# End-to-End Echocardiographic Analysis Pipeline: AI-Powered Assessment and Reporting of Heart Function

Rana Hany*, Sarah Ibrahim*, Sarah Mohammed*, Luna Eyad*, Yasmin Tarek*, PhD.Ahmed Fahmy*, PhD.Tamer Basha*

*Department of Systems and Biomedical Engineering, Cairo University, Giza, Egypt
rana.hany102@gmail.com, sarahibrahim15156@gmail.com, sarahragaei06@gmail.com, lunaeyad02@gmail.com, yasmin.elgamal02@gmail.com

***Abstract*—An echocardiogram is a sophisticated ultrasound imaging technique used to diagnose heart conditions. Cardiovascular diseases (CVDs) continue to be a leading cause of death worldwide. Additionally, interpreting echocardiogram images manually is a complex and time-intensive task that requires a high level of expertise. In this study, we present an end-to-end echocardiographic analysis pipeline that automates the entire process, including cardiac segmentation, clinical parameters calculation, and structured report generation, all integrated within an interactive web application. A variety of deep learning methods were employed for left ventricle (LV) segmentation, including supervised models trained on the CAMUS dataset and self-supervised models such as SAM and MedSAM. Ejection fraction (EF) was estimated using the Area-Length method, while large language models (LLMs) were used to generate medical impressions based on calculated parameters. The overall goal of the system is to help cardiologists in automating and analyzing the echocardiographic workflow.**

***Index Terms*—Echocardiography, web application, deep learning, segmentation, left ventricle, supervised learning, self-supervised learning, ejection fraction, large language models, reporting**

## I. Introduction

CVDs are the leading cause of death globally, according to World Health Organization (WHO). An estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke, making an early and accurate diagnosis essential for effective treatment [1].

Medical imaging plays a crucial role in cardiac assessment, with echocardiography standing out as a fast, cost-effective, and non-invasive alternative to computed tomography (CT) and magnetic resonance imaging (MRI), which are often expensive, time-consuming, and less accessible. Despite its advantages, analyzing echocardiogram images manually is a time-consuming process that relies on expert cardiologists.

Existing AI-based echocardiography products are prevalent in regions such as Europe and North America, often coming at a high cost per scan. This reveals a notable market gap in the Middle East, particularly in Egypt.

To address this challenge, we propose a comprehensive, end-to-end echocardiographic analysis pipeline that automates the process, starting with raw echocardiographic images and ending with a structured report generation. Our web-based application enables cardiologists to upload images for automatic segmentation to identify the blood pool, myocardium, and atrium within the heart chambers, followed by the calculation of key parameters such as ejection fraction (EF), cardiac blood volume, myocardium volume, and stroke volume. These parameters are then fed into a large language model (LLM) to generate medical impressions depending on the previous data. Finally, a structured report is generated according to international standards.

## II. Literature Review

### A. Pre-Deep Learning Era

In 1978, the early automation of echocardiography analysis was demonstrated by Chu and Raeside [2] through the use of Fourier analysis for evaluating mitral valve leaflets using M-mode ultrasound. This period highlights the beginning of computational methods in echocardiography.

Till 2019, only one echocardiographic datasets—CETUS—had been broadly validated [3]. However, no challenger implemented a deep neural network, only mathematical, statistical, and machine learning methods, due to the limited dataset size—only 3D 45 sequences—which are insufficient for training a deep neural network without the risk of overfitting.

### B. Deep Learning Era

The use of deep learning (DL) in echocardiography began with Carneiro *et al.* in 2012 [4], who applied a Deep Belief Network (DBN) within a maximum a posteriori (MAP) framework to segment the left ventricle in 2D echocardiographic images.

In 2019, the release of two large-scale datasets—CAMUS [5] and EchoNet-Dynamic [6]—enabled broader adoption of DL approaches in cardiac ultrasound analysis. Leclerc *et al.* [5] evaluated several architectures, including ACNN, SHG, U-Net, and U-Net++, with U-Net consistently achieving the best segmentation performance in terms of accuracy and robustness. In 2024, Sanjeevi *et al.* [7] published a comprehensive review for echocardiogram analysis using DL, presenting open research challenges in automated Echocardiogram analysis using DL, including: lightweight model development, end-to-end cardiac analysis, including cardiac chambers and valves, in a single clinical tool, would be highly beneficial, exploring unsupervised learning and transfer Learning Model.

