of attention impacts model performance, as illustrated in Fig. 5. The results are summarized in Table 4.

Figures 6-8 demonstrate examples of segmentation, where representative MRI slices are presented with masks corresponding to the left ventricle (LV), right ventricle (RV), and myocardium (MYO) and each structure is color-coded (LV in green, RV in blue, And MYO in red) This highlights the model's ability to clearly define heart boundaries and reinforce quantitative results with qualitative evidence. Combining self-attention and spatial attention improves the overall performance of the model compared to the previous individual attention approach. The dice score remains high at 0.3, but the average Hausdorff distance (HD) decreases significantly, especially for LV. The average dice score for the model across the three structures is 0.98208, indicating a clear improvement in hash accuracy.

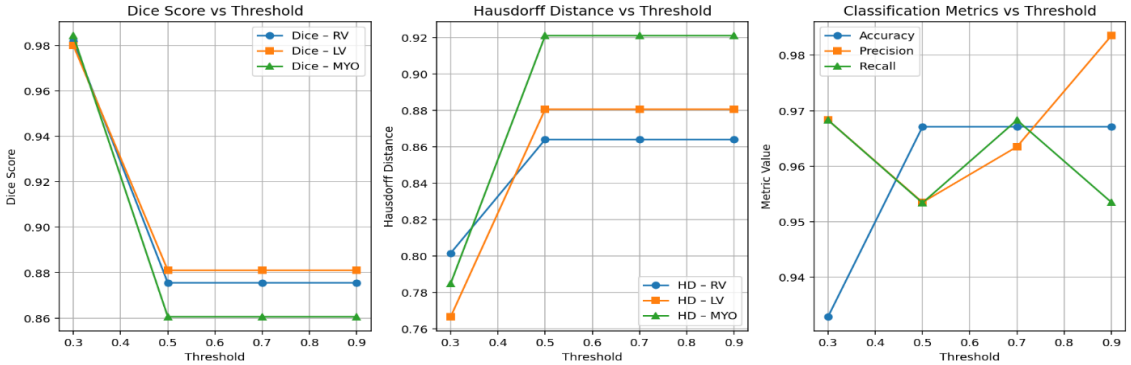

Fig. 4: Effect of Threshold Variation on Segmentation Metrics Using Single Attention Mechanism in Nested U-Net with Fuzzy Pooling.

Table 4: Gradual Threshold Adjustment Impact on Segmentation and Classification (Case 2-1: Dual Attention; Self-Attention and Spatial Attention).

Metric	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7	Threshold = 0.9
Dice – RV	0.9819	0.9819	0.9465	0.9287
Dice – LV	0.9800	0.9800	0.9470	0.9305
Dice – MYO	0.9843	0.9843	0.9430	0.9224
HD – RV	0.8016	0.8016	0.5344	0.4008
HD – LV	0.7667	0.7667	0.5111	0.3834
HD – MYO	0.7851	0.7851	0.5233	0.3925
Accuracy	0.9822	0.9471	0.9574	0.9572
Precision	0.9683	0.5165	0.6722	0.7500
Recall	0.9683	0.9434	0.9483	0.9435

From Table 5, which includes the results of using channel attention and spatial attention, it is observed that these results are very similar to those in Table 4, confirming the advantage of using multiple attention mechanisms to improve segmentation accuracy, as also seen in Fig. 9.

As shown in Table 6, the combination of channel and self-attention results in high scores on Dice at lower thresholds (for example, 0.98427 for MYO at 0.3), but the scores gradually decrease as the threshold increases. Correspondingly, the Hausdorff distance (HD) improves as the thresholds rise, suggesting better parametric accuracy and fewer outliers. On average, this configuration achieves a strong balance with Dice scores above 0.92 and average HD Values of around 40 pixels, making it suitable for applications that require precise and spatially coherent segmentations. Fig. 10 illustrates the relationship between performance measures and the change in threshold values.

