Left Heart Segmentation in Echocardiographic Images Using DNN Model

Anshika Jain, IEC2020099
Department of Electronics and Communication
Indian Institute of Information Technology Allahabad

Abstract*—Echocardiography stands as the cornerstone in evaluating the function of the left and right ventricle, a crucial component responsible for pumping oxygen-rich blood throughout the body. Accurate assessment of the left ventricle's structure and function is paramount for guiding treatment decisions and predicting patient outcomes. In this research, we analyze and compare the performance of three segmentation models: U-Net, NN-Unet, and ACNN on LV endocardium, LV epicardium, and left atrium using the CAMUS dataset. The results demonstrate the superiority of NN-Unet, underscoring its potential clinical impact in facilitating more accurate segmentation of the ventricles based on parameters such as Haus-Dorff, MAD, Correlation, Jaccard Index, MAE.

Keywords*—Echocardiography, left ventricle segmentation, deep neural networks, U-Net, NN-Unet, ACNN, CAMUS dataset.

I. INTRODUCTION

The main non-invasive method for evaluating cardiac illness is two-dimensional echocardiogram (2D echo), which is distinguished by its quick acquisition times and high temporal resolution (usually between 50 and 250 frames per second). With the use of grayscale images produced by this imaging method, physicians may assess heart function and recognise anatomical features. In particular, the diagnosis of cardiovascular illnesses and the measurement of heart function depend on the segmentation of the left ventricle (LV) walls in 2D echo. Because the left ventricle is in charge of pumping oxygen-rich blood throughout the body, any impairment to its operations can have a major negative impact on health.

Precise division of LV enables comprehensive approximation of its dimensions, form, and functionality. This involves taking measurements of vital signs of heart health such the ejection fraction (EF) and left ventricular sizes. For example, myocardial infarction, also referred to as heart failure, happens when the left ventricle (LV) is unable to pump blood efficiently, resulting in symptoms such as weariness, fluid retention, and dyspnea. Therefore, early identification of LV failure may improve therapy results and even save lives.

Currently, experts primarily rely on semi-automatic or manual delineation techniques to identify LV boundaries in 2D echo images. However, these approaches are time-consuming and subjective, leading to potential intra-observer and inter-observer variability.

Automatic LV segmentation methods included pixel classification, image-based methods, deformable methods, active appearance and shape models (AAM/ASM), and atlas models prior to the development of deep learning techniques. Although these techniques produced excellent segmentation models for CT and MRI images, they had difficulty with 2D echo images because of issues with speckle noise, brightness inhomogeneities, the presence of trabeculae and papillary muscles, and changes in shape and motion.

Furthermore, earlier approaches had trouble developing models that could manage the LV's varied forms and dynamics. This research explores the segmentation analysis of echocardiograms utilising three distinct DNN architectures: U-Net, NN-UNet, and ACNN, in light of these difficulties. We assess these architectures using multiple metrics: Hausdorff distance, correlation coefficient, mean dice coefficient, jaccard index, MAD, MAE, accuracy, precision, and recall.

This research attempts to shed light on the possibilities of several DNN architectures for improving echocardiographic analysis by examining their performance. These discoveries may contribute to improvements in automated cardiac image analysis, which would ultimately improve cardiology patients' access to clinical diagnosis and treatment.

A. Motivation

Since echocardiography, often known as an echocardiogram, is non-invasive, versatile, and allows for real-time visualisation, it is a vital diagnostic tool in cardiology. Echocardiography gives physicians precise views of the structure and function of the heart, including its chambers, valves, and blood flow patterns. This thorough evaluation makes it possible to identify cardiac anomalies early on, even before symptoms appear, which allows for prompt intervention and treatment. Additionally, by offering vital details regarding chamber size, wall motion,

ejection fraction, and valvular function, echocardiography helps to guide therapy decisions. It is an essential tool in contemporary cardiology practice due to its broad range of cardiac condition detection and monitoring capabilities, ranging from congenital heart defects to heart failure and valve illnesses.

The left ventricle in particular may be precisely delineated from MRI data thanks to research. By producing quantitative analysis of characteristics like volume and wall thickness, these algorithms improve treatment planning and diagnostic precision. In comparison to manual techniques, machine learning-based segmentation is more efficient and consistent, improving workflow and lessening the workload of clinicians. Large datasets can also be used to help these models adapt and learn, which gradually increases accuracy. Segmented echocardiograms are also a useful source of training data for strong model development.

Overall, machine learning-driven echocardiogram segmentation is promising for advancing cardiac imaging analysis, offering rapid, standardised, and accurate assessment for improved patient care.

II. LITERATURE SURVEY

Numerous studies have explored the application of deep learning (DL) approaches to segment echocardiography images. In 2019, [1] conducted a comparative analysis of various DL methods for segmenting the left ventricular endocardium and myocardium, highlighting the effectiveness of encoder-decoder-based architectures over traditional non-DL methods and the DL model accurately replicates expert assessments for end-diastolic and end-systolic left ventricular volumes, achieving an average correlation of 0.93 and a mean absolute error of 9.5 ml. The results are less consistent for the left ventricular ejection fraction, with a mean absolute error of 5.6% and an average correlation value of 0.80.

In 2020, [2] presents a novel method for employing a modified U-Net architecture to segment the left ventricle in echocardiographic pictures. Remaining learning and dilated convolutions are combined in this architecture to increase the efficiency and accuracy of segmentation. The vanishing gradient issue is resolved with the use of residual connections, which improves the model's capacity to acquire more detailed characteristics. Dilated convolutions are very useful for acquiring context in medical images because they broaden the receptive field without adding more parameters. On the other hand, a hausdorff distance of 6.5 and a dice value of 0.867 were obtained.

In 2020, [19] utilised Convolutional Neural Networks (CNNs) to segment the 3D left ventricular endocardium (LVEndo) with an anatomically constrained neural network (ACNN). This network, based on a 3D U-Net architecture [20], ensures the segmentation outputs fit a non-linear, compact anatomical representation derived from an auto-encoder. On the CETUS dataset, the method achieved average Dice scores of 0.912 (ED) and 0.873 (ES), mean absolute distances of 1.9 mm (ED) and 2.1 mm (ES), and Hausdorff distances of 7.0 mm (ED) and 7.7 mm (ES) [19]. These results are similar to those by [7]. The use of only 15 patients for training highlights the significant potential of deep learning techniques in analysing echocardiographic images.

In 2022, [5] introduced a novel segmentation technique that made precise predictions about optical flow (OF) fields in order to use motion information. In order to facilitate effective information interchange, it extracts characteristics for the segmentation and optical flow subtasks. Additionally, it established a new orientation congruence constraint for the OF estimation subtask, which guarantees optical flow orientation constancy between consecutive frames. Furthermore, better segmentation accuracy was offered by the motion enhanced segmentation module. On echocardiographic sequences from 450 patients, it does, however, achieve a correlation value of 0.893 and a mean absolute error of 5.20%.

In 2023, [6] proposed a lightweight segmentation model, LVNet, for segmenting the left ventricle. LVNet offers the advantage of requiring fewer parameters while achieving improved segmentation performance. Compared to state-of-the-art methods like UNet, MiniNetV2, and fully convolutional dense dilated network (FCdDN), LVNet shows superior performance. It includes a post-processing pipeline that further refines segmentation results. [28] proposed a pipeline for real-time segmentation of the left ventricle (LV) on a mobile device, using input from point-of-care devices. This pipeline employs a lightweight model based on UNet to perform the segmentation efficiently. These produce similar results: dice score is 0.902 and Jaccard Index is 0.823.

Recently, [20] explores reducing manual annotation by pretraining CNNs with data generated from an automatic Kalman filter-based segmentation method. The CNN achieved a dice similarity coefficient of 0.86 ± 0.06, comparable to the Kalman filter's 0.87 ± 0.06, and a better Hausdorff distance of 5.9 ± 2.9 mm versus 7.5 ± 5.6 mm. This approach could allow CNNs to surpass current methods with minimal expert annotations for fine-tuning. [26] aimed to segment the left ventricular endocardium and myocardium, focusing on algorithm design and testing on both porcine and human images. While their method showed promise, limitations arose from its reliance on open-chest pig images, which offer higher quality than human echocardiograms. [11] proposed a hybrid architecture combining modified

U-Net and FCN encoder, emphasising improved feature extraction and error learning capabilities.

