CardioSmartAssist: A customisable AI framework for echocardiography-based cardiac assessment

Biomedical Signal Processing and Control 122 (2026) 110363

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

CardioSmartAssist: A customisable AI framework for echocardiography-based cardiac assessment

Francesca Giada Antonaci [a,*], Piera Ciaramella [c,1], Giorgia Marullo [b], Luca Ulrich [b], Vincenza Papa [d,1], Walter Grosso Marra [e,1], Alessandro Depaoli [e,1], Riccardo Miraglia [b], Sandro Moos [b], Enrico Vezzetti [b]

[a] Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin (TO), 10129, Italy
[b] Department of Management and Production Engineering (DIGEP), Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin (TO), 10129, Italy
[c] Università degli studi di Roma La Sapienza policlinico Umberto I, Viale del Policlinico, 155, Roma (RM), 00161, Italy
[d] ASL Benevento, Via Oderisio da Benevento, 1, Benevento (BN), 82100, Italy
[e] Ospedale di Ivrea, Piazza Credenza, 2, Ivrea (TO), 10015, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Left ventricular ejection fraction
Computer-aided diagnosis
Clinical decision support systems
AI-assisted diagnosis

ABSTRACT

Background and Objective: Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for approximately 20.5 million deaths annually, nearly one-third of all global deaths. Despite its central role in cardiology, echocardiographic assessment remains subject to significant inter- and intra-observer variability, particularly based on manual frame selection and segmentation. This limitation has driven increasing interest in Deep Learning (DL) solutions capable of enabling more objective, reproducible, and efficient analyses.

Methods: This paper introduces CardioSmartAssist, a deep learning framework for automatic left ventricular segmentation and multi-cycle ejection fraction estimation relying solely on echocardiographic videos. The system integrates frame-by-frame segmentation visualisation, anomaly detection, and volume tracking to enhance clinical usability. Moreover, a key feature is its continuous learning mechanism, which allows clinician-corrected segmentations to be stored and used for progressive model refinement.

Results: The framework is based on a MultiResUNet architecture, trained on public (EchoNet-Dynamic) and proprietary (CardioSmartSet) datasets, achieving Dice Coefficient Scores of 0.9328 and 0.9189, respectively. On the held-out test set, the EF estimated by the system showed a mean absolute difference of 10% compared with clinically reported EF values, which is lower than the typical inter-operator variability of approximately 13%.

Conclusion: CardioSmartAssist resulted to be a promising tool for consistent cardiac evaluations, improving access to diagnostics, and enhancing clinical decision-making through smart assistance.

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of morbidity and mortality worldwide, with an estimated 4.2 million deaths in Europe, accounting for 42.5% of all deaths [1]. For this reason, accurate evaluation of cardiac function is essential for diagnosing cardiovascular diseases, screening for cardiotoxicity, and determining appropriate therapeutic interventions for critically ill patients [2].

Among the available imaging modalities, echocardiography is the most commonly employed technique in clinical practice due to its portability, cost-effectiveness, and real-time imaging capabilities [3]. This modality utilises ultrasound technology to capture dynamic spatiotemporal sequences of the heart and surrounding structures, making it an indispensable tool in cardiovascular medicine [4]. Furthermore, echocardiography is particularly valued for its rapid data acquisition, non-invasive nature, and absence of ionising radiation [2]. A critical aspect of echocardiographic assessment is the interpretation of left ventricular ejection fraction (LVEF), which serves as a fundamental indicator of cardiac function [3]. LVEF is calculated as the ratio of the change in left ventricular end-systolic and end-diastolic volumes and holds a fundamental position in determining eligibility for life-prolonging treatments [2].

* Corresponding author.
E-mail address: francesca.antonaci@polito.it (F.G. Antonaci).
[1] M.D.

https://doi.org/10.1016/j.bspc.2026.110363
Received 26 January 2026; Received in revised form 16 March 2026; Accepted 16 April 2026
Available online 24 April 2026

LVEF has emerged as a key prognostic marker in conditions such as myocardial infarction, heart failure, and valvular disease [5]. Current clinical guidelines incorporate LVEF thresholds to inform critical decisions, such as recommending valve replacement for severe valvular heart disease or implanting cardioverter-defibrillators (ICDs) in patients with heart failure and an ejection fraction of 35% or less, where ICDs have been shown to improve survival when combined with optimal medical therapy [6].

Despite its clinical significance, the manual assessment of echocardiograms is highly dependent on operator expertise and is susceptible to significant inter- and intra-observer variability. Manual ventricular size tracking is labour-intensive and prone to errors and inconsistencies, particularly in the presence of irregular heart rhythms [3]. Moreover, although guidelines from the European Association of Cardiovascular Imaging and the American Society of Echocardiography recommend averaging multiple cardiac cycles to mitigate variability, many assessments rely on single-cycle tracings or visual estimates, leading to inter-observer variability of 7.6% to 13.9% [2]. Additionally, echocardiographic quality can be compromised by factors such as weak acoustic windows and suboptimal measurement techniques [7].

In the intensive care unit, visual estimation of cardiac function from echocardiographic images may be essential due to time constraints, and the COVID-19 pandemic has highlighted the need for solutions that require non-specialist physicians to perform echocardiographic assessments in infection control environments [7].

