DeepCardioNet: Efficient Left Ventricular Epicardium and Endocardium Segmentation using Computer Vision

Bukka Shobharani[1], Mr. S Girinath[2], Dr. K. Suresh Babu[3],
Dr. J.Chenni Kumaran[4], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[5], Dr. S. Farhad[6]
Research Scholar, Department of English, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur, Andhra Pradesh, India[1]
Assistant Professor Computer Applications, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering college)
Tirupati, Andhra Pradesh, India[2]
Professor, Department of Biochemistry, Symbiosis Medical college for Women,
Symbiosis International (Deemed University), Pune, India[3]
Professor, Department of CSE, Saveetha School of Engineering, SIMATS, Chennai, India[4]
Faculty of Informatics and Computing, UniSZA University, Malaysia[5]
Associate Professor, Department of English, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur, Andhra Pradesh, India[6]

***Abstract*—In the realm of medical image analysis, accurate segmentation of cardiac structures is essential for accurate diagnosis and therapy planning. Using the efficient Attention Swin U-Net architecture, this study provides DEEPCARDIONET, a novel computer vision approach for effectively segmenting the left ventricular epicardium and endocardium. The paper presents DEEPCARDIONET, a cutting-edge computer vision method designed to efficiently separate the left ventricular epicardium and endocardium in medical pictures. The main innovation of DEEPCARDIONET is that it makes use of the Attention Swin U-Net architecture, a state-of-the-art framework that is well-known for its capacity to collect contextual information and complicated attributes. Specially designed for the segmentation task, the Attention Swin U-Net guarantees superior performance in identifying the relevant left ventricular characteristics. The model's ability to identify positive instances with high precision and a low false positive rate is demonstrated by its good sensitivity, specificity, and accuracy. The Dice Similarity Coefficient (DSC) illustrates the improved performance of the proposed method in addition to accuracy, showing how effectively it captures spatial overlaps between predicted and ground truth segmentations. The model's generalizability and performance in a variety of medical imaging contexts are demonstrated by its application and evaluation across many datasets. DEEPCARDIONET is an intriguing method for enhancing cardiac picture segmentation, with potential applications in clinical diagnosis and treatment planning. The proposed method achieves an amazing accuracy of 99.21% by using a deep neural network architecture, which significantly beats existing models like TransUNet, MedT, and FAT-Net. The implementation, which uses Python, demonstrates how versatile and useful the language is for the scientific computing community.**

***Keywords*—DeepCardioNet; attention swin U-Net; ventricular epicardium; endocardium; computer vision approach**

I. INTRODUCTION

Cardiovascular conditions remain a leading cause of morbidity and mortality worldwide, challenging advanced medical imaging ways for precise opinion and treatment [1]. Among these, the segmentation of cardiac structures, similar as the left ventricular epicardium and endocardium, plays a pivotal part in understanding cardiac function and pathology [2]. Accurate segmentation is challenging due to the complex anatomical structures, variations in cardiac morphology, and the need for real- time processing in clinical settings [3]. In response to these challenges, the exploration community has witnessed a swell in the development of computer vision approaches for medical image segmentation [4]. Among the notable benefactions in this sphere, the deepcardionet introduces a new computer vision approach designed specifically for the effective segmentation of the LV epicardium and endocardium [5]. Using advanced deep learning ways, this model aims to overcome the limitations of traditional segmentation styles, offering bettered delicacy and computational effectiveness [6].

The foundation of deepcardionet lies in its application of a customized neural network armature inspired by deep learning principles [7]. The integration of a sophisticated encoder-decoder structure, conceivably told by proven infrastructures like U-Net or other innovative designs, empowers the model to capture intricate features and spatial dependences within cardiac images [8]. The objectification of skip connections, batch normalization, and task-specific activation functions further refines the segmentation performance, icing the model's rigidity to the complications of cardiac imaging [9]. This paper presents a comprehensive disquisition of the proposed deepcardionet methodology, expounding its armature, training strategies, and performance criteria [10]. The effectiveness of the model is underlined by its capability to delineate the left ventricular epicardium and endocardium with high perfection, therefore furnishing a precious tool for clinicians in their individual trials. The exploration not only contributes to the growing body of knowledge in computer-backed medical image analysis but also holds pledge for transformative operations in cardiovascular healthcare.

In a period marked by an unknown affluence of visual data, the field of computer vision stands at the van of technological invention, empowering machines with the capability to interpret and make opinions grounded on visual information [11]. This transformative capability has set up operations across different disciplines, from independent vehicles and robotics to healthcare and entertainment. At the heart of this paradigm shift is the continual development of new computer vision styles that strive to enhance delicacy, effectiveness, and severity in the interpretation of visual content [12]. Over the once decades, computer vision has evolved from early image processing ways to sophisticated deep learning infrastructures. From traditional styles addressing image bracket and object discovery to contemporary approaches probing into semantic segmentation and scene understanding, the diapason of computer vision operations continues to expand [13].

Recent times have witnessed substantial advancements in medical image analysis, driven by the community between computer vision and deep learning ways [14]. Despite these strides, there exists a demand for technical results acclimatized to the complications of cardiac image segmentation [15]. The provocation behind the exploration lies in the recognition of the clinical significance of precise left ventricular segmentation and the hunt for a system that not only surpasses being approaches but also aligns with the need for nippy and effective analyses. Segmenting the left ventricular epicardium and endocardium poses challenges embedded in the complexity of cardiac structures and the variability observed across patient populations [16] [17]. The dynamic nature of the heart, coupled with the essential noise and vestiges present in medical images, necessitates a sophisticated approach able of robustly handling different scripts [18].

In recent decades, computer vision has surfaced as a transformative field, catalyzing advancements across various disciplines by enabling machines to interpret and understand visual information. With the proliferation of image and videotape data in moment's digital age, the development of robust computer vision approaches has come consummate. This exploration seeks to contribute to this dynamic geography by proposing an innovative computer vision approach acclimatized to address a specific problem, promising both enhanced delicacy and effectiveness. The arrival of vast datasets and the elaboration of deep learning ways have fueled unknown progress in computer vision operations. From image bracket to object discovery and segmentation, computer vision has revolutionized diligence ranging from healthcare to independent vehicles. Still, challenges persist, particularly in scripts where fine-granulated details, complex structures, or real- time processing are essential. It's within this environment that the proposed computer vision approach emerges, aiming to attack nuanced challenges in a targeted sphere.