III. METHODOLOGY

When we segment echocardiographic images, we go through a series of steps. First, we clean up the images to reduce any noise and make the details clearer. Then, we use different techniques, like setting thresholds or looking at specific regions, to pinpoint and outline important structures like the heart muscle, chambers, and valves. This process helps us get a clear picture for analysis and diagnosis based on the echocardiogram results. Detailed methodology is shown in Fig. 1.

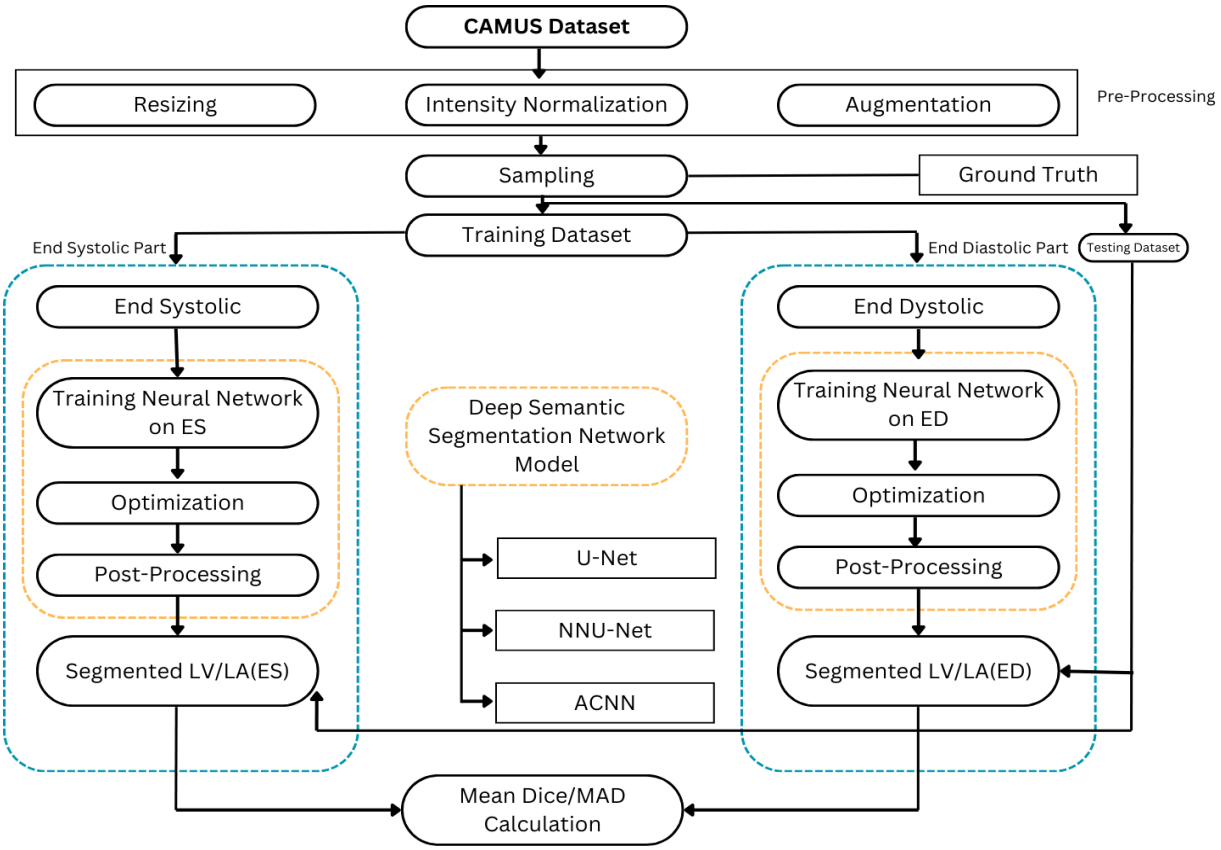


Fig. 1 illustrates the overall block diagram representing the complete methodological flow of echocardiographic images segmentation process using DNN models such as U-Net, NN-Unet, and ACNN.

A. Dataset Preparation:

We have used the CAMUS[1] dataset for the segmentation.

1) Dataset Description:

CAMUS Dataset: The dataset comprises echocardiographic scans from 500 patients conducted at the University Hospital of St Etienne, France, adhering to ethical guidelines. Utilising GE Vivid E95 ultrasound scanners with a GE M5S probe, the images captured two-dimensional apical four-chamber and two-chamber views for each patient. The dataset was divided into 10 folds for cross-validation, with each fold containing 50 patients.

Considering the dataset consists of two and four-chamber acquisitions, this implies that each fold comprises a total of 100 echocardiographic images (50 patients x 2 acquisitions per patient).

The dataset, available to the community, includes:

1. A training set of 450 patients, accompanied by manual references derived from expert clinical analysis. Out of 450 patients, 400 are used as a train set and 50 patients were used as a validation set.
2. A testing set comprising 50 additional patients. Raw input images are provided in the raw/mhd file format.

One of the notable features of the CAMUS dataset is its large size, containing thousands of high-resolution images captured from multiple viewpoints and cardiac cycles. This extensive coverage enables researchers to explore cardiac dynamics comprehensively and facilitates the development and evaluation of advanced image analysis algorithms, such as segmentation and motion tracking techniques.

The CAMUS dataset includes annotated ground truth data, delineated by expert clinicians, which serve as reference standards for algorithm validation and performance assessment. These

annotations provide precise delineation of cardiac structures, including the left ventricle endocardium, left ventricle epicardium and left atria enabling accurate evaluation of segmentation algorithms' efficacy. Annotation of End Diastole and End systole has been shown in Fig. 2. The dataset include 206 good quality images, 200 medium quality images and 44 poor quality images. The dataset consists of a total 150 female and 300 male patients varying from age 20 to 85 years. Detailed description is given in Fig. 3.

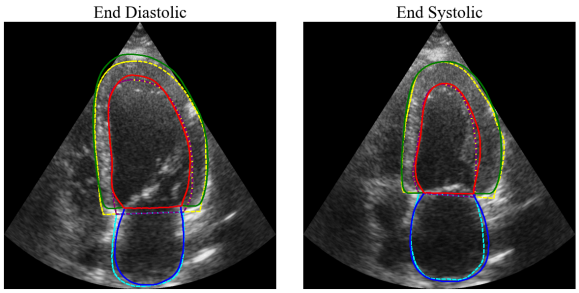
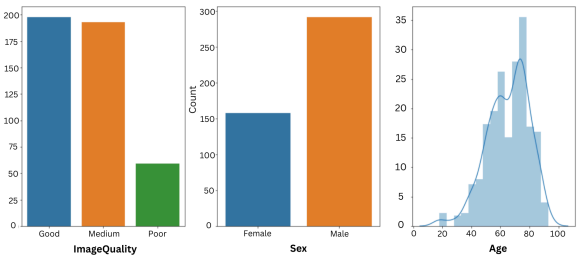


Fig. 2 Standard images taken from dataset. They show the endocardium and epicardium of the left ventricle and left atrium wall, marked in green, red, and blue, respectively. On the left side is the annotated ED, and on the right is the annotated ES.



a. Image Quality and count b. Sex and count c. Age and count

Fig. 3 illustrates the distribution of CAMUS dataset images on the bases of (a) Image Quality Vs number of CAMUS echocardiographic images, (b) Sex Vs number of patients and (c) Age Vs number of patients.

2) Data Cleaning:

The training dataset, which includes 19% of all images, was filtered to remove low-quality echocardiogram pictures in order to avoid any potential detrimental effects on the functionality of deep learning models. The training process might concentrate on high-quality data by eliminating these photos, which would improve the model's capacity to efficiently pick up pertinent features and patterns. Removing poor quality photos contributed to the deep learning models' continued resilience against noise and image quality fluctuations. This method made sure the models could handle the variability found in echocardiographic pictures and generalise well to new data.