To address these challenges, efforts have been made to standardise LVEF measurements by integrating Artificial Intelligence (AI)-based solutions [7]. Recent advancements in AI and Deep Learning (DL) have demonstrated significant potential in overcoming the limitations of human interpretation [8], enabling accurate beat-to-beat quantification without the need for manual segmentation [2], enhancing the assessment of various cardiac conditions, including valvular heart disease and ischaemic heart disease [9].

CardioSmartAssist has been designed to address these gaps by introducing an integrated pipeline that mitigates operator variability, enables physiologically consistent multi-cycle analysis, and enhances robustness across acquisition domains.

The intended clinical use case of CardioSmartAssist is to support routine echocardiographic quantification of left ventricular function and point-of-care cardiac assessment in settings where rapid and reproducible measurements are required. In particular, the system is designed to assist clinicians in estimating left ventricular volumes and ejection fraction during routine outpatient examinations, emergency department evaluations, and point-of-care ultrasound assessments performed by non-expert operators. These scenarios may include emergency triage situations with high patient inflow or domiciliary assessments where rapid cardiac function estimation is required but access to expert echocardiographers may be limited.

The primary clinical decision points supported by this workflow include the identification of clinically relevant ejection fraction thresholds used in current guidelines, such as the detection of left ventricular systolic dysfunction (LVEF $< 40\%$) and severe dysfunction (LVEF $\leq 35\%$), which may guide therapeutic decisions including pharmacological therapy optimisation or eligibility for implantable cardioverter-defibrillators.

A significant part of this work is the patient-centred mapping of the entire workflow, designed to fit separate DL tasks into the collaborative framework of clinical care. The framework uses the Unified Modelling Language (UML) to ensure that experts review algorithmic outputs, such as segmentation masks, before they are added to the medical record. The framework also puts the diagnosis in the context of the person's physiology by using patient-specific biometrics to normalise Body Surface Area (BSA). This architecture enables an expert-in-the-loop system, in which manual refinements can quickly correct anatomical errors and are systematically recorded to retrain the model using a continuous learning methodology.

To identify the most suitable methodology for the task, the study compares three DL architectures for image semantic segmentation: MultiResUNet, You Only Look Once (YOLO), and Segment Anything Model (SAM) in terms of segmentation performance and clinical applicability.

Each DL model has been fine-tuned on real-world clinical data, ensuring its performance aligns with current clinical practice standards at high resolution. The selected model achieved satisfactory results on a proprietary dataset comprising only 27 patients, of whom 5 had confirmed altered cardiac conditions. These results provide initial evidence that the approach can operate effectively even in data-constrained scenarios.

This paper is structured as follows. The Methods section details the acquisition and preprocessing of the dataset, followed by a brief discussion of DL-based segmentation approaches, the computation of the Ejection Fraction (EF) methodology, and the design of the proposed solution using UML and the development of the Graphical User Interface (GUI). The Results and Discussion section evaluates system performance, including comparative analyses of segmentation accuracy, ejection fraction calculations, and assessments of end-systolic and end-diastolic volumes. Finally, the Limits and Further Development and Conclusion sections summarise the main findings and outline potential future improvements and research directions.

2. Related works

This section reviews the main research directions that form the methodological foundation of CardioSmartAssist. It first discusses deep learning approaches for automated left ventricular segmentation, followed by advances in video-based models for direct and multi-cycle estimation of ejection fraction. The discussion then broadens to include end-to-end architectures in medical imaging and the development of AI-assisted clinical workflows, with particular emphasis on physician-centred decision support through expert-in-the-loop paradigms. Overall, the literature indicates that many existing solutions address isolated computational tasks, often lacking seamless integration into real clinical workflows. This analysis highlights the translational gap between algorithmic performance and practical deployment, positioning CardioSmartAssist as a holistic framework designed for clinical integration and continuous learning.

The accurate segmentation of the left ventricle represents a foundational step in the automated analysis of echocardiographic images, enabling the extraction of key cardiac measurements such as ventricular volumes and ejection fraction [10–12]. To overcome the inherent inter- and intra-observer variability associated with manual boundary tracing, the field has increasingly adopted deep learning approaches, predominantly leveraging Convolutional Neural Networks and encoder–decoder architectures [11]. In particular, U-Net and its variants have emerged as the cornerstone for cardiac chamber delineation, demonstrating strong performance in accurately segmenting the LV endocardium and epicardium across various standard apical views [11]). These robust segmentation models have been successfully integrated as core modules within broader automated pipelines. The extracted spatial features directly support downstream clinical tasks such as automated volume quantification, longitudinal strain calculation, and disease detection [10,12]. Recent trends have further refined these capabilities by incorporating multi-stage attention mechanisms to precisely localise the LV and reduce geometric outliers, as well as employing frame-level weak supervision and spatiotemporal convolutions to track ventricular boundaries continuously across the entire cardiac cycle [2]. Despite these significant methodological advancements, many existing approaches still treat ventricular segmentation as an isolated task rather than a component of a cohesive clinical tool.