## III. Methodology

The proposed system follows a multi-stage pipeline as illustrated in "Fig. 1":

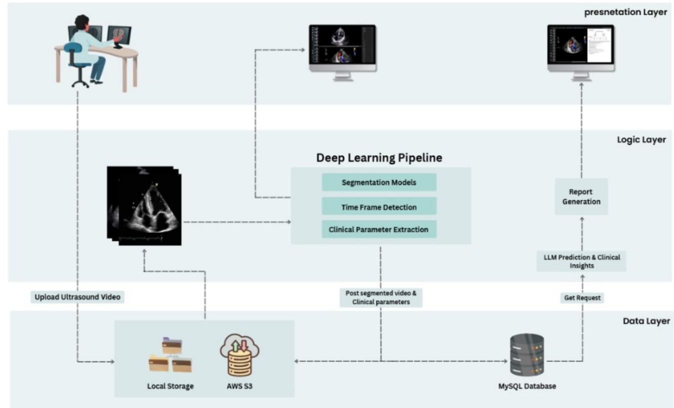



Fig. 1: System workflow.

### A. Data Preprocessing

*1) CAMUS:* It consists of 2D apical four-chamber view acquired from 500 patients, each patient has End Diastolic (ED) and End Systolic (ES) frames, including images of varying qualities (poor, medium, and good), and is manually annotated by three cardiologists. The applied preprocessing steps included:

*a) Resizing:* All images are converted from NIfTI (.nii) format to PNG format for faster processing, and resized to $256\times256$ pixels to standardize input dimensions.

*b) Data Splitting:* The dataset is divided into 80% for training, 10% for validation, and 10% for testing.

*c) Data Augmentation:* To enhance model generalization, augmentation is applied, including rotation by 5 degrees, translation by (0.05, 0.05), and scaling within (0.8, 1.2).

*2) EchoNet-Dynamic:* It consists of 10,030 apical four-chamber labeled echocardiogram videos. The applied preprocessing steps included:

*a) Frames Extraction:* ED and ES frames were extracted and identified from each video using the coordinates in the CSV file, designating the frame with the larger left ventricular (LV) area as ED and the smaller as ES.

*b) Ground Truth Generation:* We performed complete filling of the traced left ventricle (LV) to obtain full annotations used as ground truth.

### B. Cardiac Structure Segmentation

The objective is to analyze how different segmentation methods influence cardiac segmentation, including left ventricle endocardium ($LV_{Endo}$), the myocardium ($LV_{Epi}$) and the left atrium (LA).

*1) Supervised Learning Methods:*

*a) U-Net from Scratch:* Building on previous work, our goal is to develop a lightweight architecture by reducing the number of trainable parameters, computational complexity (GFLOPs), and training time per epoch.

*b) Pre-Trained U-Net:* We transitioned from a lightweight U-Net to a pre-trained U-Net with a ConvNeXt backbone for better generalization. This shift allows us to evaluate the trade-off between computational cost and model performance.

*c) Testing on EchoNet-Dynamic Dataset:* To evaluate the zero-shot generalization of our pre-trained U-Net, we applied it directly to the EchoNet-Dynamic test set without any fine-tuning. The model, originally trained on the CAMUS dataset, was applied to 1,277 test videos (2,552 frames in total). Since EchoNet-Dynamic provides ground truth only for the LV cavity, we treated all non-LV classes—myocardium and left atrium (LA)—as background during both inference and evaluation. This effectively reduced the task to binary segmentation: LV vs. background.

*2) Self-Supervised Learning Methods:* The Segment Anything Model (SAM) and its medical counterpart, MedSAM, represent significant advances in image segmentation for their ability to generalize to images without requiring labeled data.

Our primary goal was to explore the zero-shot and fine tuning capabilities of SAM and MedSAM, and to evaluate their performance relative to our trained U-Net model. We investigated the impact of different prompt types on segmentation accuracy.