Cross-Validation for Error Detection:

1. By dividing the dataset into 10 folds for cross-validation, the process enabled the identification of potential inconsistencies or outliers in the data distribution.
2. Through this iterative validation approach, any data cleaning issues could be detected and resolved, ensuring the dataset's overall integrity for subsequent analyses.

3) Equal distribution of data:

1. Cross-Validation Approach:
   a. The researchers adopted a cross-validation technique by dividing the dataset into 10 folds. This method ensured through training and evaluation coverage across different data subsets.
   b. Utilising this strategy enhanced the reliability of the model assessment by validating its performance on diverse data partitions.
2. Allocation of Training and Validation Sets:
   a. The dataset was segregated into distinct training and validation sets. This segregation facilitated model training and performance evaluation.
   b. Training on one subset while validating on another aided in preventing overfitting and provided insights into the model's ability to generalise to unseen data.
3. Progressive Model Assessment:
   a. Models underwent training on varying patient sample sizes, with evaluations conducted on fold 5 to monitor segmentation performance.
   b. This iterative evaluation process allowed for a comprehensive analysis of model performance with increasing training dataset sizes, offering valuable insights into the model's learning progression and convergence patterns.

B. Data Augmentation:

1. Data augmentation techniques were crucial for enhancing the robustness and generalisation of the deep learning models trained on the echocardiographic dataset.
2. The study employed various augmentation strategies such as rotation, scaling, and flipping to artificially increase the diversity of the training data [1].
3. These augmentations help the models learn invariant representations of the cardiac structures, making them more adaptable to variations in orientation and size within the images.
4. By augmenting the dataset, the models can better capture the underlying patterns in the echocardiographic images, leading to improved segmentation performance and clinical index estimation [3].
5. Various techniques includes -

a. Rotation: Images were rotated to different angles to simulate variations in orientation, helping the models learn invariant representations of cardiac structures.
b. Scaling: Resizing images to different scales allowed the models to capture variations in size and perspective, enhancing their ability to segment structures accurately.
c. Flipping: Horizontal and vertical flipping of images provided additional training samples, aiding the models in learning robust features regardless of the image orientation.

C. Image Processing:

1. The study implemented preprocessing techniques to enhance the echocardiographic images before feeding them into the deep learning models.
2. Techniques like normalisation were used to standardise pixel values across images i.e. their original scale (ranging from 0 to 255 for 8-bit images) to a grey scale, promoting consistency and aiding model training.
3. Resizing to 256 * 256 pixels and cropping were applied to ensure uniform image dimensions, facilitating model input standardisation.
4. Noise reduction methods were employed to improve image quality by reducing artefacts and enhancing the visibility of cardiac structures.
5. Tailored preprocessing steps addressed challenges unique to echocardiographic imaging, such as contrast issues and brightness variations, to improve segmentation accuracy.
6. Overall, the preprocessing steps aimed to optimise the images for accurate segmentation of cardiac structures and clinical index estimation, enabling the deep learning models to effectively learn from the data.

D. Theoretical basis:

While evaluating and building our segmentation model we will require various mathematical and general terms which are explained below.

1) Activation Function:

1.1 ReLU Activation Function: Rectified Linear Unit activation function is a popular choice in neural network which can be represented as:

$$f(x) = \max(0, x)$$

Here, x is the input to the function and f(x) is the output. The ReLU function outputs the input value *x* if it is positive, and zero otherwise. This means that the function passes through any positive input unchanged, while setting all negative inputs to zero. The key advantage of ReLU lies in its ability to introduce non-linearity into the network, allowing it to learn complex patterns and relationships in the data. Additionally, ReLU's computational efficiency and avoidance of the vanishing gradient problem.

1.2 SoftMax Activation Function: It is a type of sigmoid function applied to the output of a neural network to produce a probability distribution for each pixel across different classes or categories. For every pixel in the image, the neural network computes raw scores (logits) indicating the likelihood of belonging to each class. These logits are then processed through the softmax function independently for each pixel position. Essentially, the softmax function normalises the logits into probabilities, ensuring that they sum up to one. This transformation enables pixel-wise classification, where each pixel is assigned a probability of belonging to each segmentation class, facilitating the segmentation process by identifying the most probable class for each pixel. Softmax function is given by:

$$p(y = j/x) = \exp(z\_j) / \Sigma \exp(z\_k)$$

It maps the linear combination of input features to a probability distribution over the different classes. For j = 1,2,...,K classes, where K is the number of classes. Here, z_j is the linear combination of input features for class j, and Σ exp(z_k) sums over all K classes.

E. Model Evaluation:

1) Dice Loss: It measures the similarity or overlap between two sets. It quantifies how well the predicted region (e.g., object boundaries) aligns with the ground truth region (the actual location of objects in the image). The Mean dice coefficient ranges from 0 to 1. Higher the dice coefficient higher the similarity between predicted and ground truth mask.

$$\text{Dice coefficient} = 2 * |X \cap Y| / (|X| + |Y|)$$

Where,
|X| represents the number of pixels in predicted mask
|Y| represents the number of pixels in Ground Truth mask
|X ∩ Y| represents the number of overlapping pixels in both masks

2) Mean Hausdorff: It measures the average distance between the boundaries of the predicted segmentation mask and the ground truth mask. And then averages these distances over all points in the mask in both masks. The value ranges from 0 to infinity. Lower values indicating better segmentation accuracy. Mathematically, the Mean Hausdorff Distance between two sets X and Y is defined as:

$$d_{AHD(x,y)} = \left(\frac{1}{X} \sum_{x \in X} \min d(x, y) + \frac{1}{Y} \sum_{y \in Y} \min d(x, y) \right) / 2$$

Where,
x represents the set of points or contour pixels in the

predicted segmentation mask.
y represents the set of points or contour pixels in the ground truth segmentation mask.
|x| and |y| denote the cardinalities of sets X and Y, respectively.
d(x, y) calculates the distance between points x, y in Euclidean space.

3) MAD: It is the average distance between the observation points and the mean. The value ranges from 0 to infinity. Lower values are associated with closely related data points.

$$MAD = \sum \frac{|x_i - \overline{x}|}{n}$$

Where,
MAD = average value of the data set
n = number of data values
xi = data values in the set
$\bar{x}$ = mean value

4) Correlation Coefficient: It is used to measure the extent of the relationship between two or more variables. The correlation coefficient ranges from -1 to + 1. -1 indicates the predicted and ground truth masks are uncorrelated, whereas +1 indicates the predicted and ground truth masks are completely correlated.

$$r_{xg} = \frac{n\sum_{i=1}^{n} x_i g_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} g_i}{\sqrt{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} g_{i^2} - \left(\sum_{i=1}^{n} g_i\right)^2}}$$

Where,
n = Data quantity or number of data available
Σx = Total of the predicted data point
Σg = Total of the ground truth data point
Σxg = Sum of the Product of predicted & ground truth data points
Σx2 = Sum of the Squares of the predicted data point
Σg2 = Sum of the Squares of the ground truth data points

5) Precision: It measures the accuracy of the model's positive predictions by evaluating the proportion of correctly segmented pixels relative to all pixels predicted as positive. Higher precision values indicate lesser false positives and better accuracy.

Precision = (pred_mask * ground_mask) / pred_mask

Here, pred_mask and ground_mask are the predicted mask and ground truth mask respectively.

6) Recall: It assesses the model's ability to capture all positive instances, measuring the proportion of correctly segmented pixels relative to all the pixels that should have been identified as positive. A higher recall value specifies lesser missed detections and better performance in capturing relevant pixels of the target class.

Recall = (pred_mask * ground_mask) / ground_mask

Here, pred_mask and ground_mask are the predicted mask and ground truth mask respectively.