Building upon advancements in chamber segmentation, the automated estimation of left ventricular ejection fraction has undergone a significant evolution, evolving from static frame-based calculations to dynamic video-based analysis [2,13]. Early computational methods

typically required the manual or semi-automated isolation of representative end-systolic and end-diastolic frames, a process that inherently neglects the rich temporal information and dynamic motion embedded across the full cardiac cycle [2,13]. More recently, research has moved toward continuous video processing using spatiotemporal convolutions and transformer-based architectures. These models can analyse echocardiographic sequences of arbitrary length [2,13]. Propelled by the introduction of large-scale, thoroughly annotated echocardiography datasets, these video-based deep learning models can aggregate functional data across multiple beats, thereby mitigating the impact of irregular heart rhythms and isolated acquisition artefacts [2]. In parallel, several researchers have successfully developed volume-independent, direct regression models that estimate EF by mimicking the holistic visual assessment of an expert clinician, completely bypassing the need for explicit endocardial border tracing [14]. While these direct EF estimation systems achieve impressive accuracy and reduce processing time, they frequently operate as black-box prediction models that output a single functional metric without providing interpretable, intermediate anatomical boundaries. Treating EF estimation as an isolated computational task, decoupled from the collaborative clinical workflow, ultimately hinders physician oversight, anomaly detection, and clinical trust.

The paradigm of medical image analysis is increasingly shifting toward end-to-end deep learning systems that unify complex feature extraction and diagnostic prediction into a single, cohesive pipeline [15]. Recent advancements illustrate the efficacy of data-driven fusion frameworks, which dynamically integrate multi-scale image resolutions and hierarchical feature representations to capture intricate anatomical patterns without relying on hand-crafted heuristics [15]. By employing specialised multi-route architectures and integrated feature fusion strategies, these end-to-end models demonstrate significant robustness to real-world imaging variability, seamlessly managing inherent clinical artefacts, illumination changes, and differing acquisition protocols with minimal requisite preprocessing [15,16]. Furthermore, the strategic implementation of advanced regularisation techniques and comprehensive structural designs enables these networks to generalise effectively, even when confronted with highly heterogeneous imaging, skewed class distributions, or limited dataset sizes.

Beyond isolated computational tasks, the successful clinical translation of artificial intelligence in echocardiography and medical imaging necessitates the development of comprehensive, physician-guided workflows. To streamline the diagnostic process directly from the point of care, recent innovations have introduced real-time, deep learning-based guidance systems that assist operators in acquiring diagnostically acceptable images through active, prescriptive probe manipulation cues [17]. Once the imaging data is acquired, integrated pipelines can automatically classify standard views, segment relevant anatomical structures, and extract critical functional parameters, functioning as real-time decision-support tools that reduce operator workload and mitigate subjective diagnostic variability [10,18]. The clinical viability of these automated systems has been increasingly validated in real-world settings, including randomised trials demonstrating that AI-assisted preliminary assessments improve diagnostic consistency and significantly reduce the time required for subsequent expert interpretation [18]. However, practical clinical deployment requires addressing significant ongoing challenges, such as ensuring seamless interoperability with existing electronic health records, preventing algorithmic bias, and maintaining robust performance across highly different clinical datasets equipment and patient populations [18,19].

While previous studies have demonstrated substantial progress in isolated tasks, such as automated left ventricular segmentation or direct ejection fraction estimation, they frequently lack the cohesive integration necessary for real-world clinical deployment. To bridge this translational gap, CardioSmartAssist is proposed as a deployment-oriented framework that unifies automated segmentation and a physiologically consistent multi-cycle EF aggregation strategy directly from echocardiographic videos within an interactive clinical workflow. To ensure clinical trust and diagnostic safety, the system incorporates a structured expert-in-the-loop mechanism that keeps physicians actively involved in the evaluation process. Clinicians can seamlessly review, validate, and manually correct the AI-generated anatomical masks, establishing an expert correction loop within the diagnostic workflow before results are incorporated into the medical record. These manual refinements are systematically captured, versioned, and incorporated into a continuous learning pipeline. This ongoing feedback loop facilitates progressive model refinement and dataset expansion, allowing the system to adapt dynamically to the high-resolution, pathological heterogeneity encountered in diverse clinical environments. By intertwining robust deep learning with clinician-guided refinement, CardioSmartAssist transitions AI from a static computational benchmark to an adaptable clinical decision-support system, ultimately enhancing diagnostic precision and reducing inter-operator variability while strictly maintaining physician oversight.

3. Methods

This section details the materials and methods employed in developing the proposed framework, which is structured to reflect the key phases of the study. The methodology pipeline is presented, as illustrated in Fig. 1, mirroring the sequential steps undertaken in the workflow: (1) dataset acquisition and preprocessing, (2) the DL model architecture used for LV segmentation, (3) the technique for selecting end-systolic and end-diastolic frames, (4) the approach for ventricular volume assessment, and (5) the design and implementation of the application, including its underlying system architecture (modelled through UML diagrams) and the GUI.

The computational resources and software environment are described in Section 2 of the Supplementary Material.

3.1. Data acquisition and preprocessing

Two datasets were utilised in this study: the publicly available *EchoNet-Dynamic* dataset [20] for training the DL model and a custom proprietary dataset, referred to by the authors as the CardioSmartSet dataset, to fine-tune the previously trained DL model. This dataset was collected from clinical sources with informed consent, including private studies and clinics, and has been anonymised by authorised personnel with access to the data.

Ethics and governance. The EchoNet-Dynamic dataset used in this study is publicly available [20]; thus, no additional approval was required for its use.

Data included in the CardioSmartSet dataset were retrospectively collected during routine clinical examinations and fully anonymised prior to being provided to the research team. The study was conducted in accordance with the principles of the Declaration of Helsinki. No identifiable patient information was accessible to the authors at any stage of the study.