### C. Clinical Parameter Estimation

Ejection fraction (EF) is a critical metric for assessing left ventricular function, defined as the percentage of blood ejected from the ventricle relative to its end-diastolic volume (EDV):

$$\mathrm{EF} = \frac{\mathrm{EDV} - \mathrm{ESV}}{\mathrm{EDV}} \times 100$$

where ESV is the end-systolic volume. Stroke volume (SV), defined as $\mathrm{SV} = \mathrm{EDV} - \mathrm{ESV}$, myocardial volume (MV), and blood volume (BV, assumed equal to EDV) are also key indicators of cardiac health.

EF is used to classify heart failure as follows: heart failure with reduced EF (HFrEF, $< 40\%$), heart failure with mildly reduced EF (HFmrEF, $40\% \leq \mathrm{EF} < 50\%$), and heart failure with preserved EF (HFpEF, $\geq 50\%$).

The Area-Length method, implemented in this study, is simpler and relies solely on A4C views, but potentially less precise due to its assumption of an ellipsoidal LV shape, as described by Leclerc et al. (2019) [5].

## D. LLM Evaluation for Clinical Impression Generation

To identify the optimal large language model (LLM), we evaluated freely available models based on their ability to generate accurate clinical impressions.

Four representative cases were adapted from real-world clinical sources: Case 1 (new-onset HFrEF) from *Patient Management in the Telemetry/Cardiac Step-Down Unit* [8]; Case 2 (HFpEF in an elderly female) from the American College of Cardiology (ACC) Case Quizzes [9]; and Cases 3 and 4 (HFpEF due to amyloidosis and advanced HFrEF, respectively) from the Metro North Health Heart Failure Case Study Handbook [10]. Each case comprised structured patient data—including demographics, medical history, imaging, and laboratory findings—paired with a concise reference clinical impression.

We queried free-access LLMs that support text-based input and output. Most models were evaluated using the OpenRouter API, with a system instruction—"You are a cardiologist generating concise, accurate clinical impressions based on patient case data"—followed by the structured case input. **Gemini 2.5 Flash** was accessed via the Google Generative AI API. The generated outputs were compared to the reference impressions using Sentence-BERT (SBERT) cosine similarity to quantify semantic similarity.

## E. Web Application Integration

*1) System Design and Development Approach:* To ensure accessibility and scalability, we developed the solution as a web application, accessible from any device with an internet connection—no additional software required. The platform is OS-independent, browser-compatible, and supports remote access, easy updates, and lower maintenance costs.

Following the Software Development Life Cycle (SDLC), we began with requirement analysis, preparing the SRS document, designing the database (ERD), and planning architecture using the C4 model. The user interface was designed in Figma for usability.

The frontend was built using Angular and Tailwind CSS, while the backend used Node.js with Express. A MySQL database managed structured data, and echocardiographic images were stored on AWS S3 using the AWS SDK. Frontend-backend communication used RESTful APIs, and the AI model was integrated via FastAPI and Axios.

Jira was used for sprint-based task management, supporting iterative development with regular feedback and testing.

*2) Features:* The system offers a set of features that assist cardiologists in analyzing echocardiography images and generating reports more efficiently. The main features include:

- **Authentication & Access Control** Restricts access to authorized cardiologists through a secure login system.
- **Cloud-Based Storage** Stores reports, images, and patient data securely on AWS S3 cloud.
- **AI-Based Segmentation** Automatically segments echocardiography images to highlight heart structures using the AI model.
- **Quantitative Clinical Parameter Calculation** Calculates important values such as EF, stroke volume, EDV, ESV, and other key echocardiographic parameters.
- **Automated Report Generation** Creates a draft report combining AI-generated impressions and measurements with patient data.
- **Manual Editing and Report Management** Enables cardiologists to edit the report content to finalize it, delete the report, and export it as a PDF.

# IV. Results

## A. Segmentation Performance Evaluation

To evaluate the segmentation output, we used the Dice Similarity Coefficient (DSC), which measures the overlap between the prediction and ground truth images, Intersection over Union (IoU), Average Symmetric Surface Distance (ASSD), and Hausdorff Distance (HD). The lightweight U-Net model with 1.9M parameters, which is optimized for speed, performs well on good-quality images, while the pre-trained U-Net model with 93M parameters achieves the best DSC and is optimized for performance and generalization across various image qualities.