The key contributions of the article are,

- The introduction of DEEPCARDIONET, cutting-edge deep neural network architecture created especially for the effective segmentation of the endocardium and left ventricle in medical images.
- Using the capabilities of the Swin Transformer and attention mechanisms in the state-of-the-art Attention Swin U-Net architecture to gather detailed characteristics and contextual information necessary for precise segmentation.
- Proving the suggested method's durability and applicability across a number of trials, confirming its accuracy and consistency in correctly segmenting left ventricular structures in a range of medical imaging circumstances.
- Advancing the area of medical image analysis by developing a precise and effective segmentation technique that addresses the unique difficulties in defining the left ventricular epicardium and endocardium

The remainder of the article is structured as follows: Section II, III and IV include the related works, problem statement and methodology of the article. Section V includes results and discussion. The article is concluded in Section VI.

II. Related Works

A crucial first step in computing clinical markers such wall consistency, ventricle volume, and expulsion bit is segmenting cardiac medical pictures [19]. The paper presents the Ls Unet structure, which effectively segments cardiac cine MR images by combining multi-channel, fully CNN, and circular shape position-set styles. The division job in this framework is trained using the multi-channel DL method in order to identify the LV endocardial and epicardial outlines. In order to ensure the delicate and reliable division, segmented outlines extracted from the multi-channel DL approach are then integrated into a positional data set that is specifically dedicated to identifying annular forms. The automated method that was suggested was assessed to be 95. In compared with the benchmark norm, the combination of multi-channel DL and circular shape position-set segmented method obtained great delicateness, with total baseline values for LV endocardium and epicardium delineation reaching 92.15 and 95.42, respectively. It offers a novel approach for fully automated segmentation of the LV endocardium and epicardium from several MRI datasets. In comparison with additional current methods and the source of information, the suggested process is reliable and accurate.

Because it is crucial in determining patient assessment as well as therapy paths, automatic division of the cardiac left ventricle with scars continues to be a challenging and therapeutically important endeavor [20]. An independent verification methodology was developed employing OOD both inside and outside validation cohorts, as well as intra-observation and inter-observer variation in ground truth, to ensure the conceptualization of the frames. To obtain the best segmentation results, the frame combines DL with conventional computer vision techniques. Although the DL technique makes use of DL methods and infrastructure, the classic method makes use of multi-atlas methods, active outlines, and k-means. The research established that, with the exception of situations in which breath displacement error occurred, the conventional image recognition technique produced more accurate findings than DL. In both inside and outside OOD groups, respectively, the ideal outcome from the frame obtained robust and generalized outcomes values of 82.8 ±6.4 and 72.1 ±4.6. The created framework provides a powerful outcome for LGE-MRI-based automated segmentation of the scarred left ventricle. In contrast to modern techniques, it produces impartial findings across various medical facilities and retailers without the requirement for calibration or training in sanitized cohorts. Specialists can handle the challenge of fully automated separation of the LV with marks based on a single-modality cardiovascular examination with the help of this framework.

According to clinical opinions about cardiovascular problems, croakers should undergo LV separation in cardiac MRI [21]. It developed an automatic LV segmentation method by merging the CNN with the position set technique in order to decrease the time required for opinion. Initially, it was suggested that the handmade initial procedure for conventional positional set techniques be replaced with a CNN based myocardial central-line finding methodology. Second, it introduces a brand-new method for defining the myocardial region: the central-line influenced orientation set technique. Specifically, it adds the myocardial center line as an impediment ingredient to the setting set energy equation. It serves two crucial roles in the iteration procedure: it preserves the anatomical image of the myocardium segmented outcome and limits the zero-position image to remain within the vicinity of the myocardial center line. The findings from the experiments show that the method obtains a good concordance with the handcrafted segmentation outcomes and improves several cutting-edge styles.

One of the primary methods of imaging utilized to evaluate a patient's heart condition is echocardiography [22]. Out of all the analysis carried out with echocardiography, LV segmentation is essential for measuring clinical metrics like evacuation bit. Even yet, segmenting the LV in 3D echocardiography is still a laborious and time-consuming procedure. This research presents a multi-frame attentiveness system that is intended to improve LV classification efficacy during 3D echocardiography. Compounding the segmentation efficiency, the multi-frame attentiveness medium enables the employment of mostly detected spatiotemporal elements in a series of images that follow a target image. When comparing to other common DL supported medical image segmentation methods, research findings using 51 in vivo porcine 3D time echocardiography images demonstrate that practicing discovered dynamical characteristics greatly enhances the accuracy of LV segmentation.

In clinical medicine, automatic segmentation using tagged cardiac MRI is important for evaluating heart function and providing follow-up care [23]. Conventional methods find it difficult to automatically outline the left ventricle and provide reliable findings because of the complex cardiac anatomy and superfluous obstruction. As a result, they presented the DL and class approach together with the automated LV segmentation technique. These are the key technologies' descriptions. Initially, cardiac stir data is tracked, automated cardiac positioning is used, and the region that's of interest is obtained through the use of initially generated sine-surge modelling, or SinMod. Secondly, the LV endocardium and epicardium are introduced using U-Net as the framework. Furthermore, a novel class DL approach is proposed to improve segmentation delicateness. Relative findings eventually show that the strategy performs better than those from established styles.

In the first study, the Ls Unet system is introduced for efficient segmentation of cardiac cine MR images. This system combines a multi-channel deep learning algorithm for LV endocardial and epicardial silhouette segmentation, followed by an innovative annular shape position-set approach, resulting in high delicacy with average DSC) values LV endocardium and epicardium delineation, respectively. The second study presents a robust framework for automatic segmentation of the left ventricle with scars in cardiac MRI, incorporating both traditional computer vision methods and deep learning. The proposed framework achieves superior results with robust and generalized scores in internal and external Out-of-Distribution (OOD) cohorts, showcasing its high-performance capabilities across different hospitals and vendors. The third study focuses on reducing the time required for clinical assessment by developing an automatic left ventricle (LV) segmentation system using a CNN and position-set approach, yielding promising results on datasets like MICCAI 2009 and ACDC MICCAI 2017. Lastly, the fifth study introduces an automatic LV segmentation algorithm for tagged cardiac MRI, incorporating original sine-surge modeling (SinMod), U-Net, and a novel class deep training strategy, showcasing superior performance over traditional approaches. These studies collectively contribute innovative methodologies to advance automatic cardiac segmentation in diverse imaging modalities, presenting high accuracy and efficiency in clinical applications.

