International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 7, pp 82-88, July-2024

https://doi.org/10.47001/IRJIET/2024.807008

Advanced Left Ventricle Segmentation from Cardiac MRI Using U-Net with MobileNetV3 Encoder

[1]Rafeef Khalid Hasan, [2]Alaa Ghaith

[1,2]Department of Computer and Communications, Faculty of Engineering, Islamic University of Lebanon, Wardanieh, Lebanon

Abstract **- Assessing cardiac function and diagnosing different heart illnesses rely on accurate left ventricle (LV) identification using cardiac magnetic resonance imaging (MRI). To efficiently and accurately segment the left ventricle from 2D cardiac MRI data, this study introduces a novel method that combines a U-Net model with a MobileNetV3 encoder. The ACDC dataset, which includes MRI images and associated ground truth masks, underwent rigorous preprocessing and hyperparameters were adjusted to improve model performance. The evaluation resulted in an average dice score of 92.13%, with the LV segment receiving a dice score of 96.16%, displaying greater performance compared to previous studies. The combination of MobileNetV3 and U-Net has been proven to be effective for medical image segmentation, thereby enhancing diagnostic procedures and ultimately improving patient outcomes.**

Keywords: Left Ventricle, Segmentation, Magnetic Resonance Imaging, U-Net, MobileNetV3, ACDC, Dice Score.

I. INTRODUCTION

Cardiovascular diseases continue to be a significant worldwide health issue, responsible for a considerable number of fatalities each year [1]. Precise identification and prompt action are crucial to minimize the consequences of these illnesses. Cardiac magnetic resonance imaging (MRI) is crucial for assessing heart function and identifying different cardiac diseases [2]. An essential aspect of this evaluation is the accurate demarcation of the left ventricle (LV), which acts as a crucial sign of cardiac well-being.

The manual segmentation of the LV from cardiac MRI images is typically carried out by skilled professionals [3]. However, this process is demanding in terms of effort, time, and is prone to differences in interpretation among different observers. There is an increasing clinical demand for robust and effective tools that can simplify the imaging analysis process, enhance diagnostic accuracy, facilitate early detection of cardiac abnormalities, support clinicians in making informed treatment decisions and reduce the workload of radiologists. The introduction of automated segmentation techniques offers a promising solution to these issues [4], providing the potential for consistent and reproducible segmentation of the LV across various datasets and clinical environments.

Through the automation of complicated tasks like organ delineation and tumor detection, deep learning has completely transformed the efficiency and accuracy of medical picture segmentation, leading to a revolution in diagnostics [5]. Utilizing convolutional neural networks (CNNs), specifically designs like as U-Net and its variations, allows for the extraction of exact features and the acquisition of spatial context awareness, both of which are essential for correct segmentation. The capacity to learn hierarchical representations makes these models well-suited to medical imaging. Nevertheless, there are certain challenges that hinder the implementation of deep learning-based image segmentation techniques. These include the need for extensive and varied datasets, the requirement for precise identification without overfitting, and the considerable computational resources and time needed to train deep learning models. These limitations restrict scalability and hinder the practical application of these methods in clinical settings. In this study, we will develop a robust deep learning framework for automated medical image segmentation, with a particular focus on cardiac MRI, with the aim of enhancing diagnostic accuracy and supporting more informed clinical decisions in cardiovascular medicine.

II. RELATED WORK

Efficient segmentation of medical pictures for evaluating cardiac function and diagnosing heart illnesses has been accomplished using traditional approaches, machine learning methods, and more recently, deep learning techniques. Fully Convolutional Networks (FCNs), which were introduced by Long et al., enabled the training of pixel-by-pixel categorization in a seamless manner. By directly learning hierarchical characteristics from the data, this strategy greatly enhanced the accuracy of segmentation. FCNs can predict segmentation tasks at the pixel level because they preserve spatial records across the network. Creating these drawings by hand is tedious and can result in large, difficult designs. To address this problem, the AdaEn-Net architecture was proposed in [6] which is a self-adaptive FCN ensemble for 3D clinical image segmentation. AdaEn-Net's 2D FCN pulls

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 7, pp 82-88, July-2024

https://doi.org/10.47001/IRJIET/2024.807008

information from a single slice, while 3D FCN exploits information between multiple slices. The structure and hyperparameters of 2D and 3D structures are determined using a multi-objective evolutionary approach to improve segmentation accuracy while minimizing mesh parameters.