3.1.1. EchoNet-dynamic dataset

The *EchoNet-Dynamic* dataset is a benchmark for echocardiographic analysis, comprising 10030 echocardiographic video sequences acquired using apical four-chamber 2D echocardiography. Videos were recorded at an average of 51 frames per second with a resolution of 112 × 112 pixels. Each video incorporates expert-annotated ground truth data, including LV boundaries, systolic and diastolic frames, and EF values calculated using certified medical equipment. The standardised process of segmenting significant frames using a certified echograph will be called "manual segmentation" in this paper. The dataset was sourced from the EchoNet website [20], and the split into training, validation, and test sets followed the original division provided by the EchoNet-Dynamic dataset.

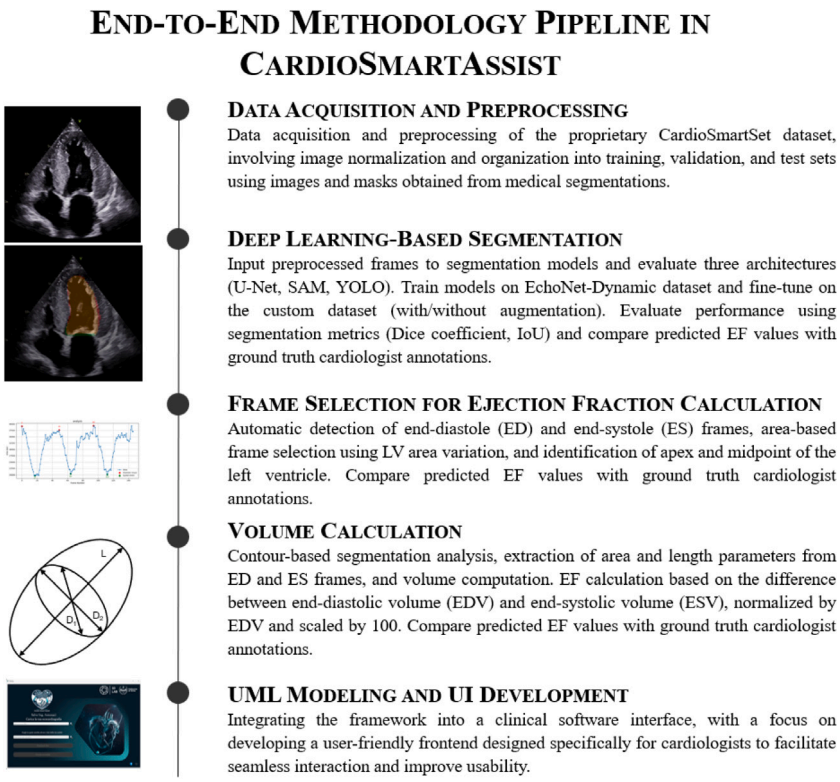

Fig. 1. Methodology pipeline. Illustrative examples (left column) corresponding to the stages of the end-to-end pipeline starting from the top: original A4C frame, LV segmentation mask overlaid on the frame, area-frame curve with ED/ES peak detection, geometric Area-Length model, prototype GUI of CardioSmartAssist.

In addition to the version of the EchoNet-Dynamic dataset distributed from the official website [20], a version of the EchoNet dataset used in this study, where appropriate, is available on Kaggle [21]. This includes a version in which the systolic and diastolic keyframes, along with their corresponding binary masks, have been extracted. The dataset is organised into three main directories: training, validation, and test, with each directory containing subdirectories for images and masks.

3.1.2. CardioSmartSet dataset

Data and echograph details. The CardioSmartSet dataset comprises 27 four-chamber 2D echocardiographic videos obtained from clinical sources, recorded with the ultrasound scanners GE Vivid E95 and Philips Affiniti 30 under standardised acquisition conditions, and split into clips of 3 cardiac cycles. The dataset for each patient includes the original video in AVI format, recorded at 1016 × 708 pixels, and the corresponding frames in JPG format for end-systole and end-diastole. It also contains expert segmentation annotations for these frames, provided in JPEG format and later converted to a PNG binary mask to meet project requirements, as well as a list of clinical parameters, such as apical height, left ventricular volume, and ejection fraction. The data were fully anonymised to ensure compliance with privacy regulations.

The patient characteristics of the dataset are reported in Section 1 of the Supplementary Material and summarised in Table 1.

Frame identification and segmentation. A team of four medical professionals, hereafter referred to as the focus group, reviewed the same videos and identified the frames corresponding to end-systole and end-diastole for each analysed video cycle. Once a unanimous consensus on frame selection was reached, the frames were integrated into the dataset. In cases of differing opinions on frame identification, if the discrepancy was confined to a few frames (±3), the frame selected by the majority of the group (over 75%) was adopted and incorporated into the dataset. In instances of substantial disagreement during segmentation, a supplementary review was conducted, involving discussions among senior experts to guarantee precise frame selection. Half of the focus group carried out the segmentations, while the other half validated them.

Data preprocessing. Preprocessing involved specific transformations to optimise data for model training. Original video frames were cropped from 1016 × 708 pixels to a square 708 × 708-pixel format, centring the LV while preserving essential echocardiographic features.

To ensure architectural consistency between pretraining and fine-tuning, all images were subsequently down-scaled to match the spatial resolution used for training on the EchoNet-Dynamic dataset. EchoNet-Dynamic videos are natively provided at 112 × 112 pixels, and the initial MultiResUNet model was pretrained at this resolution. CardioSmartSet frames were therefore resized to 112 × 112 pixels before both training and inference, while EchoNet-Dynamic data were used at their native resolution.