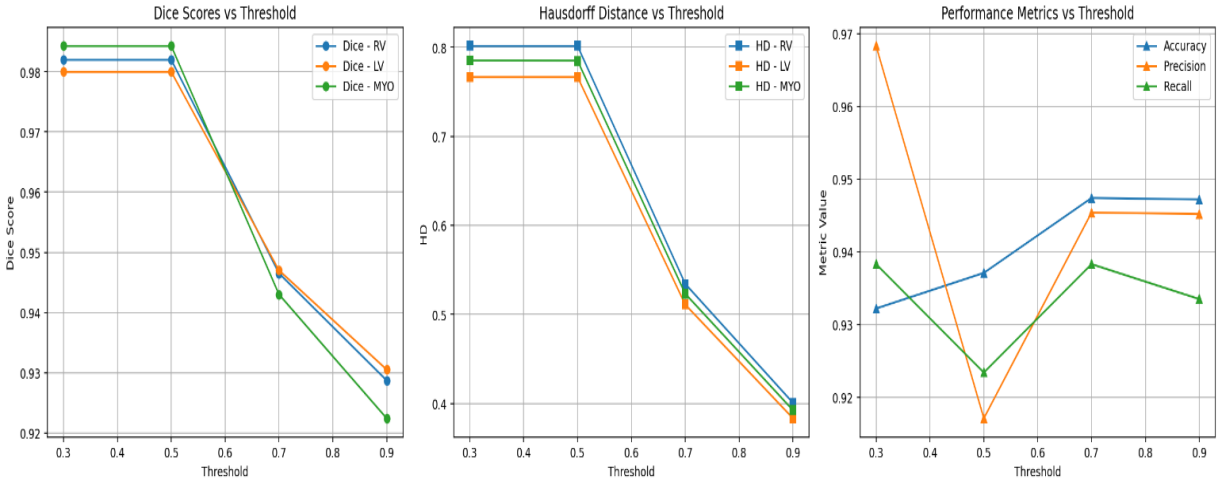

Fig. 5: Effect of Threshold Variation on Segmentation Metrics Using Self-Attention and Spatial Attention in Nested U-Net with Fuzzy Pooling.

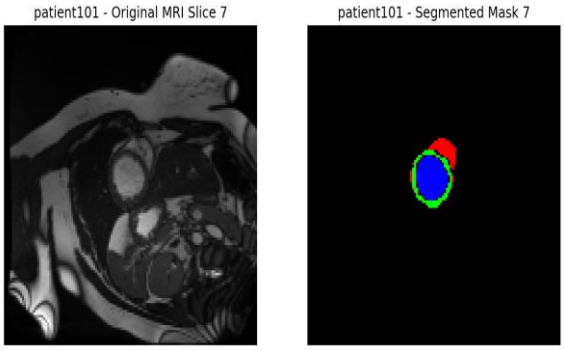

Fig. 6: MRI slices with segmented LV (green), RV (blue), and MYO (red) for patient 101.

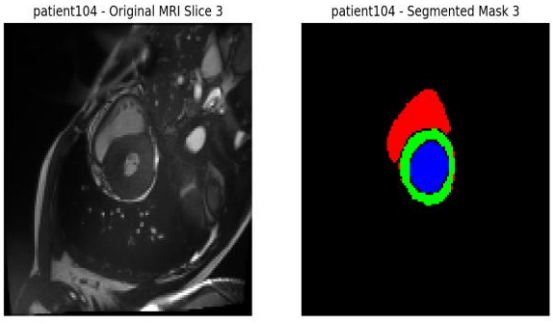

Fig. 7: MRI slices with segmented LV (green), RV (blue), and MYO (red) for patient 104.

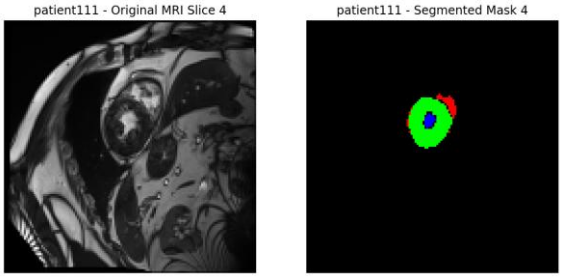

Fig. 8: MRI slices with segmented LV (green), RV (blue), and MYO (red) for patient 111.

Table 5: Gradual Threshold Adjustment Impact on Segmentation and Classification (Case 2-2: Dual Attention; Channel and Spatial Attention).

Metric	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7	Threshold = 0.9
Dice – RV	0.9820	0.9820	0.9465	0.9287
Dice – LV	0.9800	0.9800	0.9470	0.9305
Dice – MYO	0.9843	0.9843	0.9430	0.9224
HD – RV	0.8016	0.8016	0.5344	0.4008
HD – LV	0.7667	0.7667	0.5111	0.3834
HD - MYO	0.7851	0.7850	0.5233	0.3925
Accuracy	0.9322	0.9371	0.9474	0.9472
Precision	0.9683	0.9171	0.9454	0.9452
Recall	0.9383	0.9234	0.9383	0.9335

4.3. Case 3: Using three types of attention mechanisms together with multiple threshold values

In this case, all three types of attention mechanisms are integrated: channel attention, self-attention, and spatial attention. It is expected that this configuration will provide a comprehensive improvement to the segmentation process by leveraging the strengths of each type of attention. As shown in Table 7 and illustrated in Fig. 11, the results confirm this expectation.

The results showed superior performance for using the three attentional mechanisms to detect and interfere with the baseline at low thresholds (e.g,. 0.3) but worse parametric accuracy (high HD).

Table 6: Gradual Threshold Adjustment Impact on Segmentation and Classification (Case 2-3: Dual Attention; Channel and Self-Attention).

Metric	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7	Threshold = 0.9
Dice – RV	0.9820	0.9287	0.9110	0.9021
Dice – LV	0.9800	0.9305	0.9140	0.9058
Dice – MYO	0.9843	0.9224	0.9018	0.8915
HD – RV	0.8016	0.4008	0.2672	0.2004
HD – LV	0.7667	0.3834	0.2556	0.1917
HD – MYO	0.7851	0.3925	0.2617	0.1963
Accuracy	0.9309	0.9359	0.9364	0.9172
Precision	0.8165	0.7500	0.8278	0.8668
Recall	0.9113	0.9114	0.9313	0.9225

4.4. Comparison of attention types

Finally, compared the average Dice and HD values for the three different attentional mechanisms evaluated in this study. The results indicate that combining spatial attention and channel attention provides the best overall performance as measured by the average dice score and high accuracy, as shown in Table 8.