The U-Net architecture, introduced by Ronberger et al., has emerged as the fundamental framework for medical picture segmentation. A new approach was presented in [7] that involves integrating a CNN and U-Net model for LV segmentation where CNN was used to identify the specific region of interest (ROI) and U-Net to perform the segmentation process. In [8], a deep learning-based segmentation method called nnU-Net was developed. Basic design choices are modeled as parameters, rules, and empirical decisions. Using 23 publicly available datasets from global biomedical retail competitions, nnU-Net outperforms highly specialized techniques without human interaction.

Although transformer architecture is common in natural language processing, its use in computer vision is limited. Attention is used in vision to supplement or replace convolutional networks while preserving structure. In [9] a pure transformer was used directly on image patch sequences that can improve the image classification performance. Vision Transformer (ViT) outperforms state-of-the-art convolutional networks while using less CPU resources when pre-trained on large datasets and applied to medium or small image recognition benchmarks such as ImageNet, CIFAR-100, VTAB, etc. In [10] TransUNet is presented which combines Transformers and U-Net. Starting from a CNN feature map, TransUNet encodes feature image patches using transformers. This technique captures global situations well. The 3D transformer nnFormer was used for volumetric image segmentation in [11]. Convolution helps build hierarchical object concepts and encode fine-grained spatial information. Large receptive fields and hierarchical evolutionary advantage are the main targets of local and global scale-based self-interest.

Swin-UNet was presented in [12] which is similar in architecture to UNet and is intended for medical image segmentation. Using feature image patches, a transformer-based Encoder-Decoder architecture with skip connections learned global-local semantic features. A hierarchical encoder with variable windows extracts context information and symmetrical decoders with corrective expansion layers to restore spatial resolution. LeViT-UNet was introduced in [13] to improve the efficiency and accuracy of medical image segmentation. This architecture uses multistage converter blocks in the LeViT encoder to investigate the benefits of mixing local and global features.

III. METHODOLOGY

Accurate and effective segmentation of the left ventricle (LV) in cardiac MRI images is urgently required for the diagnosis and management of cardiovascular disorders. The U-Net architecture, which capitalizes on the capabilities of deep learning, will be utilized in conjunction with the MobileNetV3 encoder that has been pre-trained on the ImageNet dataset. This approach will ensure correct results by employing carefully selected hyperparameters that strike a balance between accuracy and computational efficiency. Figure 1 shows the proposed methodology, which includes reading the dataset files to extract images and masks and then performing pre-processing. The U-NET model is then initialized using the MobileNetv3 encoder. Finally, the model is trained on the training data and then the model is evaluated on the test data.

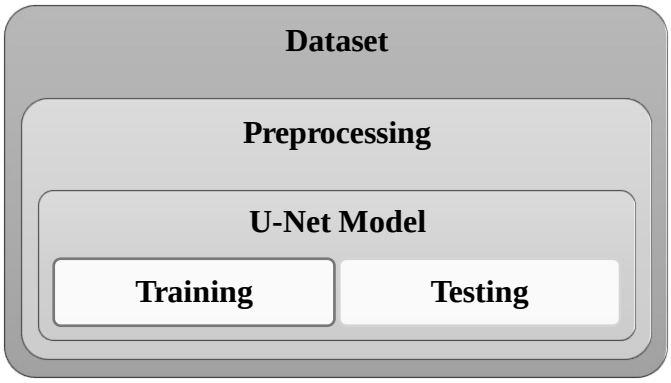

Figure 1: Proposed methodology

3.1 Dataset

The ACDC dataset was tested, which contains sufficient examples to demonstrate key physiological parameters obtained from cardiac MRI such as diastolic volume and ejection fraction. Patients whose clinical etiology was unclear were removed, and data for 150 patients were used in nifti format, which is the 4D image representation format for cardiac MRI images, as well as nifti files for the ground truth images (masks) corresponding to each MRI image. Figure 2 displays a collection of MRI pictures, which are grayscale representations of the patient's heart, together with associated masks. These masks are grayscale images that depict the segmentation of the MRI image into four distinct labels. The label values range from 0 to 3 and correspond to different regions within the image. A value of 0 represents the backdrop, a value of 1 represents the cavity of the right ventricle (RV), a value of 2 represents the myocardium (MYO), and a value of 3 represents the left ventricle (LV).