7) Jaccard Index (I0U): It measures the similarity between finite sample sets A,B as the Intersection over Union (IoU). Its value ranges from 0 to 1 where 0 indicates no overlap and 1 indicates complete overlap between predicted and ground truth mask.

$$\text{IoU} = \frac{\text{predMask * groundtruthMask}}{\text{predMask + groundtruthMask - predMask * groundtruthMask}}$$

8) **MAE** - It measures the average of the errors' magnitude between the predicted and actual values. Its value ranges from 0 to infinity where lower values are better.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \overline{y_i}|$$

Where,
n: number of observation
yi: the actual value of the ith observation
yi': the predicted value of the ith observation

F. Model Selection:

In this study, our aim is to evaluate the effectiveness of Convolutional Neural Networks (CNNs) in segmenting 2D echocardiographic images. To accomplish this, we decided to concentrate on encoder-decoder networks (EDNs), which have proven to be fundamental in various CNN architectures that have shown success in medical imaging tasks [23]. EDNs utilise a two-stage convolutional network design, making them particularly suitable for segmentation tasks. The encoder, the initial segment, comprises a sequence of convolutional layers and downsampling steps. These processes are aimed at capturing image features while reducing spatial dimensions, facilitating the extraction of abstract features. Following the encoder, the decoder utilises the extracted features and employs a series of convolutional layers and upsampling steps. This gradual transformation of feature maps by the decoder ultimately produces the final segmentation output.

The selection criteria include model performance, computational efficiency and scalability. Each model is implemented using popular DL framework TensorFlow, and experimentally tuned hyperparameters.

1) Model Architecture:

For cardiac image segmentation, following fully connected layers are used:
A. ReLU activation functions to introduce non-linearity.
B. Batch normalisation layers to stabilise and accelerate training.

C. Softmax activation function in the final layer for semantic segmentation.
D. Dropout layers are removed to enhance the model capacity to learn intricate features.
E. Global average pooling layers to reduce spatial dimensions and extract features.

2) Training Parameters:

Training consists of 30 epochs with variable batch size, indicating that the entire dataset is divided into batches of samples each for training.

ADAM optimizer is used with the following parameters:

- Learning rate: 1e4
- Decay: 1e6
- Momentum: 0.9

2.1. U-Net:

Convolutional neural networks (CNNs) using the U-Net architecture are commonly used for image segmentation tasks, especially in biomedical imaging. Its unique architecture is based on an encoder-decoder architecture with a contracting path and an expanding path.

At the core of the U-Net architecture lies its encoder, which is responsible for extracting features from input images. A total of 3 downsampling layers in encoder were used which comprises a series of 3*3 convolutional layers while doubling the number of features for downsampling interspersed with rectified linear unit (ReLU) activation functions, which introduce non-linearity to the network. These convolutional layers are followed by max pooling layers, which downsample the feature maps, thereby reducing spatial dimensions and extracting high-level features.

After the encoder, the network transitions to the decoder phase, where the upsampled feature maps are gradually reconstructed to produce the final segmentation output. In the decoder, total of 3 upsampling layers were used with 3*3 convolutional layers with ReLU activation functions are employed to learn intricate spatial patterns, while upsampling operations are utilised to restore spatial dimensions. Number of features are reduced in half per sampling. Additionally, skip connections are incorporated between corresponding encoder and decoder layers, facilitating the retention of fine-grained spatial information and aiding in the recovery of detailed features. The final 3*3 convolutional layers of 384 dimension each are used and softmax classification technique is used. Detailed Architecture of U-net is shown in Fig. 4.

Batch size [3] of 8 has been implemented. The loss function [5] used in the training of this architecture is multi-class dice loss. Adam [12] optimizer is used in loss minimization. Total number of trainable parameters for this architecture is 18M. Standard evaluation matrices in the form of dice coefficient, mean hausdorff, MAD, accuracy, recall and precision are used in both training and testing data.

The encoder-decoder structure of the U-Net architecture is utilised by its convolutional layers, max pooling layers, ReLU activation functions, and normalisation layers. Because of its design, the network can efficiently extract and incorporate spatial information from input images while maintaining minute features that are essential for precise segmentation outputs.

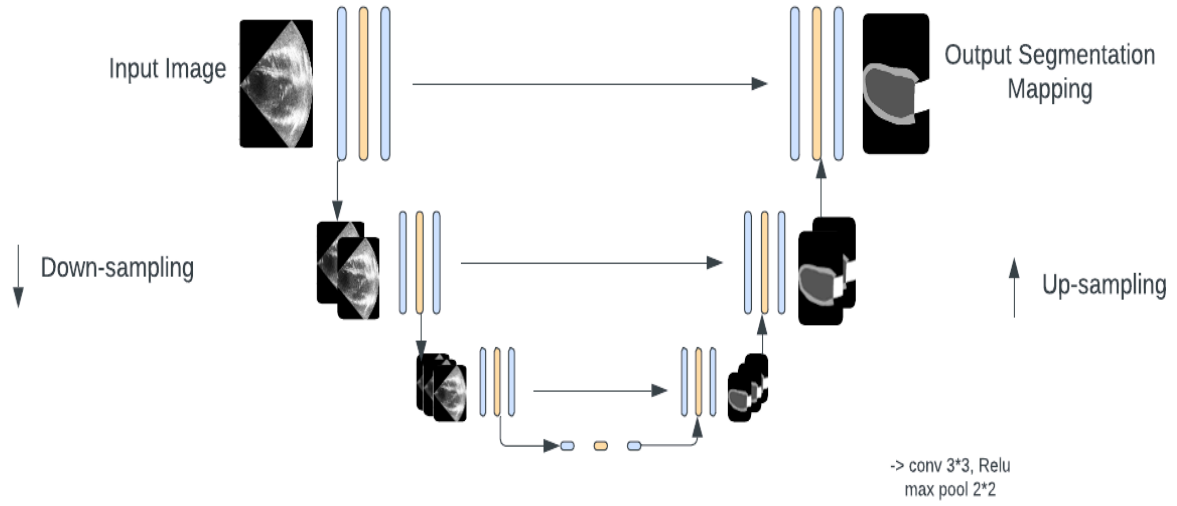


Fig. 4 The U-Net encoder-decoder deep learning architecture representation

B. ACNN [3]:

The Attention Convolutional Neural Network presents a new method for obtaining full resolution feature processing. It deliberately uses the best dilation setting to increase the receptive field while minimising the number of parameters. Thanks to this breakthrough, ACNN can now outperform other methods, achieving superior segmentation Intersection over Union (IoU) scores while requiring a far smaller number of trainable parameters and overall model sizes. This emphasises the benefits of using full-resolution feature maps for extensive feature processing jobs.

The ACNN (Auxiliary Convolutional Neural Network) model implemented in the study integrated an auxiliary loss to enhance segmentation accuracy by aligning the segmentation output with a compact representation of the underlying anatomy[5]

The ACNN model utilised the U-Net 1 architecture as the segmentation module, incorporating specific design choices to optimise performance on the dataset. Receptive field (RF) values are chosen as 1537, giving us a total of 192 Atrous II-blocks[3].

Key parameters in the ACNN model included a code of 32 coefficients for the auto-encoder network, enabling an average reconstruction accuracy of 97, and a hyperparameter balancing segmentation and shape regularisation losses for improved accuracy[12]

With a total of 5.5 million parameters, the ACNN model aimed to refine the segmentation process by leveraging anatomical constraints derived from the auto-encoder network, enhancing the model's ability to accurately delineate cardiac structures in 2D echocardiographic images.

Through these tailored design elements, the ACNN model sought to improve segmentation precision and shape regularisation, contributing to more accurate and reliable analysis of cardiac structures in clinical applications. Detailed architecture is shown in Fig. 5.

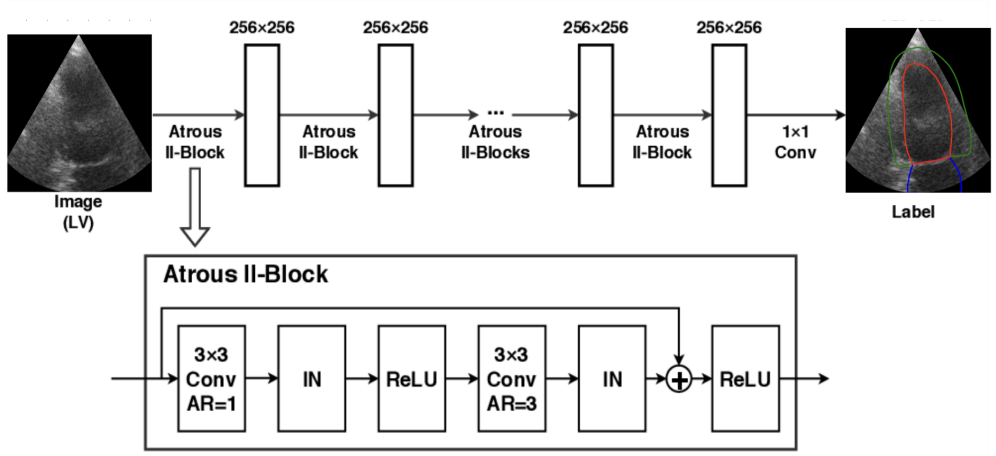


Fig. 5 The architecture of ACNN. It involves a series of residual II-blocks, with the number of these blocks determined by a formula (RF - 1) / 8, based on the targeted receptive field (RF). The atrous rate (AR) is used for atrous convolutions, with a kernel size of 3 for 3x3 Conv and a kernel size of 1 for 1x1 Conv1.