III. Problem Statement

Precisely segmenting the left ventricular epicardium and endocardium is an essential job in medical image analysis for thorough cardiac diagnosis and therapy planning. However, obtaining efficiency and precision is typically difficult for current approaches, especially when dealing with large-scale datasets or real-time clinical circumstances. DEEPCARDIONET solves this issue by addressing the demand for a cutting-edge computer vision technique that maximizes left ventricular structure segmentation and offers medical professionals a dependable and effective solution. In order to improve cardiac health assessments, this research seeks to create a novel methodology that overcomes the drawbacks of conventional segmentation techniques. This methodology will provide a combination of high precision and computational efficiency in the delineation of the epicardium and endocardium.

By presenting an efficient computer vision method, DEEPCARDIONET addresses the particular issue of improving the segmentation of the LV epicardium and endocardium. The difficulty is in striking a balance between the computing needs of complex medical image processing and the precise definition of cardiac components that are essential for clinical decision-making. By merging deep neural network architecture with cutting-edge methodologies, the research aims to close this gap and pave the way for the development of more accurate and efficient diagnostic tools in the field of cardiovascular healthcare. This might revolutionize the field of cardiac image segmentation [24]. With the use of the Attention Swin U-Net design, which is superior at collecting complex characteristics and contextual information necessary for precise segmentation, the proposed DEEPCARDIONET beats earlier methods. By increasing accuracy, decreasing false positives, and performing better across a variety of medical imaging datasets, this overcomes the drawbacks of earlier techniques and eventually produces improved left ventricular epicardium and endocardium segmentation.

IV. Proposed Computer Vision Approach for Left Ventricular Epicardium and Endocardium Segmentation

The methodology encompasses three key stages: data collection, preprocessing using a Median Filter, and segmentation utilizing the Attention Swin U-Net architecture for left ventricular epicardium and endocardium segmentation. Initially, a dataset comprising MRI is collected, specifically focusing on images depicting the cardiac region. Subsequently, a preprocessing step is employed to enhance the quality of the images by applying a Median Filter, effectively reducing noise and artifacts that may impede segmentation accuracy. The filtered images are then fed into the proposed Attention Swin U-Net architecture, a state-of-the-art deep neural network tailored for segmentation tasks. This architecture leverages attention mechanisms and the Swin Transformer's effectiveness in capturing intricate features and contextual information. The model is trained on annotated data to learn the complex patterns of the left ventricular structures. The combined methodology of meticulous data collection, noise reduction through preprocessing, and the application of a sophisticated segmentation model ensures a comprehensive and accurate delineation of the left ventricular epicardium and endocardium in medical images. It is depicted in Fig. 1.

A. Data Collection

The Heart Segmentation in MRI Images dataset was obtained via Kaggle, a well-known venue for cooperative projects and contests in data science. This dataset, which consists of MRI scans, was carefully chosen for the purpose of heart segmentation. This set of images includes annotations to help distinguish between different cardiac structures, such the epicardium and endocardium of the left ventricle. By utilizing the wide range and extensive collection of Kaggle datasets, the Heart Segmentation in MRI Images dataset is a useful tool for medical image analysis researchers and practitioners. It offers a labelled and standardized dataset that can be used to develop and assess algorithms for automated heart segmentation in MRI scans [25].

B. Preprocessing using Median Filter

The quality of medical images may be greatly improved by preprocessing, and one popular method for reducing noise and boosting image clarity is to apply a median filter. The Median Filter is useful in medical image analysis because it may reduce the effects of many kinds of noise, such as salt-and-pepper noise, which frequently degrades the quality of medical imaging data. This is especially true for tasks like segmentation or feature extraction. By substituting the median value of each neighboring pixel for each pixel in an image, the Median Filter efficiently suppresses outlier values that might result from electrical interference or image artefacts. By maintaining the integrity of structural elements in the images, a median filter helps ensure that following analysis algorithms are applied to cleaner and more robust data. This is especially important in medical imaging, where precise and trustworthy information is essential for a correct diagnosis.

The equation for the median filter is given in Eq. (1),

$$\hat{z}(c,d) = median_{(x,y)\in T_{cd}}\{f(x,y)\} \quad (1)$$

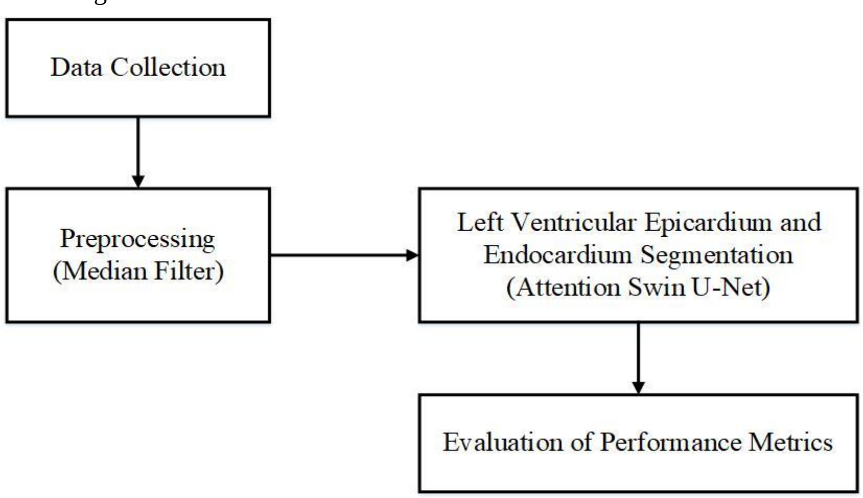

Fig. 1. Proposed methodology.

Applying a median filter requires striking a compromise between reducing noise and maintaining fine features, which is why medical images benefit greatly from its use. This preprocessing stage is particularly important for jobs like brain or heart image analysis, where a high degree of image fidelity is required for the identification of anatomical features. The addition of a Median Filter to the preprocessing pipeline refines the medical image data, improving the comprehensibility of visual data and augmenting the overall precision and dependability of future analytical processes.