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 8, Issue 7, pp 82-88, July-2024
https://doi.org/10.47001/IRJIET/2024.807008

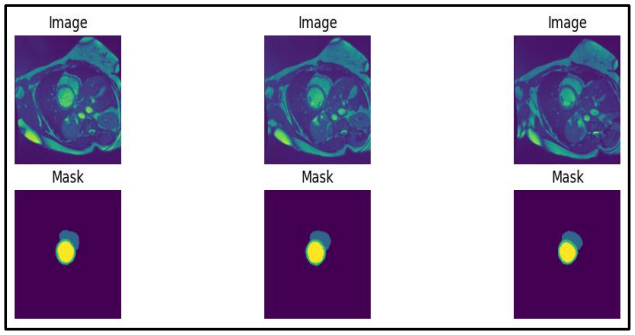

Figure 2: Some sample images with corresponding masks

Upon analyzing images of masks, it is observed that certain photos solely consist of the label 0, representing the backdrop. Consequently, MRI images and mask images encountering this issue are disregarded. Upon completion of the reading process for all MRI and mask images, a total of 2402 MRI images and their corresponding 2402 mask images were obtained.

3.2 Preprocessing

The photos have been optimized for the application of deep learning techniques. The size of the MRI pictures was standardized at (224, 224) in order to ensure uniformity and to optimize memory usage during training. Next, the photos underwent normalization, resulting in the pixel values being confined to a unified range. The masks also were resized to (224, 224) to standardize their size for processing. However, normalization was not applied to the mask images as they serve as labels for each segmentation. The mask images underwent a conversion process to become 3-channel color (RGB) images. In this conversion, label 1 was assigned to the first channel (R), label 2 to the second channel (G), and label 3 to the third channel (B). During this procedure, each channel exhibits a distribution of its respective label, including label 0, which represents the background. Figure 3 depicts the mask processing, with the upper portion illustrating the masks prior to processing and the lower portion illustrating the masks after processing. This procedure is crucial as it enables efficient training and enhances the visibility of each label.

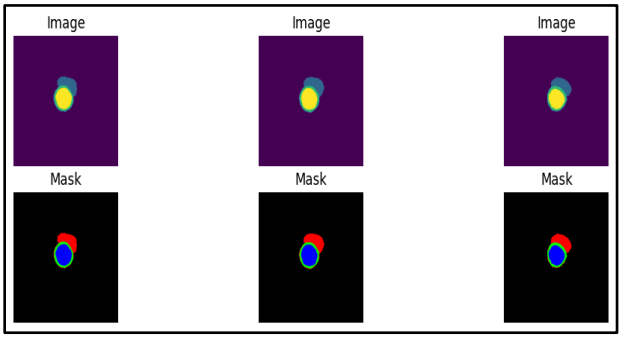

Figure 3: Masks before and after preprocessing

The dataset was partitioned into two subsets: a training set including 80% of the data, and a testing set comprising the remaining 20%. The training set has 1921 pictures and 1921 associated mask images for each MRI. The test set comprises 481 pictures and 481 associated mask images for each MRI.

3.3 U-Net Model

U-Net architecture consists of two distinct pathways, each serving a specialized purpose, as depicted in Figure 4. The initial pathway, sometimes referred to as the encoder, exhibits a structure that closely resembles that of a conventional convolutional neural network. This pathway plays a crucial role in extracting significant attributes from the image and providing the segmentation process with essential classification information. Conversely, the second pathway, commonly known as the decoder pathway or the synthesis pathway, utilizes a sequence of convolutions and concatenations to progressively enhance the precision of the output. The network achieves a thorough comprehension of the visual environment by integrating elements from the initial path into the expansion path, hence enhancing the precision of segmentation at a detailed level. The presence of this symmetry guarantees a seamless transmission of data, allowing the network to identify intricate patterns and nuanced characteristics in the input image.