This resizing step allowed the model to leverage the feature representations learned from the large-scale public dataset during transfer learning. Previous studies have shown that aggressive downsampling of echocardiographic images can substantially reduce computational burden while preserving the macroscopic anatomical features required for automated analysis [4]. In particular, reducing image resolution can decrease file size and computational cost by up to 96%–99% without significantly degrading diagnostic accuracy in echocardiographic interpretation [4].

Patient dataset composition. The patient sample, retrospectively collected, consists of 22 healthy individuals and 5 patients diagnosed with specific cardiac conditions: 1 with amyloidosis, 2 with ischaemic heart disease-related heart failure, and 2 with hypertension. The part of the dataset containing healthy subjects includes 9 male and 13 female patients, with an average age of 30 years (±4.63 years standard deviation). The other part of the dataset, containing 5 male patients diagnosed with specific cardiac conditions, has an average age of 75 years (±2.1 standard deviations).

Table 1
Patient demographic summary of the CardioSmartSet dataset.

Group	N° Patients	Sex (M/F)	Mean Age (years) ± SD	Status
Healthy subjects	22	9/13	30 ± 4.63	Healthy
Pathological subjects	5	5/0	75 ± 2.1	Specific cardiac conditions

Dataset split for fine-tuning purpose. The CardioSmartSet dataset comprised 54 images and their corresponding segmentation masks (systole and diastole available for each patient video). It was divided into training, validation, and test sets comprising 63%, 15%, and 22% images, respectively, derived from the conventional 70–15–15 split and subsequently adjusted to obtain a test set of sufficient size given the limited dataset.

To prevent the model from learning patient-specific features, all images from a given patient were kept together in the same split.

3.2. Deep learning-based segmentation

The MultiResUNet architecture was selected due to its effectiveness in addressing issues like anatomical scale variability, speckle noise, and precise boundary delineation for volumetric quantification [3,11,22,22–24]. This architecture directly meets these needs through three design features: multi-scale feature extraction, skip connections, and improved sensitivity to low-contrast boundaries. Echocardiography is characterised by significant variability of regions of interest (e.g., dilated vs. hypertrophic ventricles) [11,23]. In contrast to U-Net architectures, which typically rely on fixed kernel sizes that limit their ability to contextualise features of varying sizes, the MultiRes blocks utilised in MultiResUNet approximate 5 × 5 and 7 × 7 convolutions using sequences of 3 × 3 filters [22]. This multi-resolution analysis enables the model to reconcile global spatial features (ventricular shape) with local details, making it more robust to perturbations and artefacts compared to standard architectures [22]. In standard encoder–decoder networks, a semantic gap exists between the low-level, high-resolution features from the encoder and the high-level semantic features in the decoder [22]. Direct concatenation of these incompatible features can impede learning, whereas MultiResUNet employs Residual Paths, which introduce non-linear convolutional layers within the skip connections. This architectural choice mitigates vanishing gradient issues and stabilises training, ensuring consistent performance across heterogeneous clinical data maps [22]. It also keeps the model geometrically stable frame by frame and allows us to compute a representative EF averaged over multiple cardiac cycles without introducing excessive noise from segmentation [11].

3.2.1. Problem formulation

The semantic segmentation task is a pixel-wise classification problem where each pixel is assigned a class label indicating its category. The input is an image, and the output is a segmentation mask of the exact spatial dimensions.

This study employs the Dice loss derived from the DSC. The Dice Loss is used in image segmentation tasks to measure agreement between the predicted segmentation and the ground-truth annotations.

The loss function is defined as:

$$L_{\text{Dice}} = 1 - \frac{2\sum_i \hat{y}_i y_i}{\sum_i \hat{y}_i + \sum_i y_i}$$

where:

- $\hat{y}_i$ represents the predicted segmentation value for pixel i, expressed either as a probability between 0 and 1 or as a binary classification.
- y_i is the corresponding ground truth value for pixel i.
- The numerator accounts for the correctly predicted foreground pixels, enhancing the overlap between predictions and annotations.
- The denominator normalises the loss by considering the total number of predicted and actual foreground pixels, ensuring scale invariance.
- The summation runs over all pixels in the image.

By maximising the overlap between predicted and actual segmentations, Dice Loss balances precision and recall, making it particularly suitable for handling class imbalance in segmentation tasks.

3.2.2. Training hyperparameters

A systematic exploration of critical training hyperparameters of the MultiResUNet architecture was performed on the EchoNet-Dynamic dataset to determine the optimal configuration for semantic segmentation. The final training was extended to 30 epochs, with early stopping configured to terminate after 5 epochs to mitigate overfitting; batch sizes for training and validation were four; optimisation employed the Adam algorithm with Dice Loss as the loss function.

3.2.3. Finetuning hyperparameters

During fine-tuning, the MultiResUNet model pretrained on EchoNet-Dynamic was trained with a batch size of 8 for up to 60 epochs, using early stopping with a patience of 5 epochs. The optimisation was performed using the AdamW optimiser with a weight decay of 5×10^{-4} and a base learning rate of 5×10^{-5}.