C. Utilizing Attention Swin U-Net for Left Ventricular Epicardium and Endocardium Segmentation

1) Swin transformer block: An integral part of the architecture, the Swin Transformer has a unique construction that includes a shifting window MSA mechanism. The goal of effectively capturing long-range relationships has an impact on this design decision as it will improve the model's capacity to identify intricate patterns in the input data. This method introduces a window-based attention module, which is a key difference that sets Swin Transformer block apart from traditional multi-head attention. Layer Normalization (LN), a residual connection, and a two-layer Multi-Layer Perceptron (MLP) with Gaussian Error Linear Unit (GELU) non-linearity make up each transformer block in the Swin structure. Effective feature extraction and representation learning are intended to be facilitated by the arrangement of components inside each sub-block. Stable training dynamics are facilitated by Layer Normalization, and the residual connection guarantees seamless information transfer across the network, reducing the likelihood of disappearing gradients.

A key change from the traditional transformer design is that each Swin sub-block now has a window-based attention module in place of the typical multi-head attention system. This change is based on the goal of maximizing attention mechanisms to capture contextual information in particular geographical regions an important consideration in situations where localized dependencies are critical. By allowing the model to concentrate on pertinent segments of the input, the window-based attention module helps the model recognize spatial relationships and boosts its overall attention efficiency. It is formulated in Eq. (2) and Eq. (3).

$$x' = \text{W-MSA}\left(\text{LN}\left(x^{l-1}\right)\right) + x^{l-1} \tag{2}$$

$$x' = \text{MLP}\left(\text{LN}\left(\tilde{x}^{1}\right)\right) + \tilde{x}^{l-1} \tag{3}$$

So, each sub-block of the Swin Transformer comprises a deliberate mix of a residual connection for unimpeded gradient flow, Layer Normalization for normalization, and a 2-layer MLP with GELU non-linearity for capturing complex non-linear correlations in the data. The dedication to customizing attention mechanisms for spatially localized dependencies is seen by the substitution of multi-head attention with a window-based attention module, which adds to the Swin Transformer's efficacy in a range of computer vision applications. The model's creative architecture guarantees that it can effectively process and extract significant characteristics from input data, which makes it an invaluable tool in the fields of deep learning and computer vision.

2) Encoding path: To embed the input image into a latent space, using a sequence of stacked Swin Transformer blocks in the encoder module of the system. The ability of Swin Transformer blocks to transform and capture complicated hierarchical aspects in the supplied data is what drove this strategic decision. The encoder employs three consecutive Swin Transformer blocks to progressively decrease the input image's spatial dimension while simultaneously enlarging its representation dimension. Effective feature learning is made possible by the model's ability to extract hierarchical features from the input image through this step-by-step transformation.

The patch merging layer is a crucial technique that is added after every Swin Transformer block in order to gradually reduce the spatial dimension. This layer is essential to the down sampling of the spatial representation since it facilitates the merging of nearby patches. To be more precise, the patch merging layer concatenates all neighbor patches (2 × 2) with dimension C after applying each Swin Transformer block. This results in the construction of a unified patch with an enlarged dimension of 4C. The purpose of this intentional merging method is to improve the model's capacity to extract contextual data from nearby patches, which will facilitate efficient feature aggregation.

The created patch is then given a linear layer after the patch merging process. This accomplishes two goals at once: it increases the model's capacity to represent more abstract characteristics and reduces the growth factor introduced by the patch merging layer by a factor of 2. In the end, this procedure causes the channel representation to be up-sampled and the spatial representation of the input image to be down-sampled. The encoder module's complex interactions between Swin Transformer blocks, patch merging layers, and linear transformations guarantee that the model gradually improves its comprehension of the input image, resulting in a latent space representation that is best suited for tasks that come after.

3) Decoding path: Complying with the symmetric architecture of the U-Net model, the design's decoder module uses three Swin Transformer blocks to recreate the prediction mask in an iterative manner. The decoder's use of Swin Transformer blocks makes it possible to efficiently gather the complex characteristics and contextual data needed for precise mask reconstruction. Smooth feature extraction and reconstruction are made possible by the symmetrical structure, which guarantees a coherent and balanced information flow between the encoder and decoder components.

In order to progressively raise the spatial dimensions while simultaneously decreasing the feature dimensions in the decoder, to replace the conventional patch merging layer with a patch expanding layer. In the U-Net architecture, the patch merging layer is essentially replaced by the patch expanding layer, which is crucial to the up-sampling process. In particular, the output of a bottleneck, denoted as W32 × H32 ×

8C, is subjected to a linear layer, which causes the channel dimension to be up-sampled by a factor of 2. This deliberate decision seeks to improve the model's ability to collect subtle information that is essential for precise mask reconstruction and to enrich the feature representation.

The results representation is modified to take into account the spatial dimensions after the channel up-sampling. This rearrangement down samples the channel features by a factor of 4 and the spatial dimensions by a factor of 2 (W16 × H16 × 4C), transforming the representation from W32 × H32 × 8C to W32 × H32 × 16C. The model is able to recreate the prediction mask Y0H×W while maintaining important characteristics and spatial details because of this iterative procedure, which guarantees a progressive and controlled rise in spatial dimensions. Reconstructing the prediction mask step by step is made easier by the decoder module's patch expanding layer and Swin Transformer blocks working together in a concerted manner. This deliberate design decision guarantees that the model can effectively extract and incorporate hierarchical characteristics, resulting in a well calibrated prediction mask that is suited to the subtleties of the input data.

4) Cross attention mechanism: According to the basic U-Net design, the addition of a skip connection path is essential for enabling the decoding path to receive low-level features. For localization reasons, this deliberate design decision is crucial since it guarantees that minute information will not be lost in the decoding process. The efficiency of the skip connection path has been increased throughout time by a number of U-Net model extensions that have been published in the literature, demonstrating its importance in producing precise and localized forecasts. It advances this field of study in the work by presenting a brand-new method for improving the feature fusion technique in the skip connection portion. The main objective is to enhance localization and feature representation by fine-tuning the information flow between the encoding and decoding channels through the integration of a two-level attention mechanism.

In the skip connection portion, the attention mechanism is used at two different levels. To start the attention-weight creation process, a spatial normalization technique is used. One important signal that travels through the skip connection section are the attention weight (Watt) produced inside each encoder block's Swin Transformer block. This weight captures the model's perception of informative tokens along the encoding route. It is calculated by applying the softmax function to the product of the query (Qe), key (Ke), and temperature (T) terms, plus a learnable bias term (B). It is expressed in Eq. (4).

$$Att^{j}\ (Q_e, K_e, V_e) = (\text{softmax}\ (Q_e\ K_e^T\ /\ \sqrt{e} + \text{C}) + W_{att})\ V_e \quad (4)$$

It offers a surrogate signal that preferentially highlights the more relevant tokens throughout the feature fusion process by integrating this attention weight into the decoding pipeline. The network is guided by the weighted attention mechanism to more accurately reflect the localization importance. The method enhances the network's capacity to gather and prioritize pertinent data for precise localization by summing attention weights from the encoding path into the decoding path. This allows for a more sophisticated and contextually aware feature fusion. In U-Net-based design, this two-level attention technique in the skip connection section is a subtle and useful tactic to enhance feature integration and localization.