C. NN-Unet:

The NN-Unet architecture is an extension of the U-Net architecture, designed to enhance its performance in medical image segmentation tasks. Similar to U-Net, NN-Unet also follows an encoder-decoder framework but incorporates additional features to improve segmentation accuracy.

In the encoder part of NN-Unet, a total of 5 times down-sampling is done until the feature map reaches either 4 voxels or the feature map space becomes anisotropic, followed by a rectified linear unit(ReLU). This down-sampling process occurs separately for each axis, with high-resolution axes down-sampled until their resolution is approximately half that of the lower resolution axis. Each axis is handled individually during the down-sampling process, stopping when the feature map constraints are met. By default, convolutions are performed with a kernel size of 3×3 for 2D U-Net. This increased depth allows for better feature extraction and representation, which is particularly beneficial for complex segmentation tasks in medical imaging. Additionally, batch normalisation layers are included after each convolutional layer to stabilise and accelerate the training process by normalising the activations.

Rather than traditional max pooling layers used in U-Net, NN-Unet employs more sophisticated downsampling techniques. In this case we have used 2*2 strided convolutions [7]. These alternatives help preserve spatial information better while reducing the spatial dimensions of feature maps, thereby enhancing the network's ability to capture fine details.

Upsampling techniques are used in the NN-Unet decoder section to progressively recreate the spatial dimensions of feature maps. To

concatenate feature maps from the encoder to the respective decoder layers, skip connections akin to those in U-Net are employed. This helps with the accurate localization of segmentation boundaries by enabling the combination of low-level and high-level data. Softmax probabilities are averaged during the ensemble averaging process, which unifies the model. Both training and testing data are subjected to standard assessment matrices, such as dice coefficient, mean hausdorff, MAD, accuracy, recall, and precision. Detailed architecture is shown in Fig. 6.

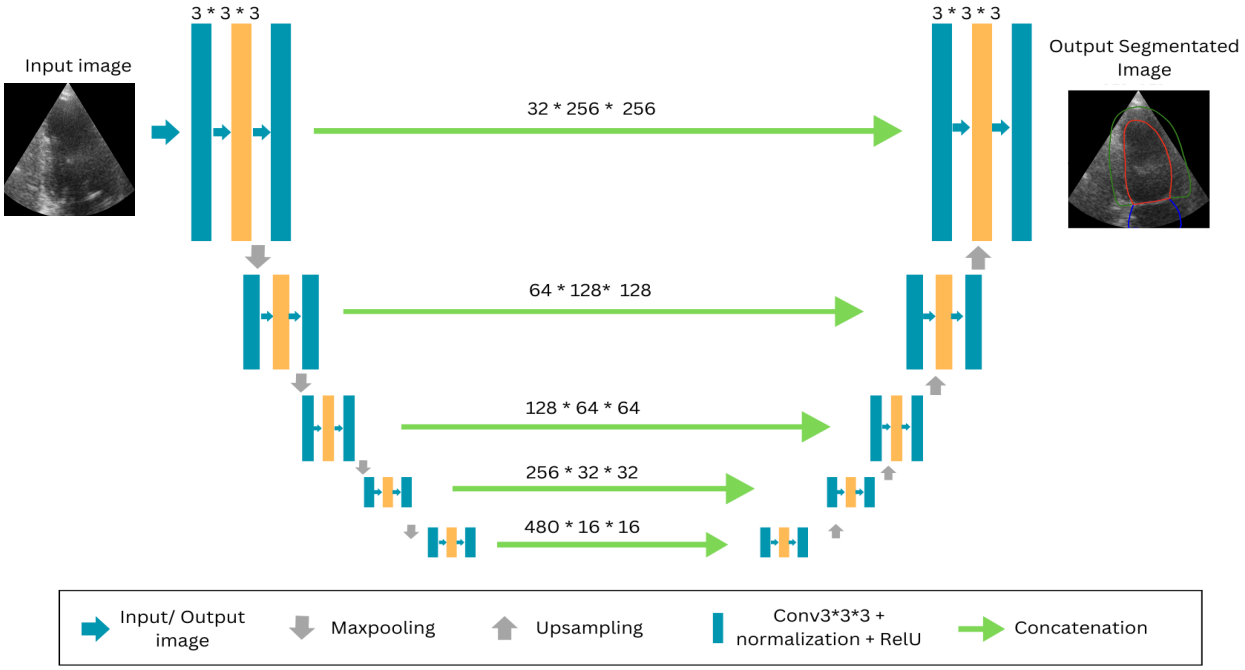


Fig. 6 illustration of NN-Unet network architect

TABLE I
Main characteristics of all 3 DNN segmentation model architecture including U-Net, ACNN and NN-Unet

Architectures	Lowest Resolution	Upsampling Scheme	Downsampling Scheme	Normalisation Scheme	Batch Size	Learning Rate	Loss Function	Trainable Parameters	Optimizer
U-Net	24 * 24	3 * 3 repeats	3 * 3 repeats	None	32	1e-4	Multi-class Dice Loss	18M	ADAM
ACNN	32 * 32	Deconvolution	3 * 3 Atrous convolution	Batch normalisation	32	1e-4	Multi-class Dice Loss/ Auxiliary Loss	5.5M	ADAM
NNU-Net	16 * 16	Deconvolution	convolution	Batch normalisation	32	1e-4	Multi-class Dice Loss	2.4M	ADAM

IV. Experimental Results

We excluded 19% of low-quality images from the analysis to ensure the accuracy of the metrics presented in this section. Additionally, to maintain consistency across acquisition settings, we opted to train a single model for each machine learning approach using two-chamber (2Ch) views, regardless of the specific time point in the cardiac sequence. The trained models are evaluated on the basis of accuracy, precision, recall along with mean dice coefficient, MAE, MAD, mean correlation, jaccard index and hausdorff distance.

A. U-Net:

U-Net achieved the accuracy of 95.7% along with a mean dice score of 94.7% for the left ventricle (endocardium). Similar results are also achieved for the left ventricle (epicardium) and left atrium. Mean hausdorff value of 4.1 which is significantly less. It also achieved the correlation for LVED of 0.959. Mean Dice Plot's results have been shown in Fig. 7. Fig. 8 shows the training and testing accuracy, loss curve and dice coefficient graph for ED and ES respectively.