A OneCycleLR learning rate schedule was adopted with the following parameters: max_lr = 2×10^{-4}, div_factor = 25, final_div_factor = 10 000, and pct_start = 0.2. This configuration results in an initial learning rate of 8×10^{-6}, which increases to a peak of 2×10^{-4} during the first 20% of the training iterations and then gradually decreases to a final value of 2×10^{-8} over the remaining iterations.

3.2.4. Evaluation metrics

To evaluate segmentation performance objectively, two standard metrics were used: the DSC, which measures accuracy by assessing overlap between predicted segmentation masks and ground-truth annotations, and the Intersection over Union (IoU), which quantifies spatial agreement by computing the ratio of their intersection to their union.

The DSC is defined as:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

where X represents the set of pixels in the predicted segmentation mask and Y represents the set of pixels in the ground truth mask. A higher DSC value, closer to 1, indicates greater overlap between the two masks.

Similarly, the IoU is computed as:

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \tag{2}$$

where $|X \cap Y|$ denotes the number of pixels in the intersection of the predicted and ground truth masks and $|X \cup Y|$ represents the total number of pixels in their union. IoU measures the spatial overlap between the segmentation outputs and the reference standard.

DSC and IoU metrics were computed at the frame level by comparing the predicted segmentation masks generated by the model with the corresponding ground-truth annotations. For the EchoNet-Dynamic dataset, ground-truth masks were obtained from the official dataset annotations [20], whereas for the proprietary CardioSmartSet dataset they were provided but the focus group involved in the study. The reported DSC and IoU values correspond to the mean performance across all frames in the respective test datasets.

These metrics collectively assess the model's ability to accurately delineate left ventricular structures while ensuring spatial consistency across the entire echocardiographic sequence.

3.2.5. Comparison and rationale for the selected architecture

A comparative analysis was conducted across three models: MultiResUNet [22], YOLO v11, a cutting-edge model released in September 2024, and SAM 2, a self-supervised segmentation framework introduced by Meta in July 2024.

Each of these models underwent a two-stage training process: they were pre-trained on the publicly available *EchoNet-Dynamic* dataset, leveraging its large-scale echocardiographic data to establish robust feature extraction capabilities; then they were fine-tuned on our proprietary clinical dataset.

YOLO v11 was adapted for real-time segmentation of left ventricular boundaries and selected for its well-known efficiency, a compromise between speed and precision, which is valuable for fast LV localisation in cine-loops [25]. The YOLO model was trained using the SGD optimiser with an initial learning rate of 1×10^{-3}, decayed according to a cosine annealing schedule. The model was trained for 50 epochs using a batch size of 16 images, a standard optimisation parameter unrelated to the number of subjects in the clinical dataset. The loss function used was the standard YOLO multi-task loss, combining IoU-based localisation, objectness confidence, and classification error terms.

The SAM v2 model was incorporated for its self-supervised learning capabilities and as a modern foundation-model baseline, enabling testing of a general-purpose segmentation model on ultrasound data [26]. The AdamW optimiser was used to train the model, with an initial learning rate of 1×10^{-4} and a cosine annealing schedule. Using a Dice loss function, 8 batches were used over 30 epochs.

3.2.6. Continuous learning mechanism

CardioSmartAssist implements an expert-in-the-loop continuous learning strategy that progressively enhances the segmentation model using clinician-validated feedback. When a clinician reviews and corrects a segmentation, the revised mask and the corresponding pseudonymised frame are stored in an encrypted local directory. At each application start, the software verifies the number of accumulated validated frames. The system monitors this number and triggers the export process once a predefined threshold is reached. This threshold is configurable and is currently set to fifty frames for the initial deployment. When the threshold is met, the software automatically assembles an encrypted export package containing only anonymised images, corrected masks, and non-sensitive metadata. The package is then transferred using encryption mechanisms that ensure restricted access and comply with established data protection principles. Upon receipt, authorised members of the research team integrate the validated samples into a retraining buffer. The model is subsequently fine-tuned offline on this expanded dataset, avoiding any on-device updates and preserving reproducibility and stability. After internal validation, the updated model is incorporated into the next software release and redistributed to users as part of the standard versioned update cycle. This iterative feedback mechanism enhances the system's adaptability, enabling it to learn and improve progressively. A schematic overview of this iterative expert-in-the-loop process is provided in Fig. 2.

3.3. Frame selection for ejection fraction calculation

The algorithm automates the analysis of a sequence of segmented LV images, enabling the identification of key cardiac cycle phases (systole and diastole) by tracking variations in ventricular area over time.

After the segmentation model completes the inference pass over all frames of an echocardiographic video, the resulting binary masks are saved in a dedicated directory. Each mask encodes the left ventricular cavity as foreground pixels with intensity 255 (white), while the background is assigned a value of 0 (black). This organised collection of masks constitutes the sole input to the subsequent EF computation pipeline. The algorithm sequentially processes these masks in frame order, extracts the frame index, loads each image in greyscale, and computes the ventricular area as the total number of foreground pixels. From this sequence of areas, an area–frame curve is constructed and used to identify systolic and diastolic phases. The same masks are then used to extract anatomical landmarks (apex and mitral-valve midpoint), measure ventricular length, and estimate left-ventricular volume using the area–length method. In this way, the full EF estimation relies solely on the segmentation outputs from a single cardiac acquisition.