Testing on the EchoNet-Dynamic dataset using a pre-trained U-Net model demonstrated strong generalization across all frames and robust zero-shot performance on an unseen domain. Performance was stronger at ED due to clearer LV boundaries. Higher ASSD and HD at ES reflect the difficulty in segmenting the smaller LV during contraction, as shown in Table I.

Beyond supervised learning methods, SAM and MedSAM demonstrated promising zero-shot segmentation performance without requiring labeled data. Additionally, given that image annotation requires expert cardiologists, fine-tuning SAM on only 100 labeled images reduced the training time significantly—just 8 minutes for 10 epochs—compared to 97 minutes for 150 epochs using U-Net, as shown in Table II.

For qualitative analysis, "Fig. 2" demonstrates the segmented output across the models on a sample from CAMUS dataset.

## B. Performance Analysis: EF Estimation

The Area-Length method predicted EF with 80.0% of values within ±10% of the ground truth. The mean EF difference was 6.9%, with a slight overestimation tendency (48.3% predicted vs. 44.7% ground truth), mainly due to segmentation variability and the simplified ellipsoidal heart model.

## C. LLM Evaluation for Clinical Impression Generation

Only free models with complete results (four SBERT scores across all cases) were evaluated, namely **Gemini 2.5 Flash** (via Google API free tier) and **deepseek/deepseek-r1-0528-qwen-8b:free** (via OpenRouter). For comparison, the non-free **liquid/lfm-3b** was included in Case 3. The best-performing models for each case were:

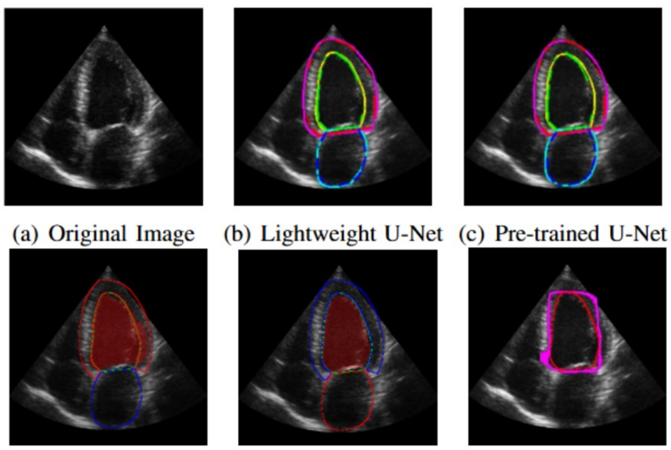
- Case 1: **Gemini 2.5 Flash** (SBERT = 0.6354)
- Case 2: **Gemini 2.5 Flash** (SBERT = 0.7742)

TABLE I: Evaluation Metrics (DCI, IoU, HD, ASSD) for LV Endo and LV Epi at ED and ES Frames

| Methods | ED | | | | | | | | ES | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $LV_{Endo}$ | | | | $LV_{Epi}$ | | | | $LV_{Endo}$ | | | | $LV_{Epi}$ | | | |
| | **DCI** | **IoU** | **HD** | **ASSD** | **DCI** | **IoU** | **HD** | **ASSD** | **DCI** | **IoU** | **HD** | **ASSD** | **DCI** | **IoU** | **HD** | **ASSD** |
| Lightweight U-Net | 0.94 | 0.89 | 8.8 | 2.09 | 0.86 | 0.75 | 10.5 | 2.10 | 0.92 | 0.85 | 14.8 | 2.81 | 0.86 | 0.76 | 13.8 | 2.34 |
| **Pre-trained U-Net** | **0.95** | **0.91** | **5.7** | **1.70** | **0.89** | **0.80** | **6.3** | **1.70** | **0.94** | **0.89** | **5.6** | **1.68** | **0.89** | **0.80** | **6.2** | **1.84** |
| Echo-Net Testing | 0.84 | 0.74 | 19.3 | 7.58 | – | – | – | – | 0.87 | 0.78 | 15.2 | 6.26 | – | – | – | – |

TABLE II: DCI for Foundation Models Versus U-Net Across the CAMUS Testing Dataset

| Methods | Prompt Type | DCI |
|---|---|---|
| SAM | Bounding Box Prompt | 0.59 |
| | Point Prompt | 0.69 |
| MedSAM | Bounding Box Prompt | 0.67 |
| | Point Prompt | 0.80 |
| Fine-tuned SAM | Bounding Box Prompt | 0.86 |
| Pre-trained U-Net | | 0.93 |



(a) Original Image (b) Lightweight U-Net (c) Pre-trained U-Net

(d) SAM (e) MedSAM (f) Fine-tuned SAM

Fig. 2: Segmentation outputs across different models.