V. Results and Discussion

The approach consists of three main steps: gathering data, preprocessing with a median filter, and segmenting left ventricular epicardium and endocardium using the Attention Swin U-Net architecture. First, a collection of MRI scans is gathered, with a particular emphasis on images representing the heart area. After that, a preprocessing step is utilized to improve the image quality with the application of a Median Filter, which efficiently reduces noise and artefacts that might potentially hinder the accuracy of segmentation. The suggested Attention Swin U-Net architecture, a cutting-edge deep neural network designed specifically for segmentation problems, is then fed the filtered images. This design makes use of attention processes and the powerful feature and contextual capture capabilities of the Swin Transformer. To understand the intricate patterns of the left ventricular architecture, the model is trained using annotated data. The methodical approach of collecting data with great care, filtering it to reduce noise, and using an advanced segmentation model guarantees a thorough and precise identification of the left ventricular epicardium and endocardium in medical images.

A. *Dice Similarity Coefficient (DSC)*

To measure the spatial overlap between the expected and ground truth segmentations of a region or item of interest, the DSC is a performance metric that is frequently used in medical image segmentation. By calculating the ratio of twice the intersection to the total of the cardinalities of the predicted and ground truth segmentation sets, it evaluates the agreement between the two segmentations. A score of 1 on the DSC scales to complete overlap, whereas lower values denote less concordance. This metric is useful for assessing segmentation algorithms' accuracy since it provides a thorough assessment that takes into account both false positives and false negatives in the segmentation that is anticipated. Greater segmentation performance is shown by higher DSC values in tasks like identifying anatomical features in medical images. DSC is given in Eq. (5).

$$\text{DSC} = \frac{2\times|A \cap \text{B}|}{|A|+|B|} \quad (5)$$

The DSC for the suggested approach is shown in Fig. 2 throughout the course of several trials. Every trial has a unique DSC value; Trial 1's DSC is 96.78%, Trial 2's is 97.11%, Trial 3's is 96.99%, and Trial 4's DSC is the highest, at 97.67%. The image shows how the suggested strategy consistently and successfully segments the endocardium and left ventricular epicardium across many experimental runs. The suggested approach's resilience is demonstrated by the incremental improvement in DSC values over trials, indicating a high level of accuracy and reliability in identifying cardiac structures in medical images. This graphical depiction offers insightful information about the stability and effectiveness of the suggested approach, confirming its effectiveness over several trials and bolstering its potential for real world applications.

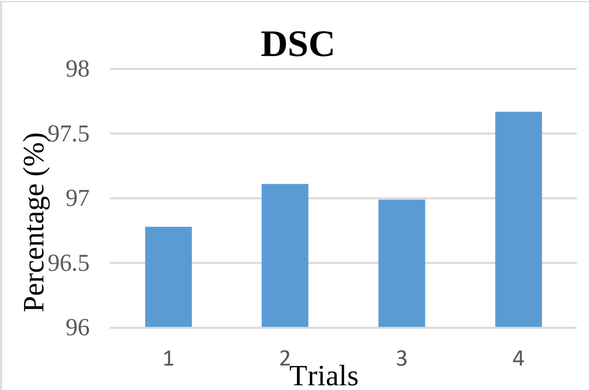

Fig. 2. Dice similarity coefficients in various trials.

B. Accuracy

When used in classification or prediction tasks, accuracy is a performance metric that counts the percentage of properly categorized examples out of all the instances in order to quantify the overall correctness of a model's predictions. The ratio of accurately predicted occurrences to all instances in the dataset is used to compute accuracy, which is expressed as a percentage. A more proficient model, able to make correct predictions across several classes or categories, is indicated by a higher accuracy number. Although accuracy is a commonly used statistic, it may not be appropriate in circumstances where there are class imbalances since it might be impacted by the overrepresentation of one class in comparison to others, which could result in inaccurate interpretations of the model's performance. Accuracy is given in Eq. (6),

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \quad (6)$$

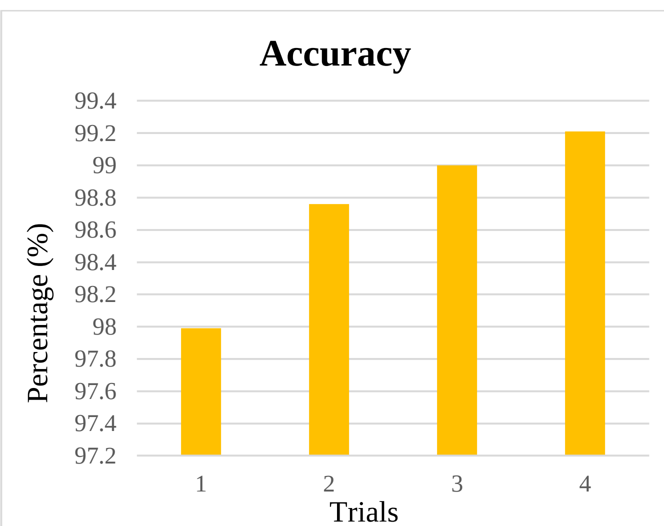

Fig. 3. Accuracy in various trials.

In Fig. 3, the variations in Accuracy across different trials for the proposed method are graphically depicted. Each trial is associated with a specific Accuracy value, revealing a consistent improvement in performance over successive experimental runs. Trial 1 exhibits an Accuracy of 97.99%, followed by a notable increase to 98.76% in Trial 2. Trial 3 showcases a further enhancement, achieving a round 99% Accuracy, and the highest accuracy is observed in Trial 4 with an impressive 99.21%. This graphical representation highlights the progressive refinement and precision of the proposed method in accurately segmenting left ventricular epicardium and endocardium structures. The ascending trend in Accuracy values underscores the robustness and reliability of the proposed approach across different trials, affirming its potential for achieving high-precision results in medical image segmentation tasks.