As shown in Fig. 12, different attention mechanisms have been compared using average dice score, accuracy, and recall. While individual units improved performance compared to the baseline, the combined use of channel, location, and self-attention achieved the best segmentation results by effectively capturing both local and global context.

Table 7: Gradual Threshold Adjustment Impact on Segmentation and Classification (Case 3: Triple Attention).

Metric	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7	Threshold = 0.9
Dice - RV	0.9820	0.9287	0.9120	0.9021
Dice - LV	0.9800	0.9305	0.9140	0.9058
Dice - MYO	0.9843	0.9224	0.9018	0.8915
HD – RV	0.8016	0.4008	0.2672	0.2004
HD – LV	0.7667	0.3834	0.2556	0.1917
HD - MYO	0.7851	0.3925	0.2617	0.1963
Accuracy	0.9320	0.9361	0.9374	0.9272
Precision	0.9165	0.7500	0.8278	0.8668
Recall	0.9283	0.9214	0.9353	0.9235

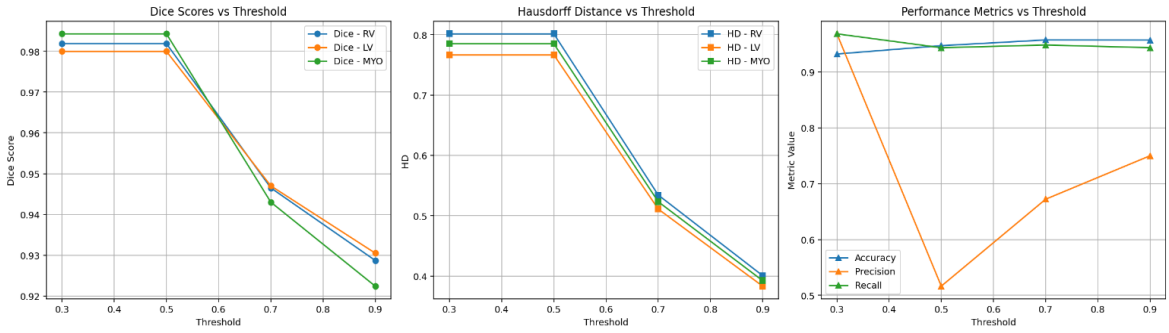

Fig. 9: Effect of Threshold Variation on Segmentation Metrics Using Channel Attention and Spatial Attention in Nested U-Net with Fuzzy Pooling.

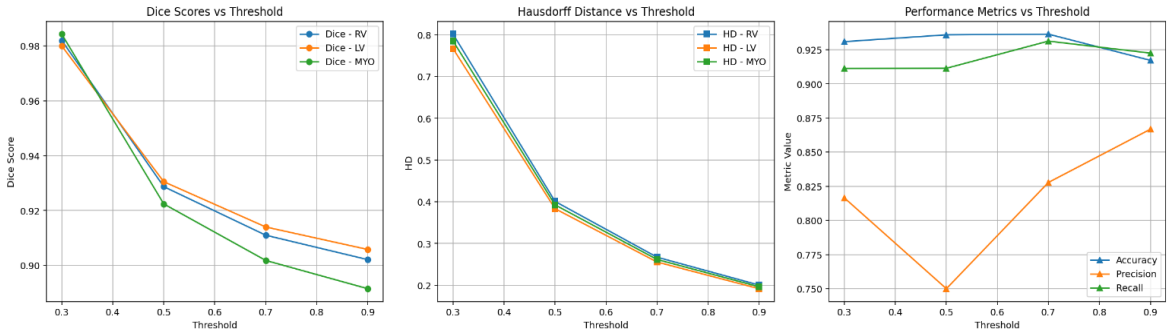

Fig. 10: Effect of Threshold Variation on Segmentation Metrics Using Channel and Self-Attention Mechanisms in Nested U-Net with Fuzzy Pooling.

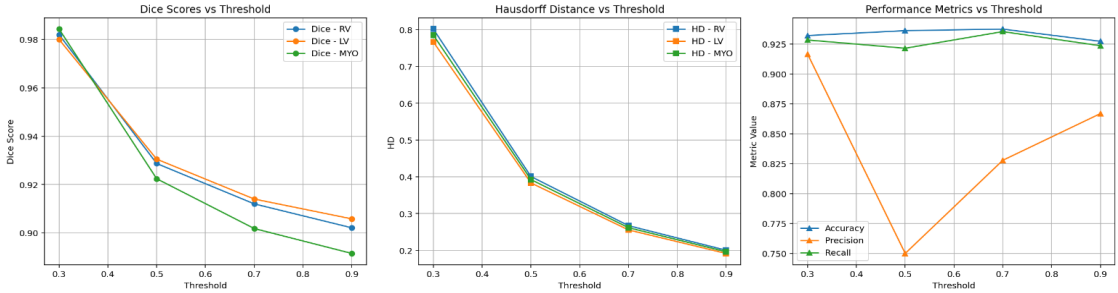

Fig. 11: Effect of Threshold Variation on Segmentation Metrics Using (Channel, Spatial, Self) Attention Mechanisms in Nested U-Net with Fuzzy Pooling.