- Case 3: **liquid/lfm-3b** (SBERT = 0.7183)
- Case 4: **Gemini 2.5 Flash** (SBERT = 0.6806)

**Gemini 2.5 Flash** was selected for deployment within the application. It achieved an average SBERT score of 0.6733 across all four cases. In comparison, **deepseek/deepseek-r1-0528-qwen-8b:free** achieved an average SBERT of 0.5979. The non-free **liquid/lfm-3b** (average SBERT = 0.6748) was included for comparison in Case 3 but was not considered for deployment due to its cost. Although **liquid/lfm-3b**'s average SBERT was comparable, its non-free status and lower performance in Cases 2 and 4 made **Gemini 2.5 Flash** the preferred choice.

Moderate SBERT scores (0.5817–0.7742) suggest that fine-tuning could enhance performance. However, the lack of publicly available echocardiography datasets with expert clinical impressions presents a key limitation. Developing such resources remains essential for improving automated clinical reporting.

## V. DISCUSSION

This study demonstrated the effectiveness of supervised and self-supervised learning methods for automated echocardiographic segmentation. The pre-trained U-Net model achieved the highest segmentation Dice score, while SAM and MedSAM showed promising zero-shot performance. Fine-tuning SAM on a small labeled set further gave reasonable results. Additionally, the Area-Length method provided reliable ejection fraction estimates, and the integration of LLMs enabled automated, interpretable report generation.

## VI. CONCLUSION

The system helps cardiologists by automating the complex echocardiographic analysis process through uploading the images by the cardiologist to the website, then providing segmentation for cardiac structures, calculating clinical parameters, and generating a structured report.


## ACKNOWLEDGMENT

Special thanks are extended to Eng. Ahmed Sharshar, and Eng. Nada Mansour for their valuable guidance and technical assistance throughout the development of this project.


## REFERENCES


[1] World Health Organization, “Cardiovascular diseases (CVDs),” 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] C. H. Chu and E. L. Raeside, “Fourier analysis of M-mode echocardiograms,” Ultrasound in Medicine & Biology, vol. 4, pp. 303–308, 1978.

[3] O. Bernard et al., “Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography,” in Proc. MICCAI Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS), 2014.

[4] G. Carneiro et al., “Segmentation of the left ventricle of the heart in 2D ultrasound images using deep belief networks,” Medical Image Analysis, vol. 16, no. 6, pp. 1197–1206, 2012.

[5] S. Leclerc et al., “Deep learning for segmentation using an open large-scale dataset in 2D echocardiography,” IEEE Transactions on Medical Imaging, vol. 38, no. 9, pp. 2198–2210, 2019.

[6] D. O. Ouyang et al., “EchoNet-Dynamic: A large-scale dataset and deep learning benchmark for echocardiographic analysis,” Nature Communications, vol. 11, no. 1, pp. 1–14, 2020.

[7] N. S. Sanjeevi et al., “Deep learning for echocardiogram analysis: A comprehensive review,” Artificial Intelligence in Medicine, vol. 148, 2024.

[8] E. Herzog, Patient Management in the Telemetry/Cardiac Step-Down Unit: A Case-Based Approach, 2020

[9] American College of Cardiology, “Case Quizzes: Heart Failure with Preserved Ejection Fraction,” [Online].Available: https://www.acc.org/Education-and-Meetings/Patient-Case-Quizzes/A-Case-of-an-Elderly-Patient-With-HFpEF

[10] Metro North Health, Heart Failure Case Study Handbook [Online]. Available :https://metronorth.health.qld.gov.au/wp-content/uploads/2019/09/hf-cs.pdf