To detect cardiac cycle phases, the algorithm identifies local maxima in the ventricular area curve as end-diastolic phases and local minima as end-systolic phases, using peak prominence thresholds to select only significant extrema. Cardiac cycles are defined by designating Cycle 0 at the initial detected diastolic peak, with subsequent cycles marked between consecutive diastolic phases. This approach aligns with medical experts and adheres to established guidelines, ensuring that the identified diastoles and systoles within each cycle are accurately categorised and that clinical validity is preserved. A plot of ventricular area versus frame number is generated to facilitate visualisation of the data, as shown in Fig. 3, and another example is provided in Section 3.1 of the Supplementary Material.

In particular, the `analyze_peaks` function implemented reads all binary masks in ascending frame order and computes the foreground area A_j for each frame j. Candidate diastolic maxima and systolic minima are detected via `find_peaks` on A_j and $-A_j$, respectively, with parameters min_distance = 20 and min_prominence = 5000. Each raw peak at index j is retained only if

$$\frac{|A_j - A_{j-1}|}{A_{j-1}} \leq 0.3$$

(to remove outliers), and frame 0 is added as a diastolic maximum only if

$$\frac{|A_0 - \overline{A}_{\max}|}{\overline{A}_{\max}} \leq 0.1,$$

where $\overline{A}_{\max}$ is the mean area of the other valid maxima The results are exported to CSV and plotted with red markers for maxima (diastolic peaks) and green for minima (systolic troughs), annotating each label with a vertical offset of $0.03\,A_j$ to improve readability.

To facilitate the transition between frame analysis and volume computation, a schematic summary of the workflow is presented in Fig. 4, showing how segmentation outputs, area variations, landmark detection, and volume calculations contribute to deriving the EF.

3.4. Volume calculation

The " Volume Calculation" section includes two subsections: the first describes the Area-Length method for calculating left ventricular volume, and the second discusses how to derive the ejection fraction from these volumes.

3.4.1. Volume computation through the area-length method

This section outlines a two-stage process for calculating left ventricular volume, beginning with the identification of key anatomical landmarks: the apex and the midpoint of the mitral valve plane. The apex and midpoint were determined using a contour-based landmark detection approach, which ensures robust and reproducible localisation.

Morphological operations facilitated contour extraction to define the left ventricular boundary, with the apex identified at the maximum curvature of the outline and the midpoint established by bisecting the long axis of the ventricle, thereby ensuring consistent landmark placement.

Further computational details, including intermediate steps and algorithmic specifications, are available in Section 3.2 of the Supplementary Material.

The ventricular length L is defined clinically as the Euclidean distance between the apex of the left ventricular chamber and the midpoint of the mitral valve line segment [27,28].

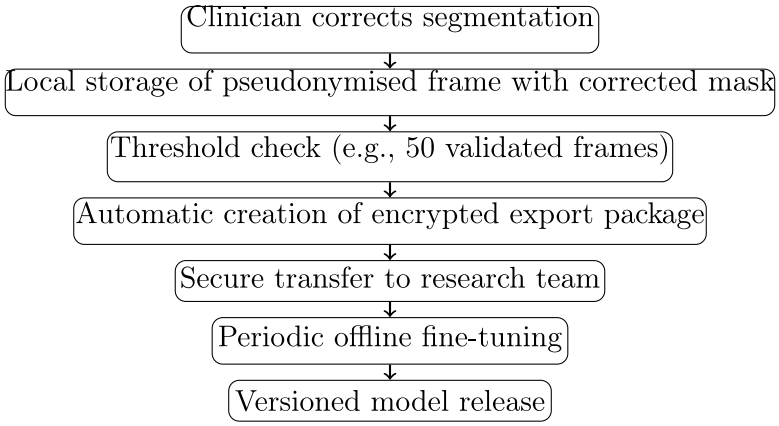

Fig. 2. Workflow diagram of the expert-in-the-loop continuous learning mechanism implemented in CardioSmartAssist.

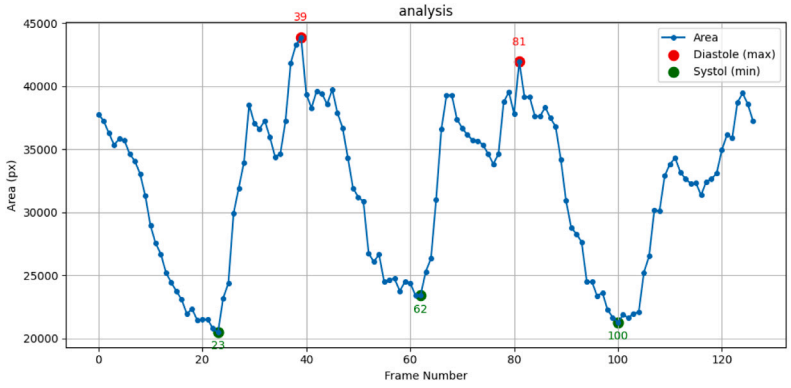

Fig. 3. Plot of ventricular area of Patient 3 versus frame number without considering the first frame, since it is not classified as complete diastole.