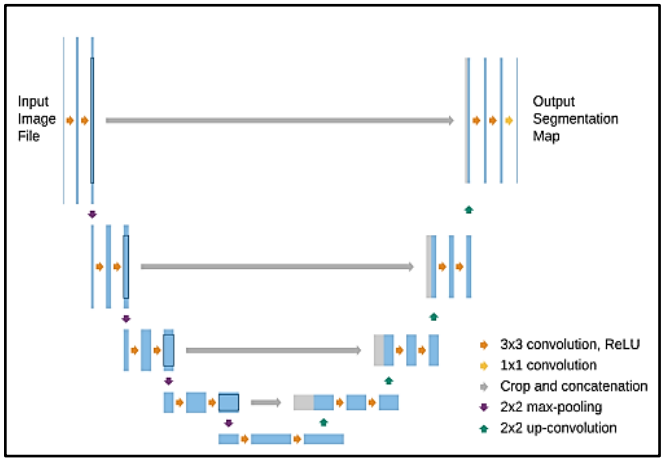

Figure 4: U-Net architecture

The MobileNet V3 architecture, a compact model well-suited for segmentation and classification tasks, was employed. Figure 5 illustrates the architecture of Mobilenet V3, in which the costly layers of the previous version have been modified in the remaining inverted design of MobileNet V2. The 1x1 expansion layer has been relocated after the pooling layer, leading to a decrease in latency and processing time. In the case of 3×3 convolution, the number of filters in the initial stage is decreased to 16, which is lower than the default value of 32 in other MobileNet models [14].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 7, pp 82-88, July-2024

https://doi.org/10.47001/IRJIET/2024.807008

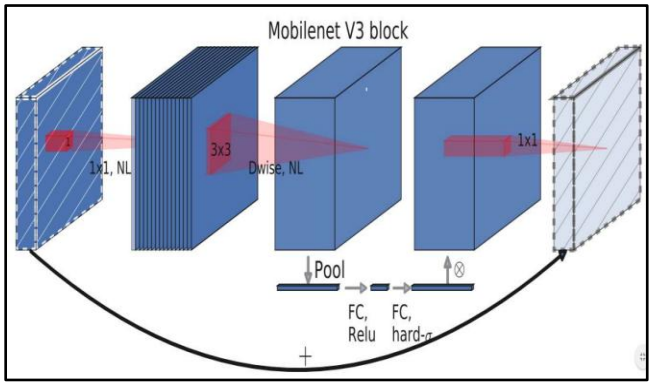

Figure 5: MobileNet V3 [14]

An encoder based on the MobileNet V3 technique, which was pre-trained on the ImageNet dataset, was used in the U-Net model. A dropout rate of 0.2 was implemented to mitigate the risk of overfitting. The input image is configured with a single channel, indicating that it is a grayscale image. The output image, on the other hand, is configured with three channels, indicating that it is a color image. These channels reflect the distinct colour components of the input image's special mask. A sigmoid activation function was employed to apply it to the output layer in order to ascertain the optimal pixel value for the channel. Table 1 displays the specific U-Net network settings that were used. The U-NET model has a total of 3,714,518 training parameters.

Table 1: U-Net parameters

Parameter	Value
U-Net Encoder	MobileNet V3 small
Input channel number	1
Output channel number	3
Activation function	Sigmoid
Dropout value	0.2
Training parameters	3714518 parameters

3.4 Training

Various techniques were employed to achieve accurate training of the U-Net model. Both the training and testing data were processed using a batch size of 8. The Adam optimizer was employed with an initial learning rate of 0.00098. The number of epochs was set to 100, indicating that the training will be repeated a maximum of 100 times. The model is trained using the training data, while the test data is employed as validation data throughout each epoch. To ensure accurate training and prevent overfitting, an early stopping technique was implemented. This involved monitoring the loss function for the training set after each epoch. If the loss did not improve by at least 0.005 over a span of 10 epochs, the training process was terminated.

In order to evaluate the U-Net model effectively, a dice score was calculated for each class separately which accounts for the overlap of two groups. The closer the value is to 1, the interference is excellent, and if it is close to 0, the interference is not excellent. Dice scores were calculated for each label using the equation 1:

$$DiceScore = \frac{2 * |MS_i \cap GT_i|}{|MS_i| + |GT_i|} \qquad Equation \quad 1$$

Where:

- i: Number of the channel.
- MS: True mask volume.
- GT: Predicted mask volume.