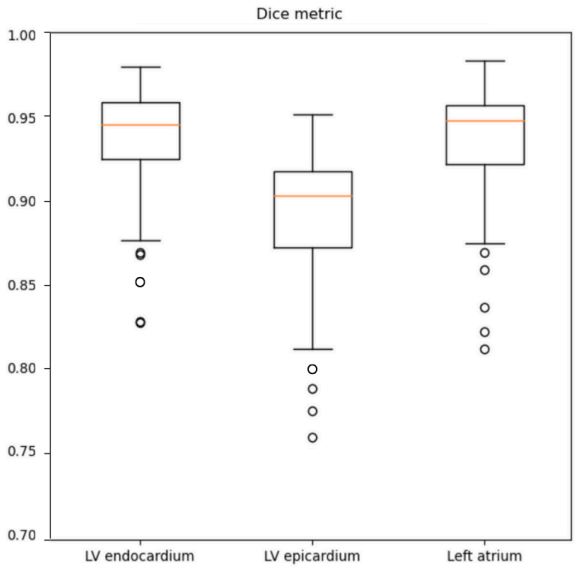
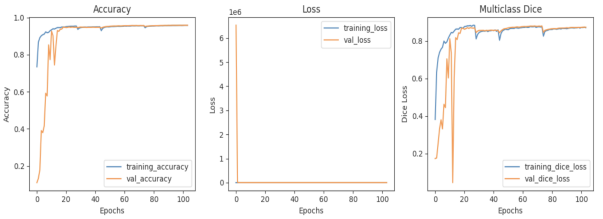
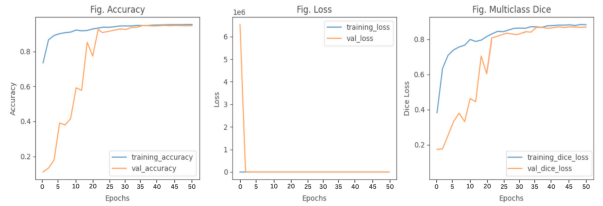


Fig. 7 Dice score for LV endocardium, LV epicardium and left atrium for U-Net architecture



a. Accuracy vs Epochs b. Loss Vs Epochs c. Mean Dice Vs Epochs



e. Accuracy vs Epochs f. Loss Vs Epochs g. Mean Dice Vs Epochs

Fig. 8 (a), (b), (c) shows the accuracy, loss and mean dice plots Vs epochs for LVED. Whereas (e), (f), (g) shows the accuracy, loss and mean dice plots Vs epochs for LVES for U-Net architecture.

B. ACNN:

U-Net achieved the accuracy of 93.6% along with a mean dice score of 93.3% for the left ventricle (endocardium). Similar results are also achieved for the left ventricle (epicardium) and left atrium. Mean hausdorff value of 5.4 which is significantly less. It also achieved the correlation for LVED of 0.930. Mean Dice Plot's results have been shown in Fig. 9. Fig. 10 shows the training and testing accuracy, loss curve and dice coefficient graph for ED and ES respectively.

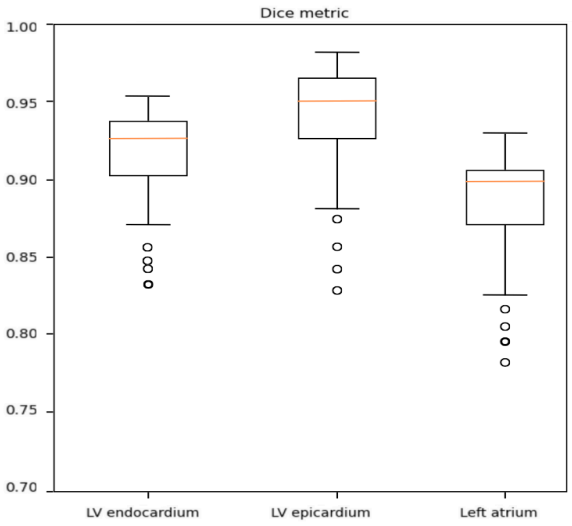
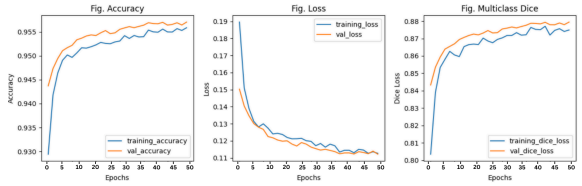
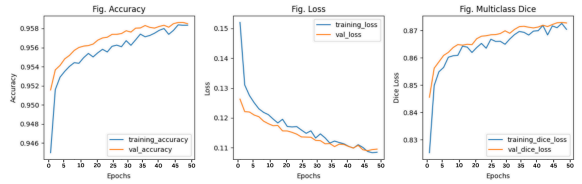


Fig. 9 Dice score for LV endocardium, LV epicardium and left atrium for ACNN architecture



a. Accuracy vs Epochs b. Loss Vs Epochs c. Mean Dice Vs Epochs



e. Accuracy vs Epochs f. Loss Vs Epochs g. Mean Dice Vs Epochs

Fig. 10 (a), (b), (c) shows the accuracy, loss and mean dice plots Vs epochs for LVED. Whereas (e), (f), (g) shows the accuracy, loss and mean dice plots Vs epochs for LVES for ACNN architecture.

C. NN-Unet:

NN-Unet performed the best as compared to U-net and ACNN.It achieved the accuracy of 96.7% along with a mean dice score of 95.4% for the left ventricle (endocardium). Similar results are also achieved for the left ventricle (epicardium) and left atrium. Mean hausdorff value of 4.5 which is significantly less. It also achieved the correlation for LVED of 0.983. Mean Dice Plot's results have been shown in Fig. 11. Fig. 12 shows the training and testing accuracy, loss curve and dice coefficient graph for ED and ES respectively.

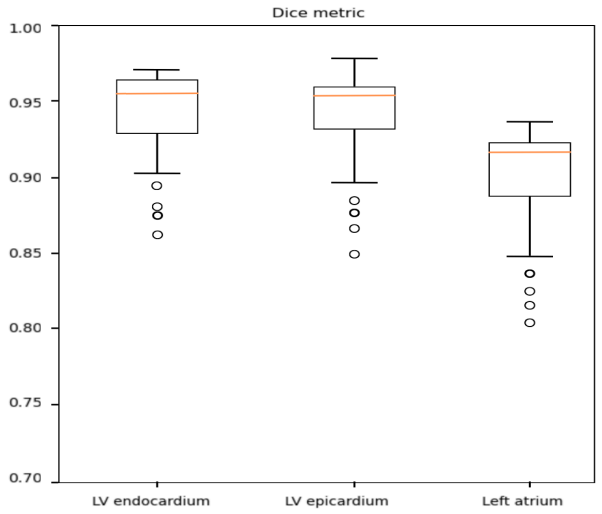
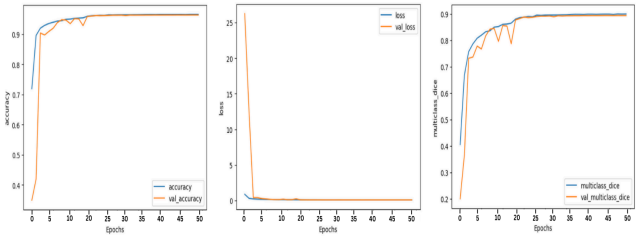
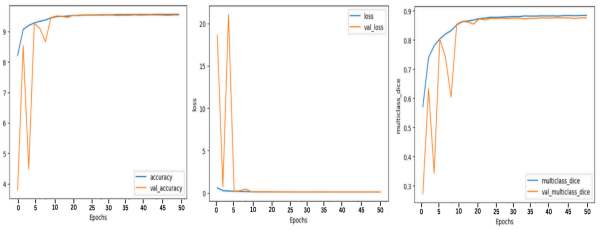


Fig. 11 Dice score for LV endocardium, LV epicardium and left atrium for NN-Unet architecture



a. Accuracy vs Epochs b. Loss Vs Epochs c. Mean Dice Vs Epochs



e. Accuracy vs Epochs f. Loss Vs Epochs g. Mean Dice Vs Epochs

Fig. 12 (a), (b), (c) shows the accuracy, loss and mean dice plots Vs epochs for LVED. Whereas (e), (f), (g) shows the accuracy, loss and mean dice plots Vs epochs for LVES for NN-Unet architecture.

Overall NN-Unet provided the best results of LVEndo, LVEpi and Left Atrium segmentation. Table II, Table III and Table IV shows the mean dice score, mean hausdorff, MAD, correlation, accuracy, precision, recall, MAE and jaccard Index for left ventricle endocardium (LVEndo), left ventricle epicardium (LVEpi) and left atrium ED and ES respectively.

Table II:
Segmentation accuracy for LV Endocardium of the 3 evaluated segmentation methods i.e. U-Net, ACNN and NN-Unet on 50 test patients having good and medium image quality. The values in bold refer to the best performance for each measure for LVEndo. All the scores obtained with NN-Unet show that it is the best performing method overall.