Table 8: Comparison of attention types based on Average Dice, Precision, and Recall.

Rank	Case	Attention Type	Threshold	Avg. Dice	Precision	Recall	Reason
1	2.1	Self + Spatial	0.3	0.9821	0.9683	0.9683	Best Dice + recall, ideal for high sensitivity use cases
2	2.2	Channel + Spatial	0.3	0.9821	0.9683	0.9383	Equal Dice to Case 2, slightly lower recall
3	3	All (Channel + Self + Spatial)	0.7	0.9089	0.8278	0.9353	Great balance, best for precision-critical tasks
4	2.3	Channel + Self	0.7	0.9089	0.8278	0.9313	Similar to Case 5, slightly lower recall
5	1	One attention only	0.5	0.8724	0.9535	0.9534	High precision & recall, but significantly lower Dice

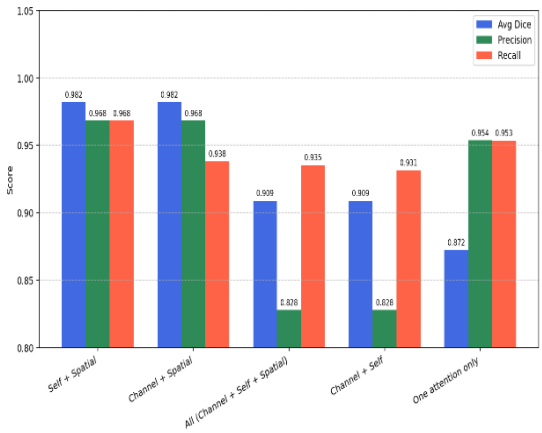

Fig. 12: Comparison of Attention Types Based on Avg.Dice, precision, and Recall.

4.5. Analysis of Training and Validation Curves

Fig. 13 shows the evolution of training and verification accuracy over the ages. The continued rise in training accuracy reflects the model's ability to accurately fit the data, while verification accuracy indicates its generalization to unseen samples. The close alignment of both curves indicates minimal overprocessing and demonstrates that the model converges effectively. Fig. 14 illustrates the reduction in loss during training and verification. Both curves are constantly decreasing, which confirms that the model reduces errors and generalizes well. Low and stable loss values indicate effective training and strong generalization performance.

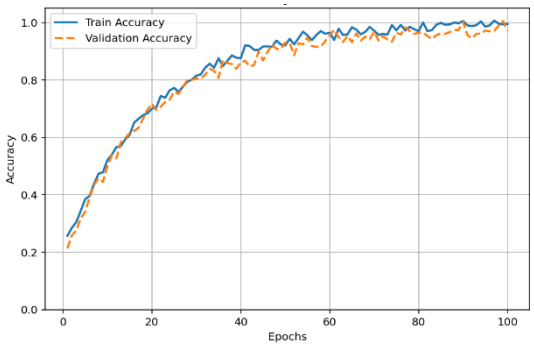

Fig. 13: Training and Validation Accuracy curves.

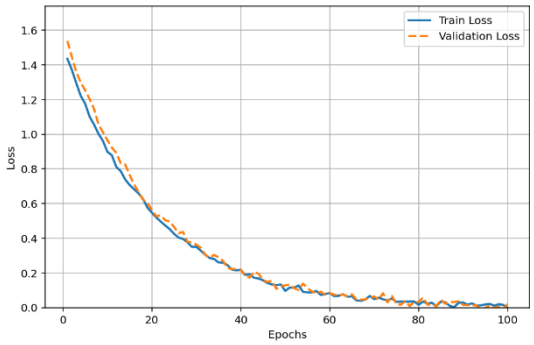

Fig. 14: Training and Validation Loss curves.

4.6. Disease classification results

The proposed model, an overlapping U-Net with attention (self and spatial) and the fuzzy pooling mechanism, has shown excellent classification performance in all five cardiac conditions: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), myocardial infarction (MINF), normal right ventricle (NOR), and abnormal right ventricle (RV). The model achieved an overall accuracy of 96%, with perfect accuracy, recall, and F1 scores (1.00) for the DCM, NOR, and RV categories. The HCM and MINF have also demonstrated strong performance, with F1 scores of 0.96, which illustrates the model's reliability in distinguishing similar heart diseases, as shown in Table 9. These results confirm the effectiveness of the classification pipeline and complement the model's high segmentation accuracy and computational efficiency. The overall accuracy is 0.96.

Table 9: Classification results.

Class	Precision	Recall	F1-score
DCM	1.00	1.00	1.00
HCM	1.00	1.00	1.00
MINF	0.90	0.90	0.90
NOR	0.90	0.90	0.90
RV	1.00	1.00	1.00

The classification results are summarized in the confusion matrix of Fig. 15, which demonstrates the model's robust classification capability, providing clear insight into true positives, false positives, and false classifications across the five categories. The model shows strong taxonomic performance, with remarkable accuracy in differentiating similar heart diseases.