Additionally, a function was employed to store the optimal weights for the model. If the Dice Score value of the validation data for a specific epoch is higher than the Dice Score of the validation data for the previous epoch, the weights of the model are stored. The learning rate value was changed every 10 epochs by multiplying it by 0.9. Figure 6 shows the loss function curve for the training and validation data. These curves show that the model has been trained correctly and has learned the pattern of MRI images to produce the correct masks.

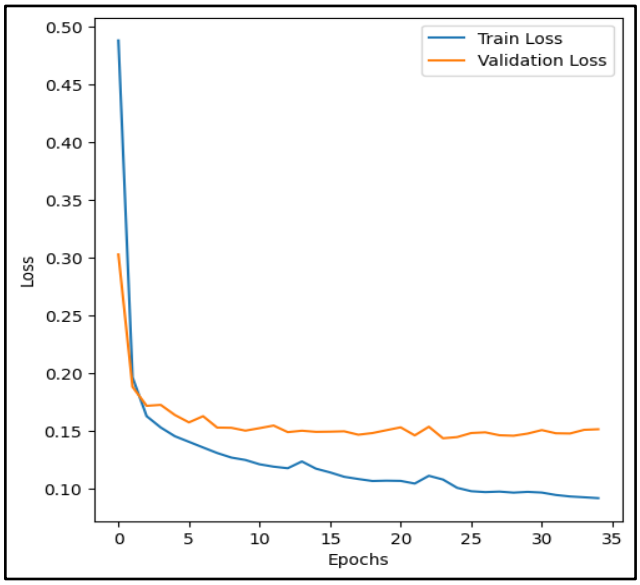

Figure 6: Loss curves for training and validation data

Training stops at epoch 35 due to the early stopping method that was applied. Figure 7 shows the dice score curve of the training and validation data.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 8, Issue 7, pp 82-88, July-2024
https://doi.org/10.47001/IRJIET/2024.807008

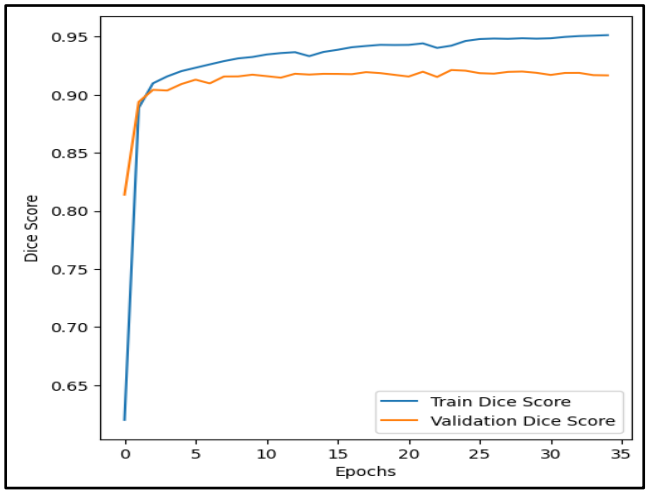

Figure 7: Dice Score curves for training and validation data

3.5 Testing

After completing training for the U-Net model, the best saved weight for the model was evaluated using the test data, where 4 values were calculated: the Dice Score value for the first label (RV), the Dice Score value for the second label (MYO), the Dice Score value for the third label (LV), and finally, the Dice Score value for all labels in order to display all the values for each label and also the average of all labels. The dice scores for RV, MYO, and LV were 90.88%, 89.34%, and 96.16% respectively. The average dice score for all labels was 92.13%. These findings clearly indicate that the model is highly effective in segmenting the LV label, since it achieves the highest dice score compared to the other labels.

IV. RESULTS

To effectively monitor the training process, a subset of the validation dataset was printed at regular intervals. Figure 8 displays the training samples at several epochs, specifically at epochs 1, 10, 20, 30, and 35. The efficacy of MobileNet V3 in feature extraction is evident from the fact that a significant amount of the image was detected in the initial epoch. After 10 epochs, the figure demonstrates a notable improvement in the learning of the U-Net model, reaching a level of similarity with the real mask. The training process was halted at epoch 35, and the model's weights were saved at epoch 24 as the optimal weights for the model.