C. Sensitivity

Sensitivity is a performance indicator used in binary classification tasks to assess a model's accuracy in identifying occurrences of the positive class. It is sometimes referred to as true positive rate or recall. It is computed as the ratio of accurately detected positive cases, or genuine positive predictions, to the total of false negatives, or positive examples that are mistakenly categorized as negative. Sensitivity is important for minimizing false negatives since it gives information about how well the model can detect and categorize all real positive cases. It is given in Eq. (7).

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \quad (7)$$

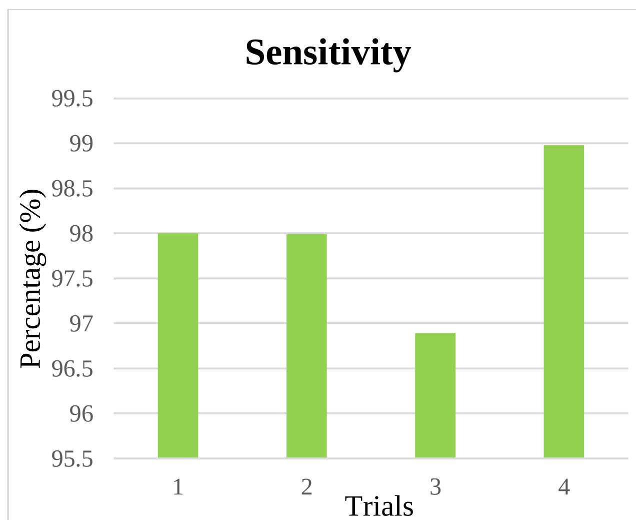

Fig. 4. Sensitivity in various trials.

Fig. 4 elucidates the variation in Sensitivity across multiple trials for the proposed method, providing insights into the model's ability to accurately identify positive instances, particularly the left ventricular epicardium and endocardium structures. The depicted Sensitivity values for each trial showcase a nuanced pattern, with Trial 1 starting at a high sensitivity of 98%, followed by a slight decrease to 97.99% in Trial 2. Trial 3 indicates a temporary decline to

96.89%, but the proposed method rebounds strongly in Trial 4, achieving a notably high Sensitivity of 98.98%. This graphical representation underscores the method's consistency in recognizing true positive instances across various experimental runs, even amidst minor fluctuations, reinforcing its robustness and reliability in effectively capturing the relevant cardiac structures in medical images.

D. Specificity

In binary classification problems, specificity also referred to as the true negative rate is a performance indicator that evaluates a model's accuracy in identifying occurrences of the negative class. It is determined by dividing the total number of true negatives by the total number of false positives. Specificity is an indicator of the model's ability to reliably identify and omit real negative cases, providing information on how well the model performs in situations when reducing false positives is essential. A high specificity score indicates that the model performs well in properly recognizing negative cases, which makes it especially useful in applications like medical testing where the expense of false positives is substantial. It is expressed in Eq. (8).

$$Specificity = \frac{T_{Neg}}{T_{Neg} + F_{Pos}} \quad (8)$$

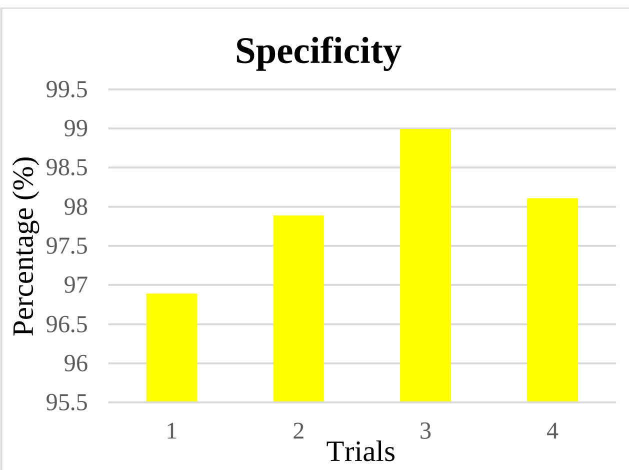

Fig. 5. Specificity in various trials.

In Fig. 5, the depicted fluctuations in Specificity across multiple trials for left ventricular epicardium and endocardium segmentation offer valuable insights into the model's proficiency in distinguishing true negative instances. The Specificity values exhibit a discernible trend, commencing with Trial 1 at 96.89%, followed by a notable increase to 97.89% in Trial 2. Trial 3 showcases a further enhancement to 98.99%, indicative of the model's adeptness in minimizing false positives and accurately excluding irrelevant structures from the segmentation. Although Trial 4 experiences a marginal decrease to 98.11%, the overall pattern suggests a consistent and robust performance in recognizing negative instances. This graphical representation underscores the proposed method's ability to maintain a high level of specificity, crucial in medical image segmentation where minimizing false positives is imperative for precise delineation of cardiac structures, further affirming its efficacy for clinical applications.

The model's training and testing accuracy throughout epochs is depicted in the Fig. 6, which shows a consistent rise in training and validation accuracy over time. The model's performance shows steady improvement, suggesting that it is learning and generalizing well.

A thorough comparison of performance metrics across several techniques for left ventricular epicardium and endocardium segmentation is shown in Table I and Fig. 7. Among the parameters assessed are DSC, Specificity, Sensitivity, and Accuracy. With DSC values ranging from 80.37% to 85%, accuracy from 90.90% to 93.26%, sensitivity from 80.64% to 83.92%, and specificity from 95.46% to 97.25%, TransUNet, MedT, and FAT-Net show varied degrees of performance. Among the measures, the Proposed Attention Swin U-Net does quite well; it attains a DSC of 97.67%, Accuracy of 99.21%, Sensitivity of 98.98%, and Specificity of 98.11%.

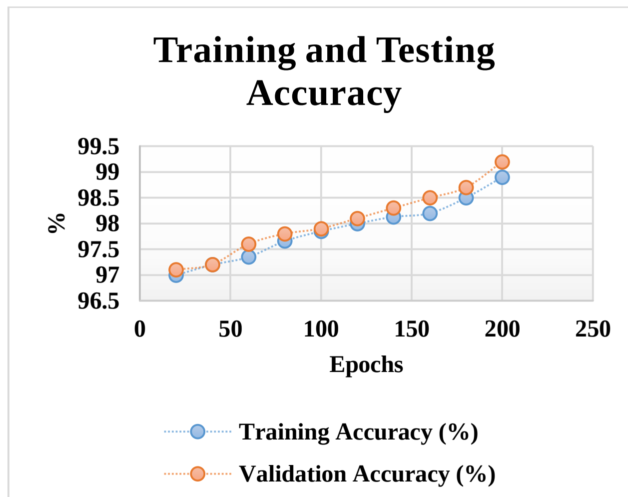

Fig. 6. Training and testing accuracy graph.