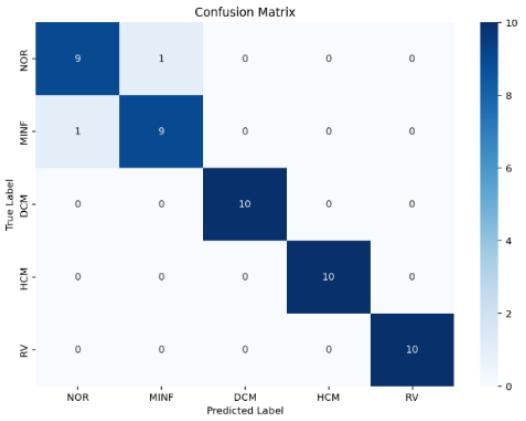

Fig. 15: Confusion Matrix for the ACDC dataset.

The model was evaluated on a new unseen dataset of 220 DICOM images across four classes: normal (N), hypertension (HYP), heart failure (HF), and ischemic heart failure (HF-I). The model achieved an overall accuracy of 96% on the new dataset, with class-level accuracy and recall ranging from 94% to 98% (Table 10).

Fig. 16 shows a confusion matrix highlighting clinically understandable misclassifications, such as between HF-I and HYP or HF and HF-I.

Table 10: Classification Report on the Sunnybrook Cardiac Dataset (SCD).

Class	Precision	Recall	F1-Score
N	0.98	0.98	0.98
HYP	0.96	0.96	0.96
HF	0.96	0.98	0.97
HF-I	0.94	0.93	0.94
Accuracy		0.96	

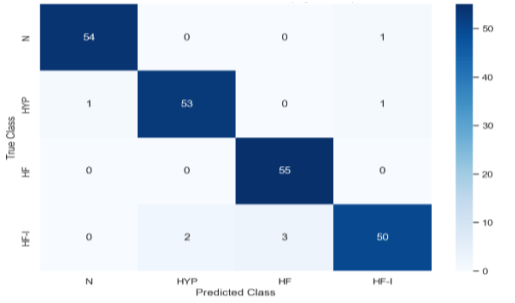

Fig. 16: Confusion Matrix for the Sunnybrook Cardiac Dataset (SCD).

4.7. Comparison with previous research

To contextualize the results of this research, a comparison is made between the proposed model, an overlapping U-Net with attention (self and spatial), and the fuzzy pooling mechanism, and previously published methods remove it. As shown in Table 11, the proposed approach significantly outperforms previous models in terms of Dice, accuracy, and recall.

Table 11: Classification performance comparison of the proposed model with recent.

Method	Dice	Precision	Recall
U-Net [7]	0.8648	0.8234	0.9106
R2U-Net [8]	0.8977	0.9360	0.8624
UNet++ [9]	0.8953	0.8637	0.9293
Attention U-Net [38]	0.9138	0.9154	0.9122
Attention R2U-Net [7]	0.8979	0.9088	0.8873
ANU-Net [7]	0.9431	0.9519	0.9344
MSF-TransUNet [39]	.09152	-	-
Proposed model	0.9821	0.9683	0.9683

In the comparative analysis, the proposed overlapping U-Net with the attention (self and spatial) and fuzzy pooling mechanism was significantly superior to previous models in the ACDC MRI dataset, as shown in Table 12. This model achieves the highest average Dice score of 98.21%, clearly surpassing all basic and modern methods, including TF-Unet (91.72%), Swin U-Net (90.00%), and TransUNet (89.71%). In addition, it exhibits superior performance in all cardiac structures: RV (98.19%), MYO (98.43%), and LV (98.00%), demonstrating its robustness in segmenting anatomical and pathological differences. This improvement highlights the effectiveness of incorporating both attention and fuzzy pooling mechanisms in enhancing feature representation and demarcation in cardiac MRI.

Table 12. Segmentation performance comparison of the proposed model with recent studies.

Method	Avg. Dice	RV (%)	MYO (%)	LV (%)
U-Net [12]	87.55	87.10	80.36	94.92
Attn U-Net [38]	86.75	87.58	79.20	93.47
VIT [14]	81.45	81.46	70.71	92.18
R50 + VIT [14]	87.57	86.07	81.88	94.75
TransUNet [40]	89.71	88.86	84.54	95.73
Swin U-Net [15]	90.00	88.55	85.62	95.83
TF-Unet [41]	91.72	90.16	89.40	95.60
MSF-TransUNet[39]	91.52	90.25	87.72	96.59
Proposed model	98.21	98.19	98.43	98.00

5. DISCUSSION

This section presents a comprehensive analysis of the experimental results obtained by evaluating the proposed model with different attentional mechanisms under multiple threshold values.

5.1. Effect of single attention mechanism (case 1)

The results of case 1 indicate that using a single attentional mechanism (whether channel, self, or spatial) will lead to robust performance at lower probability thresholds. Specifically, a threshold value of 0.3 produced the highest dice scores across all heart structures (RV: 0.9819, LV: 0.9800, MYO: 0.9843). For all that, these grades were accompanied by high Hausdorff distances (HD), indicating a less precise delimitation. As the threshold was increased from 0.3 to 0.9, there was a marked decrease in dice and recall scores, indicating a decrease in the model's sensitivity. In contrast, HD values improved, indicating clearer detection of boundaries but at the cost of a loss of fine anatomical details. This highlights the classic trade-off between hash completeness and boundary accuracy.