Table 2 presents a thorough comparison between the suggested model and other relevant studies. The comparison is conducted by using the mean dice score and dice scores for the regions of RV, MYO, and LV. The comparison results demonstrate the efficacy of the proposed U-Net model with MobileNetV3 encoder in segmenting cardiac MRI images.

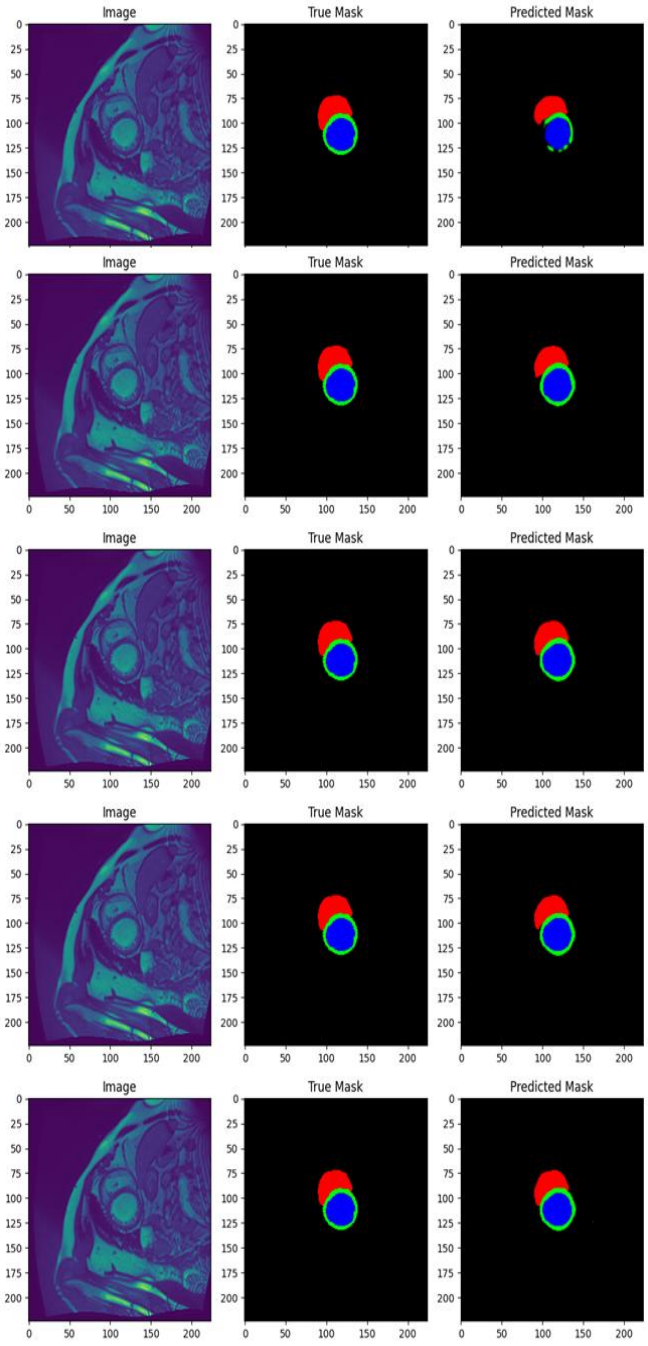

Figure 8: Sample during training from epochs 1, 10, 20, 30, and 35

MobileNetV3 is optimized for computational efficiency, enabling the model to perform well even with limited computing power. Moreover, the conversion of the mask into three-channel RGB images facilitated the segmentation process and improved the model's capacity to distinguish between distinct anatomical components. Utilizing enhanced hyperparameters and implementing early stopping resulted in enhanced model performance. The methodology's resilience and efficiency make it highly suitable for integration into clinical workflow, hence improving the speed and accuracy of cardiac MRI analysis.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 7, pp 82-88, July-2024

https://doi.org/10.47001/IRJIET/2024.807008

Table 2: Comparing proposed method results with related work

Model	Avg Dice Score	RV Dice Score	MYO Dice Score	LV Dice Score
nnUNet [8]	91.61 %	90.24 %	89.24 %	95.36 %
ViT [9]	81.45 %	81.46 %	70.71 %	92.18 %
R50 ViT [9]	87.57 %	86.07 %	81.88 %	94.75 %
TransUNet [10]	89.71 %	88.86 %	84.53 %	95.73 %
nnFormer [11]	91.78 %	90.22 %	89.53 %	95.59 %
Swin UNet [12]	90.00 %	88.55 %	85.62 %	95.83 %
LeVit-UNet384 [13]	90.32 %	89.55 %	87.64 %	93.76 %
Proposed Model	92.13 %	90.88 %	89.34 %	96.16 %