	Mean Dice ED	**Mean Dice ES**	**Mean Hausdorff ED**	**Mean Hausdorff ES**	**MAD ED**	**MAD ES**	**Correlation ED**	**Correlation ES**
U-Net	0.947	0.925	**4.1**	4.4	1.5	**1.3**	0.938	0.959
ACNN	0.933	0.915	5.4	5.2	1.7	1.7	0.930	0.947
NN-Unet	**0.954**	**0.949**	4.5	**4.3**	**1.5**	1.5	**0.978**	**0.983**

	Accuracy ED (%)	**Accuracy ES (%)**	**Precision ED**	**Precision ES**	**Recall ED**	**Recall ES**	**MAE ED**	**MAE ES**	**Jaccard Index ED**	**Jaccard Index ES**
U-Net	95.7	**95.8**	0.938	0.927	0.873	0.912	11.2	7.5	0.889	0.873
ACNN	93.6	94.8	0.932	0.911	0.890	0.934	9.7	6.9	0.905	**0.901**
NN-U net	**96.7**	94.6	**0.950**	**0.943**	**0.953**	**0.947**	**5.9**	**4.0**	**0.915**	0.891

Table III:
Segmentation accuracy for LV Epicardium of the 3 evaluated segmentation methods i.e. U-Net, ACNN and NN-Unet on 50 test patients having good and medium image quality. The values in bold refer to the best performance for each measure for LVEpi. All the scores obtained with NN-Unet show that it is the best performing method overall.

	Mean Dice ED	**Mean Dice ES**	**Mean Hausdorff ED**	**Mean Hausdorff ES**	**Mean MAD ED**	**Mean MAD ES**	**Correlation ED**	**Correlation ES**
U-Net	0.958	0.936	5.2	5.7	1.7	1.9	0.915	0.902
ACNN	0.946	0.953	5.8	5.7	1.8	2.0	0.928	0.910
NN-Unet	**0.968**	**0.928**	**4.6**	**4.4**	**1.5**	**1.5**	**0.948**	**0.953**

	Accuracy ED (%)	**Accuracy ES (%)**	**Precision ED**	**Precision ES**	**Recall ED**	**Recall ES**	**MAE ED**	**MAE ES**	**Jaccard Index ED**	**Jaccard Index ES**
U-Net	93.8	95.6	0.925	0.933	0.907	0.917	9.9	6.7	0.926	0.915
ACNN	94.6	93.2	0.945	**0.949**	0.869	0.892	10.1	7.8	0.893	0.897
NN-Unet	**97.1**	**96.3**	**0.963**	0.915	**0.947**	**0.955**	**6.2**	**5.4**	**0.945**	**0.938**

Table IV:
Segmentation accuracy for left atrium of the 3 evaluated segmentation methods i.e. U-Net, ACNN and NN-Unet on 50 test patients having good and medium image quality. The values in bold refer to the best performance for each measure for LA. All the scores obtained with NN-Unet show that it is the best performing method overall.

	Mean Dice ED	**Mean Dice ES**	**Mean Hausdorff ED**	**Mean Hausdorff ES**	**Mean MAD ED**	**Mean MAD ES**	**Correlation ED**	**Correlation ES**
U-Net	0.890	0.921	5.6	5.4	2.1	2.0	0.899	0.895
ACNN	0.881	0.870	5.9	6.0	2.3	2.3	0.926	**0.945**
NN-Unet	**0.912**	**0.929**	**5.1**	**4.3**	**2.0**	**1.8**	**0.987**	0.927

	Accuracy ED (%)	Accuracy ES (%)	Precision ED	Precision ES	Recall ED	Recall ES	MAE ED	MAE ES	Jaccard Index ED	Jaccard Index ES
U-Net	93.1	92.6	0.874	0.917	0.889	0.845	13.4	10.7	0.889	**0.895**
ACNN	93.8	91	**0.969**	**0.958**	0.937	0.943	10.9	9.8	0.88	0.890
NN-Unet	**94.2**	**94.1**	0.902	0.931	**0.950**	**0.949**	**7.4**	**6.0**	**0.894**	0.883

Runtime Performance

The three segmentation methods were implemented in Python using the same versions of TensorFlow and Keras libraries, and they were run on an Apple M2 chip with 8-core GPU, a 16-core Neural Engine and 100GB/s memory width. Due to the higher number of trainable parameters in each segmentation method (as shown in Table I), the training times for these networks vary. Time taken for the training of 400 patients and testing of 50 patients for different segmentation architectures along with their trainable parameters are discussed in Table V.

Table V:
Runtime performance and trainable parameters of 3 evaluated methods on 400 training dataset and 50 testing dataset restricted to patients having medium and good image quality.

Architectures	Trainable Parameters	Testing Time	Training Time
U-net	18M	150 ± 5 minutes	1.02 ± 0.1 minutes
ACNN	5.5M	85 ± 5 minutes	0.35 ± 0.6 minutes
NN-Unet	2.4M	54 ± 1 minutes	0.18 ± 0.3 minutes

V. COMPARATIVE ANALYSIS

We have compared our results with methods mentioned in other researches including the dataset used along with methodology implemented and different evaluation methods and its comparative analysis has shown in Table VI.

Table VI:
Segmentation accuracy comparison for different evaluated segmentation methods. The values in bold refer to the best performance for each measure. All the scores obtained with NN-Unet show that it is the best performing method overall.

Title	Dataset	Method Used	Dice	Haus-dorff	MAD	Correlation	Jaccard Index(IOU)	MAE
[1]	CAMUS	U-Net 2	0.916	5.5	1.6	-	-	-
[2]	STACOM	ResDUnet	0.867	6.5	1.6	-	-	-
[3]	Self-Collected	ACNN + DBN	-	-	-	0.86	0.8	-
[4]	UK Digital Heart Project	ACNN	0.866	-	-	-	-	-
[5]	CAMUS	SOCOF	0.814	-	-	0.918	-	-

[6]	Cardiac US image of canine LV	LVnet	0.902	-	-	-	**0.823**	-
[7]	Self-Collected	BEAS	0.825	-	-	-	-	-
[8]	Self-Collected	TransV-Net	0.902	4.9	1.5	-	-	-
[9]	EchoNet-Dynamic	CSS	0.919	4.17	-	-	-	4.9
[10]	CAMUS	PWC Net	0.79	-	-	0.84	-	**2.8**
	CAMUS	U-Net	0.947	**4.1**	**1.5**	0.938	0.889	11.2
	CAMUS	ACNN	0.933	5.4	1.7	0.930	0.905	9.7
	CAMUS	NNU-Net	**0.954**	4.5	**1.5**	**0.978**	**0.915**	**5.9**

VI. CONCLUSION

The study concludes that encoder-decoder networks, specifically the NN-Unet architecture, are effective at accurately segmenting cardiac structures in 2D echocardiography. It also shows that there is a favourable trade-off between model complexity and performance, as well as a balance between the number of parameters and the performance that is attained. Notably, it demonstrated high correlations for end-diastole (ED) and end-systole (ES), along with a minimal absolute mean distance of 1.5 for ED and 1.3 for ES, accurately replicating expert analysis for left ventricular volumes. Furthermore, NNU-Net produced remarkable average dice scores for the end-systole and end-diastole of the left atrium. These results highlight how well NNU-Net performs precise and accurate segmentation for 2D echocardiography applications.

All things considered, this work advances the field of computer-aided echocardiography diagnosis significantly and opens the door to the development of trustworthy new clinical studies. The findings emphasise how important it is to choose appropriate deep learning architectures, give computational efficiency top priority, and explore different image representations in order to improve model effectiveness and practicality.

Even with the impressive outcomes, there is still space for development to match expert manual annotations, particularly in the areas of left ventricular capacity and ejection fraction estimation. The study shows how 2D echocardiographic image processing can be done accurately and totally automatically, opening the door to improved clinical diagnosis and treatment

VII. FUTURE SCOPE

In order to further automate the processing of 2D echocardiographic pictures and achieve even greater accuracy in segmenting cardiac components such the right ventricle, right atrium, myocardium, chambers, and valves, future research may focus on developing more sophisticated deep learning models. With enhanced automation, specialists may need to exert less manual labour, and sophisticated models can offer instantaneous analysis while performing an echocardiography.