5.2. Effect of dual attention mechanisms (case 2)

The combinations of two attention mechanism types improved segmentation and classification results. For example, self-spatial attention+ (case 2-1) maintained high dice scores (~0.982) at a threshold of 0.3 while also significantly reducing HD compared to case 1. This demonstrates a synergistic effect between spatial context (spatial attention) and global relevance (self-attention).

In Case 2-2 (channel + spatial attention), performance was nearly identical to Case 2-1, confirming that different combinations of attention types lead to similar benefits. The best balance was observed across the Dice and HD scales at a threshold of 0.5, indicating that this setting is optimal for maintaining segmentation accuracy while improving geometric coherence.

Additionally, cases 2-3 (channel + self-attention) demonstrated high segmentation quality at low thresholds and a gradual improvement in HD at higher thresholds. This reinforces the observation that dual attention mechanisms help balance sensitivity and spatial accuracy more effectively than a single type of attention.

5.3. Effect of tripartite attention mechanisms (case 3)

In the third case, the three attention mechanisms were combined. Although dice scores at threshold 0.3 matched previous best results (~0.982 for RV and ~0.984 for MYO), gains compared to dual-attention configurations were marginal. HD values followed similar patterns, steadily improving as the threshold increased.

These results suggest that while triple attention provides the most comprehensive feature optimization, marginal gains compared to dual attention settings do not always justify the additional computational complexity, especially in real-time or resource-limited clinical settings.

5.4. Threshold sensitivity analysis

The threshold values in all discussed cases had a profound effect on hashing and classification performance. Low thresholds (0.3) favored higher dice and calls, indicating effective hashing coverage, although this was often accompanied by lower parametric accuracy. While the mid-range thresholds (0.5) provided the best balance between hash quality and boundary accuracy, they proved to be the most stable configuration. While high thresholds (0.7 and 0.9) enhanced the accuracy and clarity of boundaries, they significantly reduced dice and recall, making them less suitable for tasks that prioritize complete segmentation.

5.5. Comparative evaluation of attention strategies

A comparative overview of attention mechanisms indicates that individual attention models are sensitive to threshold shifts and achieve best performance only in limited circumstances. It also indicates that dual attention mechanisms provide robust and generalizable performance across thresholds, especially for constructs involving spatial attention. Furthermore, although triple attention integration is slightly more effective, the results in diminished returns compared to dual attention settings.

Finally, in general, spatial attention has consistently improved boundary accuracy, while channel attention and self-attention have enhanced feature recognition and structure segmentation.

5.6. General effects and robustness of the model

The proposed nested U-Net with a fuzzy pooling structure, especially when enhanced by attentional mechanisms, shows generalizability and strong adaptability. The results demonstrate the model's ability to capture global and local structural patterns in cardiac MRI data.

In addition, experiments highlight the importance of threshold tuning in segmentation models both in terms of achieving clinically acceptable sensitivity and reducing false boundary detections.

The results of the light classification demonstrate the model’s strength, robustness, and generalizability of the proposed model in accurately distinguishing between five clinically significant cardiac conditions. The high accuracy and recall in all categories, especially the perfect scores for DCM, NOR, and RV, confirm the model's ability to accurately identify both common and less representative conditions. It is worth noting that F1 scores of 0.96 for HCM and MINF indicate strong performance even in cases where stratigraphic overlap and similar morphological features may challenge classification. These results indicate that combining high-quality segmentation with an effective classification line enhances diagnostic accuracy. In addition, the high overall accuracy (96%) and balanced performance by category indicate that the model is well-suited for clinical applications, where accurate differentiation between heart diseases is crucial for appropriate treatment planning.

On the SCD dataset, the model achieved 96% accuracy and F1 scores $\geq$0.94, with minor HF-I misclassifications, demonstrating robustness for cardiovascular screening.

6. CONCLUSION

In this study, we propose a nested U-Net model enhanced with fuzzy clustering and attention mechanisms for cardiac MRI segmentation. Attention to channel and location improved segmentation quality, and combining them provided the best balance between dice score and Hausdorff distance; in contrast, triple attention added little benefit at a higher cost. A threshold of 0.5 yielded optimal results, achieving an average dice score of 98.2% and a rating accuracy of 96% across five cardiac cases. The model is efficient, enabling real-time clinical use. Key contributions include the development of a hybrid architecture, analysis of attention strategies, and threshold optimization. While the results are promising, limitations include untested generalizability to other imaging modalities, increased computational requirements with complex attentional configurations, and limited model interpretability. The model was trained on limited cardiac MRI data, making generalization challenging, and its performance depends on preprocessing, thresholding, and postprocessing.

Future research should focus in improving transparency, expanding multi-organ segmentation, and utilizing larger, multicenter datasets to address class imbalance, particularly in conditions such as dilated cardiomyopathy. Integrating temporal information and clinical metadata further improves reliability. Overall, this study presents a balanced and efficient model architecture for cardiac MRI analysis, demonstrating clear clinical potential. Adaptive

thresholding strategies can also be incorporated to enhance robustness across unseen datasets without manual intervention.