V. CONCLUSION

Accurate determination of cardiac parameters and diagnosis of various heart illnesses rely on correct left ventricle (LV) segmentation from cardiac MRI data. There is a lot of room for error and inefficiency in manual segmentation due to the high degree of subjectivity involved. The study works to improve upon existing techniques for LV segmentation using 2D cardiac MRI data by using a U-Net architecture with a MobileNetV3 encoder. Utilized was the ACDC dataset, which included 4D nifti files of cardiac MRI images together with their respective ground truth segmentation masks. In order to prepare the MRI pictures for training, the preprocessing steps included reducing their size, standardizing their pixel values, and transforming the mask images to three-channel RGB format. The pre-trained MobileNetV3 encoder was used to initialize the U-Net model since it is computationally efficient and has good feature extraction. The model achieved an average score of 92.13% Dice Score. These results show that the model can accurately segment left ventricles, providing a tool to help doctors measure important physiological parameters and enhance cardiovascular health assessment and treatment.

REFERENCES

[1] O. Gaidai, Y. Cao, and S. Loginov, "Global cardiovascular diseases death rate prediction," *Current Problems in Cardiology*, vol. 48, no. 5, 2023.

[2] D. Zhao, G. M. Quill, K. Gilbert, V. Y. Wang, H. C. Houle, M. E. Legget, and M. P. Nash, M. "Systematic comparison of left ventricular geometry between 3D-echocardiography and cardiac magnetic resonance imaging," *Frontiers in Cardiovascular Medicine*, vol. 8, 2021.

[3] M. M. Hadhoud, M. I. Eladawy, A. Farag, F. M. Montevecchi, and U. Morbiducci, "Left ventricle segmentation in cardiac MRI images," *American Journal of Biomedical Engineering*, vol. 2, no. 3, pp. 131-135, 2012.

[4] J. Sander, B. D. de Vos, and I.Išgum, "Automatic segmentation with detection of local segmentation failures in cardiac MRI," *Scientific Reports*, vol. 10, no. 1, p. 21769, 2020.

[5] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523-3542, 2021.

[6] M. B. Calisto, and S. K. Lai-Yuen, "AdaEn-Net: An ensemble of adaptive 2D–3D Fully Convolutional Networks for medical image segmentation," *Neural Networks*, vol. 126, pp. 76-94, 2020.

[7] B. Wu, Y. Fang, and X. Lai, "Left ventricle automatic segmentation in cardiac MRI using a combined CNN and U-net approach," *Computerized Medical Imaging and Graphics*, vol. 82, 2020.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203-211, 2020.

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[10] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[11] H. Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, "nnformer: Interleaved transformer for volumetric segmentation," *arXiv preprint arXiv:2109.03201*, 2021.

[12] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," *arXiv e-prints, arXiv-2105*, 2021.

[13] G. Xu, X. Wu, X. Zhang, and X. He, "LeViT-UNet: Make Faster Encoders with Transformer for Medical

ISSN (online): 2581-3048
Volume 8, Issue 7, pp 82-88, July-2024
https://doi.org/10.47001/IRJIET/2024.807008

Image Segmentation," *arXiv preprint arXiv:2107.08623*, 2021.

[14] P. S. Kavyashree, and M. El-Sharkawy, "Compressed mobilenet v3: a light weight variant for resource-constrained platforms," *In 2021 IEEE 11th annual computing and communication. Workshop and conference (CCWC)*, pp. 0104-0107, 2021.

Citation of this Article:

Rafeef Khalid Hasan, & Alaa Ghaith. (2024). Advanced Left Ventricle Segmentation from Cardiac MRI Using U-Net with MobileNetV3 Encoder. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(7), 82-88, Article DOI https://doi.org/10.47001/IRJIET/2024.807008

****