Optimising existing encoder-decoder architectures further to achieve better left ventricular volume and ejection fraction estimation accuracy, aiming to meet or exceed clinical inter-observer assessments' accuracy. This could entail methods like employing loss functions designed specifically for medical image segmentation, hyperparameter adjustment, or attention processes.

Adding more instances and annotations to the CAMUS dataset will aid in the creation of more reliable and broadly applicable models for the analysis of echocardiographic images. A more extensive and varied dataset would facilitate the development of more broadly applicable models, lowering the possibility of overfitting and enhancing performance in various patient demographics and scenario scenarios. An improved dataset might also include three-dimensional echocardiography pictures, which would give training and testing segmentation models a stronger basis.

The findings of this study may open the door for the application of deep learning algorithms in clinical settings, improving the effectiveness and precision of echocardiographic evaluations and maybe resulting in quicker and more accurate cardiac diagnoses.

VIII. REFERENCES

[1] S. Leclerc *et al*., "Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography," in *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198-2210, Sept. 2019, doi: 10.1109/TMI.2019.2900516.

[2] Amer, Alyaa, et al. "ResDUnet: Residual dilated UNet for left ventricle segmentation from echocardiographic images." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.

[3] Carneiro, Gustavo, Jacinto C. Nascimento, and António Freitas. "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods." IEEE Transactions on Image Processing 21.3 (2011): 968-982.

[4] Oktay, Ozan, et al. "Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation." IEEE transactions on medical imaging 37.2 (2017): 384-395.

[5] Xue, Wufeng, et al. "Improved segmentation of echocardiography with orientation-congruency of optical flow and motion-enhanced segmentation." IEEE Journal of Biomedical and Health Informatics 26.12 (2022): 6105-6115.

[6] Awasthi, Navchetan, et al. "LVNet: Lightweight model for left ventricle segmentation for short axis views in echocardiographic imaging." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 69.6 (2022): 2115-2128.

[7] Morais, Pedro, et al. "Fast segmentation of the left atrial appendage in 3-D transesophageal echocardiographic images." IEEE transactions on ultrasonics, ferroelectrics, and frequency control 65.12 (2018): 2332-2342.

[8] Zhang, Jiapeng, et al. "Dual-branch TransV-Net for 3D echocardiography segmentation." IEEE Transactions on Industrial Informatics (2023).

[9] Lin, Zihan, et al. "CLA-U-Net: Convolutional Long-short-term-memory Attention-gated U-Net for Automatic Segmentation of the Left Ventricle in 2-D Echocardiograms." 2022 IEEE International Ultrasonics Symposium (IUS). IEEE, 2022.

[10] Dai, Weihang, et al. "Cyclical self-supervision for semi-supervised ejection fraction prediction from echocardiogram videos." IEEE Transactions on Medical Imaging (2022).

[11] Mortada MJ, Tomassini S, Anbar H, Morettini M, Burattini L, Sbrollini A. Segmentation of Anatomical Structures of the Left Heart from Echocardiographic Images Using Deep Learning. Diagnostics (Basel). 2023 May 9;13(10):1683. doi: 10.3390/diagnostics13101683. PMID: 37238168; PMCID: PMC10217142.

[12] Yang, Tingyang, et al. "Segmentation of five components in four chamber view of fetal echocardiography." 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020.

[13] Østvik, Andreas, et al. "Myocardial function imaging in echocardiography using deep learning." ieee transactions on medical imaging 40.5 (2021): 1340-1351.

[14] Ahn, Shawn S., et al. "Multi-frame attention network for left ventricle segmentation in 3d echocardiography." Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021.

[15] Naghne, R., et al. "An Efficient Capsule-based Network for 2D Left Ventricle Segmentation in Echocardiography Images." 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2023.

[16] Cao, Jiawei. (2021). Research on crack detection of bridge deck based on computer vision. IOP Conference Series: Earth and Environmental Science. 768. 012161. 10.1088/1755-1315/768/1/012161.

[17–New] E. Smistad and F. Lindseth, "Real-time tracking of the left ventricle in 3d ultrasound using kalman filter and mean value coordinates," in Proc. MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), Boston, MIDAS Journal, 2014, pp. 65–72.

[18] Zhou, Guang-Quan, et al. "DSANet: Dual-branch shape-aware network for echocardiography segmentation in apical views." IEEE Journal of Biomedical and Health Informatics (2023).

[19] Sustersic, Tijana, Milos Anic, and Nenad Filipovic. "Heart left ventricle segmentation in ultrasound images using deep learning." 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON). IEEE, 2020.

[20] E. Smistad, A. Østvik, B. O. Haugen and L. Lovstakken, "2D left ventricle segmentation using deep learning," *2017 IEEE International Ultrasonics Symposium (IUS)*, Washington, DC, USA, 2017, pp. 1-1, doi: 10.1109/ULTSYM.2017.8092812.

[21] Y. Guo, Y. Wang, S. Nie, J. Yu and P. Chen, "Automatic Segmentation of a Fetal Echocardiogram

Using Modified Active Appearance Models and Sparse Representation," in *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1121-1133, April 2014, doi: 10.1109/TBME.2013.2295376.

[22] O. S. Al-Kadi, "Spatio-Temporal Segmentation in 3-D Echocardiographic Sequences Using Fractional Brownian Motion," in *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 8, pp. 2286-2296, Aug. 2020, doi: 10.1109/TBME.2019.2958701.

[23] Y. Zhu, C. Xiong, H. Zhao and Y. Yao, "SAM-Att: A Prompt-Free SAM-Related Model With an Attention Module for Automatic Segmentation of the Left Ventricle in Echocardiography," in *IEEE Access*, vol. 12, pp. 50335-50346, 2024, doi: 10.1109/ACCESS.2024.3384383.

[24] M. C. Brindise, B. A. Meyers, S. Kutty and P. P. Vlachos, "Automated Peak Prominence-Based Iterative Dijkstra's Algorithm for Segmentation of B-Mode Echocardiograms," in *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 5, pp. 1595-1607, May 2022, doi: 10.1109/TBME.2021.3123612.

[25] S. Leclerc *et al*., "LU-Net: A Multistage Attention Network to Improve the Robustness of Segmentation of Left Ventricular Structures in 2-D Echocardiography," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2519-2530, Dec. 2020, doi: 10.1109/TUFFC.2020.3003403.

[26] W. -Y. Hsu, "Automatic Left Ventricle Recognition, Segmentation and Tracking in Cardiac Ultrasound Image Sequences," in *IEEE Access*, vol. 7, pp. 140524-140533, 2019, doi: 10.1109/ACCESS.2019.2920957.

[27] G. Zamzmi, L. -Y. Hsu, W. Li, V. Sachdev and S. Antani, "Harnessing Machine Intelligence in Automatic Echocardiogram Analysis: Current Status, Limitations, and Future Directions," in *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 181-203, 2021, doi: 10.1109/RBME.2020.2988295.

[28] M. H. Jafari et al., “Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training,” Int. J. Comput. Assist. Radiol. Surg., vol. 14, no. 6, pp. 1027–1037, 2019.

[29] Moradi S, Oghli MG, Alizadehasl A, Shiri I, Oveisi N, Oveisi M, Maleki M, Dhooge J. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Phys Med. 2019 Nov;67:58-69. doi: 10.1016/j.ejmp.2019.10.001. Epub 2019 Oct 28. PMID: 31671333.

[30] Ondersma SJ, Martino S, Svikis DS, Yonkers KA. Commentary on Kim et al. (2017): Staying focused on non-treatment seekers. Addiction. 2017 May;112(5):828-829. doi: 10.1111/add.13736. PMID: 28378329; PMCID: PMC6552680.

[31] Monkam, Patrice, et al. "A Disentanglement and Fusion Data Augmentation Approach for Echocardiography Segmentation." 2022 IEEE International Ultrasonics Symposium (IUS). IEEE, 2022.

[32] T. Sustersic, M. Anic and N. Filipovic, "Heart left ventricle segmentation in ultrasound images using deep learning," *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*, Palermo, Italy, 2020, pp. 321-324, doi: 10.1109/MELECON48756.2020.9140527.