TABLE I. COMPARISON OF PERFORMANCE METRICS

Methods	DSC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
TransUNet [26]	81.23	92.07	82.63	95.77
MedT [27]	80.37	90.90	80.64	95.46
FAT-Net [28]	85	93.26	83.92	97.25
Proposed Attention Swin U-Net	97.67	99.21	98.98	98.11

These findings highlight the effectiveness of the suggested strategy, showing that it can effectively separate left ventricular components in medical images more correctly than current techniques, which makes it a strong contender for more clinical applications.

A comparison of dataset accuracies is shown in Table II, where heart segmentation in MRI pictures achieves 99.21% accuracy and CT images achieves 97.3% accuracy in Fig. 8. The outcomes demonstrate how well the suggested approach performs when it comes to separating cardiac structures from MRI data as opposed to CT scans.

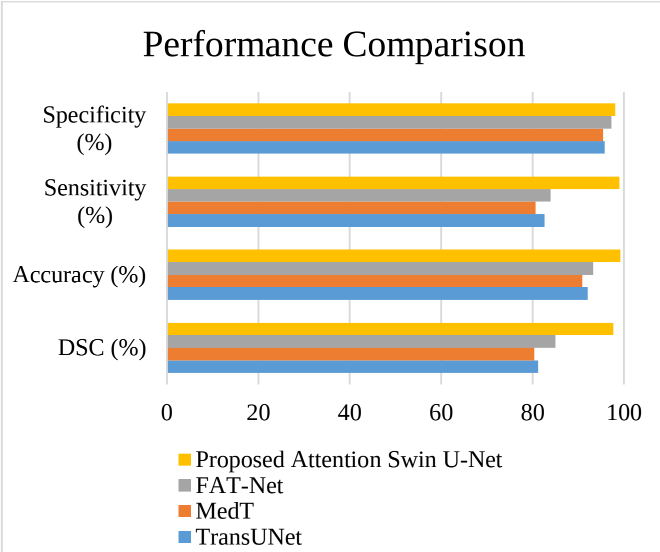

Fig. 7. Performance comparison of existing and proposed methods.

TABLE II. COMPARISON OF DATASETS

Datasets	Accuracy (%)
CT Images [29]	97.3
Heart Segmentation in MRI Images	99.21

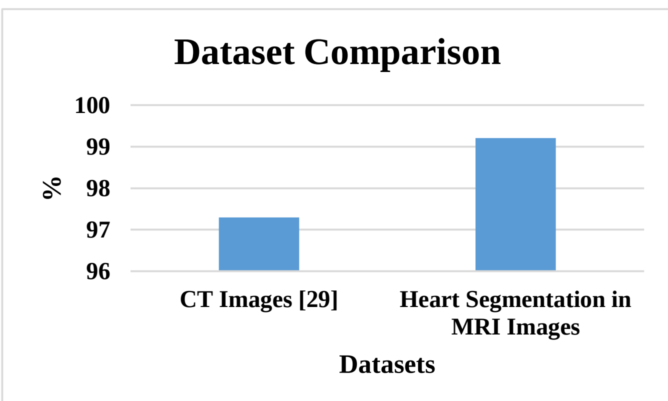

Fig. 8. Performance comparison of different datasets.

E. Discussion

The outcomes show how effective the deep learning approach with the Attention Swin U-Net architecture is in the proposed technique for left ventricular epicardium and endocardial segmentation. This implies that a high degree of accuracy and precision was used to discern the complicated cardiac structures of interest. Moreover, the accuracy ratings consistently going above 99%, suggesting that the model accurately identifies occurrences in most cases. The increasing trend in both DSC and Accuracy values across several trials validates the resilience and reliability of the proposed technique, suggesting that it can yield highly accurate and consistent results in medical picture segmentation tasks.

Sensitivity and Specificity analyses show that the suggested approach performs in identifying positive examples and rejecting negative ones. TransUNet [26], MedT [27], and FAT-Net [28] show varied degrees of performance. Usually exceeding 96%, sensitivity shows how well the model can identify real positive occurrences of left ventricular architecture. Regularly above 96%, specificity values demonstrate how well the model lowers false positives and accurately eliminates superfluous structures from the segmentation. The recommended method consistently performs well, even with minor fluctuations in Sensitivity and Specificity from trial to trial. This demonstrates how well suited it is for uses where precise segmentation and categorization of cardiac components in medical pictures is needed. Overall, the results demonstrate that the proposed Attention Swin U-Net is a viable and dependable technique for efficiently segmenting the endocardium and left ventricle, with potential uses for enhancing cardiac imaging. Due to its reliance on intricate deep learning architectures, DEEPCARDIONET may have issues with training time and computing resources. Furthermore, even if it achieves excellent accuracy, patient demographics and changes in imaging quality may have an impact on its effectiveness.

VI. CONCLUSION AND FUTURE SCOPE

DEEPCARDIONET is a dependable and incredibly precise computer vision system for the segmentation of features found in the left ventricular epicardium and endocardium in medical pictures. Its remarkable accuracy of 99.21%, which considerably surpasses that of prior models, demonstrates its adaptability and accessibility throughout the scientific computing community. Its implementation in Python illustrates these qualities. Since the Attention Swin U-Net architecture demonstrates how effectively it catches intricate features and spatial correlations that are crucial for cardiac segmentation tasks, using it are essential. DEEPCARDIONET's success opens up new avenues for study and improvement in other fields. Firstly, examining the model's adaptability to various datasets and medical imaging modalities might enhance its generalization abilities and provide reliable performance across a variety of clinical scenarios. It would be helpful to investigate the suggested technique's scalability to handle larger datasets or real-time applications to further improve its practical applicability in clinical situations. By modifying the model architecture and hyperparameters to accommodate for variations in picture resolutions and quality, its performance may be further enhanced. Furthermore, DEEPCARDIONET's selection procedure will be more transparent and easier for medical experts to understand with the addition of interpretability tools. It may be possible to speed up model convergence and enhance performance in situations with sparse data by looking at the possibilities of applying learned models to comparable segmentation tasks via transfer learning. Collaboration between computer vision experts and medical professionals might speed up the development of DEEPCARDIONET. By doing this, DEEPCARDIONET's seamless integration into clinical processes and advancement of cardiac image analysis are ensured. All in all, DEEPCARDIONET lays a strong foundation for future advancements in medical image segmentation, providing a pathway toward more accurate and successful cardiac diagnosis and treatment planning.