REFERENCES

[1] Y. M. A. Mohammed, S. El Garouani, and I. Jellouli, "A survey of methods for brain tumor segmentation-based MRI images," J. Comput. Des. Eng., vol. 10, no. 1, pp. 266–293, 2023, doi: 10.1093/jcde/qwac141.

[2] D. A. Shoieb, K. M. Fathalla, S. M. Youssef, and A. Younes, "CAT-Seg: cascaded medical assistive tool integrating residual attention mechanisms and Squeeze-Net for 3D MRI biventricular segmentation," Phys. Eng. Sci. Med., vol. 47, no. 1, pp. 153–168, 2024, doi: 10.1007/s13246-023-01352-2.

[3] J. Bleker et al., "A deep learning masked segmentation alternative to manual segmentation in biparametric MRI prostate cancer radiomics," Eur. Radiol., vol. 32, no. 9, pp. 6526–6535, 2022, doi: 10.1007/s00330-022-08712-8.

[4] D. Li, S. Yin, Y. Lei, J. Qian, C. Zhao, and L. Zhang, "Segmentation of White Blood Cells Based on CBAM-DC-UNet," IEEE Access, vol. 11, no. December 2022, pp. 1074–1082, 2023, doi: 10.1109/ACCESS.2022.3233078.

[5] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Towards Efficient Motor Imagery EEG-based BCI Systems using DCNN," 2024 Arab ICT Conf. AICTC 2024, pp. 59–66, 2024, doi: 10.1109/AICTC58357.2024.10735028.

[6] T., Zhang, J. Xiang, S. Zhenxiao, Z. Yi, and Y. Yunqiang. "Software defect prediction based on machine learning algorithms," In 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 520-525. IEEE, 2019, doi:10.1109/ICCC47050.2019.9064412.

[7] M. H. Guo et al., "Attention mechanisms in computer vision: A survey," Comput. Vis. Media, vol. 8, no. 3, pp. 331–368, 2022, doi: 10.1007/s41095-022-0271-y.

[8] H. Ye, R. Zhou, J. Wang, and Z. Huang, "FMAM-Net: Fusion Multi-Scale Attention Mechanism Network for Building Segmentation in Remote Sensing Images," IEEE Access, vol. 10, no. December, pp. 134241–134251, 2022, doi: 10.1109/ACCESS.2022.3231362.

[9] Sun, J., Chen, K., He, Z., et al., "Medical image analysis using improved SAM-Med2D: segmentation and classification perspectives," BMC Medical Imaging, vol. 24, article 241, 2024, doi: 10.1186/s12880-024-01401-6.

[10] W. Kim, Y. Ahn, J. Kim, and B. Shim, "Towards Deep Learning-aided Wireless Channel Estimation and Channel State Information Feedback for 6G," J. Commun. Networks, vol. 25, no. 1, pp. 61–75, 2023, doi: 10.23919/JCN.2022.000037.

[11] A. Al-Saegh, A. Daood, and M. H. Ismail, "Dual Optimization of Deep CNN for Motor Imagery EEG Tasks Classification," Diyala J. Eng. Sci., vol. 17, no. 4, pp. 75–91, 2024, doi: 10.24237/djes.. 2024.17405.

[12] A. Al-Saegh, "Identifying a Suitable Signal Processing Technique for MI EEG Data," Tikrit J. Eng. Sci., vol. 30, no. 3, pp. 140–147, 2023, doi: 10.25130/tjes 30.3.14.

[13] T. Tassew, B. A. Ashamo, and X. Nie, "Multimodal MRI brain tumor segmentation using 3D attention UNet with dense encoder blocks and residual decoder blocks," Multimed. Tools Appl., pp. 1–28, 2024, doi: 10.1007/s11042-024-18942-1.

[14] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation BT - Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support," Miccai, vol. 11045, no. 2018, pp. 3–11, 2018, doi: 10.1007/978-3-030-00889-5.

[15] Y. Jia, L. Liu, and C. Zhang, "Moon Impact Crater Detection Using Nested Attention Mechanism Based UNet++," IEEE Access, vol. 9, pp. 44107–44116, 2021, doi: 10.1109/ACCESS.2021.3066445.

[16] G. Huang et al., "Channel-Attention U-Net: Channel Attention Mechanism for Semantic Segmentation of Esophagus and Esophageal Cancer," IEEE Access, vol. 8, pp. 122798–122810, 2020, doi: 10.1109/ACCESS.2020.3007719.

[17] Z. Li, H. Zhang, Z. Li, and Z. Ren, "Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation," Appl. Sci., vol. 12, no. 14, 2022, doi: 10.3390/app12147149.

[18] B., Philippe, S. Billings, N. Joshi, and J. Albayda. "Automated diagnosis of myositis from muscle ultrasound: exploring the use of machine learning and deep learning methods," PloS one 12, no. 8 (2017): e0184059, doi: 10.1371/journal.pone.0184059.