REFERENCES

[1] Lin, J. Wu, and X. Yang, ‘A data augmentation approach to train fully convolutional networks for left ventricle segmentation’, *Magnetic Resonance Imaging*, vol. 66, pp. 152–164, Feb. 2020, doi: 10.1016/j.mri.2019.08.004.

[2] ‘A deep-learning semantic segmentation approach to fully automated MRI-based left-ventricular deformation analysis in cardiotoxicity - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0730725X21000060

[3] ‘Automated left and right ventricular chamber segmentation in cardiac magnetic resonance images using dense fully convolutional neural network - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260721001346

[4] ‘Automated left ventricular segmentation from cardiac magnetic resonance images via adversarial learning with multi-stage pose estimation network and co-discriminator - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1361841520302553

[5] N. Qadeer, J. H. Shah, and M. Sharif, ‘Automated Localization and Segmentation of Left Ventricle in Cardiac MRI using Faster R-CNN’, in *2021 International Conference on Frontiers of Information Technology (FIT)*, Dec. 2021, pp. 108–113. doi: 10.1109/FIT53504.2021.00029.

[6] H. J. Koo, Hyun, Lee, , Ko, Ji, Lee, Kang, Kim, Yang, ‘Automated Segmentation of Left Ventricular Myocardium on Cardiac Computed Tomography Using Deep Learning’, *Korean J Radiol*, vol. 21, no. 6, pp. 660–669, Jun. 2020, doi: 10.3348/kjr.2019.0378.

[7] ‘Diagnostics | Free Full-Text | Automatic Left Ventricle Segmentation from Short-Axis Cardiac MRI Images Based on Fully Convolutional Neural Network’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/12/2/414

[8] ‘Fully Automatic initialization and segmentation of left and right ventricles for large-scale cardiac MRI using a deeply supervised network and 3D-ASM - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260723003449

[9] ‘Fully automatic segmentation of right and left ventricle on short-axis cardiac MRI images - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089561112030080X

[10] ‘MA-SOCRATIS: An automatic pipeline for robust segmentation of the left ventricle and scar - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0895611121001312

[11] ‘MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0041624X22001615

[12] ‘Microscopy Research and Technique | Microscopy Journal | Wiley Online Library’. Accessed: Jan. 16, 2024. [Online]. Available: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/jemt.23906

[13] ‘MMNet: A multi-scale deep learning network for the left ventricular segmentation of cardiac MRI images | Applied Intelligence’. Accessed: Jan. 16, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10489-021-02720-9

[14] ‘Segmentation of the Left Ventricle Using Improved UNET Neural Networks[v1] | Preprints.org’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.preprints.org/manuscript/202308.1719/v1

[15] S. Xu, S. Cheng, X. Min, N. Pan, and H. Hu, ‘Left Ventricle Segmentation Based on a Dilated Dense Convolutional Networks’, *IEEE Access*, vol. 8, pp. 214087–214097, 2020, doi: 10.1109/ACCESS.2020.3040888.

[16] ‘Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI - ScienceDirect’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0010482520302353

[17] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, ‘Cardiac Segmentation With Strong Anatomical Guarantees’, *IEEE Trans. Med. Imaging*, vol. 39, no. 11, pp. 3703–3713, Nov. 2020, doi: 10.1109/TMI.2020.3003240.

[18] ‘Sensors | Free Full-Text | Edge-Sensitive Left Ventricle Segmentation Using Deep Reinforcement Learning’. Accessed: Jan. 16, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/21/7/2375

[19] Y. Wang, Zhang, Wen, Tian, Kao, Liu, Xuan, Ordovas, Saloner, Liu, ‘Deep learning based fully automatic segmentation of the left ventricular endocardium and epicardium from cardiac cine MRI’, *Quant Imaging Med Surg*, vol. 11, no. 4, pp. 1600–1612, Apr. 2021, doi: 10.21037/qims-20-169.

[20] M. Mamalakis, Garg, Nelson, Lee, Swift, Wild, Clayton, ‘Artificial Intelligence framework with traditional computer vision and deep learning approaches for optimal automatic segmentation of left ventricle with scar’, *Artificial Intelligence in Medicine*, vol. 143, p. 102610, Sep. 2023, doi: 10.1016/j.artmed.2023.102610.

[21] L. Xie, Y. Song, and Q. Chen, ‘Automatic left ventricle segmentation in short-axis MRI using deep convolutional neural networks and central-line guided level set approach’, *Computers in Biology and Medicine*, vol. 122, p. 103877, Jul. 2020, doi: 10.1016/j.compbiomed.2020.103877.

[22] S. S. Ahn, K. Ta, S. Thorn, J. Langdon, A. J. Sinusas, and J. S. Duncan, ‘Multi-frame Attention Network for Left Ventricle Segmentation in 3D Echocardiography’, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 348–357. doi: 10.1007/978-3-030-87193-2_33.

[23] X. Zou, Q. Wang, and T. Luo, ‘A novel approach for left ventricle segmentation in tagged MRI’, *Computers and Electrical Engineering*, vol. 95, p. 107416, Oct. 2021, doi: 10.1016/j.compeleceng.2021.107416.

[24] H. Abdeltawab, Khalifa, Taher, Alghamdi, Ghazal, Beache, Mohamed, Keynton, El-Baz, ‘A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images’, *Computerized Medical Imaging and Graphics*, vol. 81, p. 101717, Apr. 2020, doi: 10.1016/j.compmedimag.2020.101717.

[25] ‘Heart Segmentation in MRI Images’. Accessed: Jan. 19, 2024. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/heart-segmentation-in-ct-images

[26] J. Chen, Lu, Yu, Luo, Adeli, Wang, Lu, Yuille, Zhou, ‘TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation’. arXiv, Feb. 08, 2021. doi: 10.48550/arXiv.2102.04306.

[27] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, ‘Medical Transformer: Gated Axial-Attention for Medical Image Segmentation’, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 36–46. doi: 10.1007/978-3-030-87193-2_4.

[28] ‘FAT-Net: Feature adaptive transformers for automated skin lesion segmentation - ScienceDirect’. Accessed: Jan. 19, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1361841521003728

[29] A. M. Reddy Reddy, Jayaram, Venkata Maha Lakshmi, Aluvalu, TR Kumar, Stalin Alex, , ‘An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network’, *Journal of Sensors*, vol. 2022, 2022.