[19] M. Cui, K. Li, J. Chen, and W. Yu, "CM-Unet: A Novel Remote Sensing Image Segmentation Method Based on Improved U-Net," IEEE Access, vol. 11, no. June, pp. 56994–57005, 2023, doi: 10.1109/ACCESS.2023.3282778.

[20] O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," no. Midl, 2018, doi:10.48550/arXiv.1804.03999.

[21] D. Cheng et al., "Attention based multi-scale nested network for biomedical image segmentation," Heliyon, vol. 10, no. 14, p. e33892, 2024, doi: 10.1016/j.heliyon.2024.e33892.

[22] H. Wang, S. Qiu, B. Zhang, and L. Xiao, "Multilevel Attention Unet Segmentation Algorithm for Lung Cancer Based on CT Images," Comput. Mater. Contin., vol. 78, no. 2, pp. 1569–1589, 2024, doi: 10.32604/cmc.2023.046821.

[23] H. Cao et al., "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 13803 LNCS, pp. 205–218, 2023, doi: 10.1007/978-3-031-25066-8_9.

[24] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, & Y. Yu, "nnFormer: Interleaved Transformer for Volumetric Segmentation," IEEE Transactions on Image Processing, vol. 32, no. 7, pp. 4036-4045, 2023, doi: 10.1109/TIP.2023.3293771

[25] D. E. Diamantis & D. K. Iakovidis, "Fuzzy Pooling," IEEE Transactions on Fuzzy Systems, vol. 29, no. 11, pp. 3481-3488, 2021, doi: 10.1109/TFUZZ.2020.3024023.

[26] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images," Irbm, vol. 43, no. 4, pp. 290–299, 2022, doi: 10.1016/j.irbm.2021.06.003.

[27] H. Li, L. Wang, and S. Cheng, "HARNU-Net: Hierarchical Attention Residual Nested U-Net for Change Detection in Remote Sensing Images," Sensors, vol. 22, no. 12, pp. 1–26, 2022, doi: 10.3390/s22124626.

[28] X. Fan, Y. Lu, J. Hou, F. Lin, Q. Huang, and C. Yan, "DMC-UNet-Based Segmentation of Lung Nodules," IEEE Access, vol. 11, no. October, pp. 110809–110826, 2023, doi: 10.1109/ACCESS.2023.3322437.

[29] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," IEEE Access, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[30] Liu, W., Sun, Y., & Ji, Q., "MDAN-UNet: Multi-Scale and Dual Attention Enhanced Nested U-Net Architecture for Segmentation of Optical Coherence Tomography Images," Algorithms, vol. 13, no. 3, article 60, 2020, doi: 10.3390/a13030060.

[31] O. Bernard et al., "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," IEEE Trans. Med. Imaging, vol. 37, no. 11, pp. 2514–2525, 2018, doi: 10.1109/TMI.2018.2837502. Available: https://www.cardiacatlas.org/sunnybrook-cardiac-data/.

[32] P. Radau, Y. Lu, K. Connelly, G. Paul, A. J. Dick, and G. A. Wright, "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI.," Midas J., 2022, doi: 10.54294/g80ruo. Available: https://www.cardiacatlas.org/sunnybrook-cardiac-data/.

[33] I. Mohsen Asghari, D. Shi, and Y. Mike Banad. "T1-weighted MRI-based brain tumor classification using hybrid deep learning models," Scientific Reports 15, no. 1 (2025): 7010, doi: 10.1038/s41598-025-92020-w.

[34] J. Liu et al., "Multi-Scale Hybrid Attention Convolutional Neural Network for Automatic Segmentation of Lumbar Vertebrae From MRI," IEEE Access, vol. 12, no. January, pp. 77999–78013, 2024, doi: 10.1109/ACCESS.2024.3407833.

[35] S. R. Gunasekara, H. N. T. K. Kaldera, and M. B. Dissanayake, "A Systematic Approach for MRI Brain Tumor Localization and Segmentation Using Deep Learning and Active Contouring," J. Healthc. Eng., vol. 2021, 2021, doi: 10.1155/2021/6695108.

[36] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, & S. Zhang, "CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation," IEEE Transactions on Medical Imaging, vol. 40, no. 2, pp. 699-711, 2021, doi: 10.1109/TMI.2020.3035253.

[37] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," J. Med. Imaging, vol. 6, no. 01, p. 1, 2019, doi: 10.1117/1.jmi.6.1.014006.

[38] J. Schlemper et al., "Attention gated networks: Learning to leverage salient regions in medical images," Med. Image Anal., vol. 53, pp. 197–207, 2019, doi: 10.1016/j.media.2019.01.012.

[39] J. Chen et al., "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," pp. 1–13, 2021, doi:10.48550/arXiv.2102.04306.

[40] A. Dosovitskiy et al., "an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale," ICLR 2021 - 9th Int. Conf. Learn. Represent., 2021, [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[41] Z. Fu, J. Zhang, R. Luo, Y. Sun, D. Deng, and L. Xia, "TF-Unet: An automatic cardiac MRI image segmentation method," Math. Biosci. Eng., vol. 19, no. 5, pp. 5207–5222, 2022, doi: 10.3934/mbe.2022244.