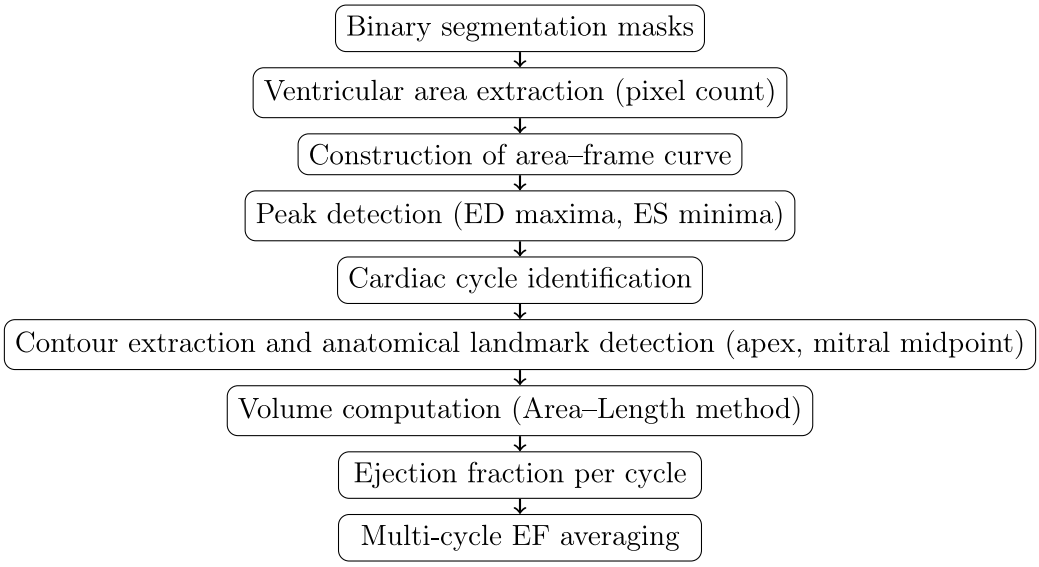

Fig. 4. Simplified data-flow diagram from segmentation masks to EF computation.

The left ventricular volume (V) was estimated using the area-length method, an echocardiographic technique that models the left ventricle as a prolate ellipsoid (see Fig. 5). This geometric approximation is widely employed in medical imaging to enhance volume estimation through 2D echocardiography [29].

This method is mathematically expressed as:

$$V = \frac{8}{3} \cdot \frac{A^2}{\pi L}$$

where A represents the left ventricular cross-sectional area, and L denotes the length of the ventricular long axis [30]. This formulation enables a reliable estimation of ventricular volume from two-dimensional imaging data.

A systematic comparison of the area–length method and the Simpson Method in EchoNet-Dynamics was conducted, using EF values from EchoNet-Dynamics as the reference. The LVEF was calculated from

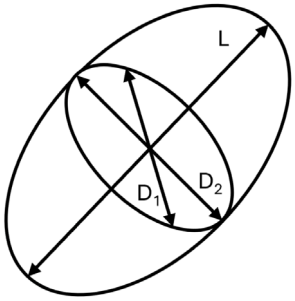

Fig. 5. Schematic representation of the LV modelled as a 3D prolate ellipsoid, as used in the area-length method for volume computation. The parameter L denotes the ventricular long-axis length measured from the apex to the midpoint of the mitral annulus. The quantities D_1 and D_2 represent the orthogonal short-axis diameters derived from the segmented cross-sectional area.

volumes obtained via the area–length approach. Descriptive statistics, including means and standard deviations, characterised central tendency and variability. Error metrics such as the MAE and RMSE were computed to evaluate the accuracy, precision, and stability of the area–length method in estimating LVEF.

3.4.2. Ejection fraction calculation

The EF was computed based on the estimated EDV and ESV according to the following equation:

$$EF(\%) = \frac{EDV - ESV}{EDV} \times 100$$

This measure quantifies the fraction of blood ejected from the LV during each cardiac cycle, thereby providing clinical insights into cardiac systolic function.

3.5. UML modelling and UI development

3.5.1. Structured system design through UML diagrams

To better understand the routine diagnostic workflow and identify which manual procedures are most time-consuming and prone to operator variability, we conducted preliminary consultations with cardiologists before defining the system architecture. The acquisition and interpretation of echocardiograms were the main topics of discussion, with a focus on the manual estimation of the left ventricular ejection fraction. The clinical team also provided thorough explanations of the time-consuming steps required to complete an examination, and we observed several processes to ensure the pipeline accurately reflected clinical practice. The design of the multi-cycle EF computation module, the selection of segmentation targets, and the ranking of UI features were all influenced by the insights gained from these consultations.

Building on the clinical requirements identified through these consultations, the system was subsequently formalised using UML diagrams, which served as standardised tools to clearly and unambiguously represent its structural and behavioural components.

The application of UML enhanced communication between engineers and clinicians, resulting in an accurate conversion of clinical requirements into technical specifications.

Following a collaborative validation phase, the diagrams that served as the foundation for application development were: Use Case Diagram, Use Case Descriptions, a Sequence Diagram, and Activity Diagram. Partial representations are included in Section 3.3 of the Supplementary Material to enhance readability, while full-resolution versions of the diagrams are available upon request.

The Use Case Diagram captures interactions between the system and external users, specifically cardiologists and sonographers, outlining their roles and permissions.

The Use Case Descriptions diagram provides detailed breakdowns of each use case, including action sequences, involved actors, and expected outcomes, helping developers understand system behaviour and user interactions.

The Sequence Diagram focuses on the dynamic interactions among system components, details the sequence of messages during operations, and maps the logical flow, aiding in identifying potential bottlenecks and enhancing clinical workflows essential for precise medical applications.

The Activity Diagram visualises system workflows by categorising tasks into swim lanes to clarify the roles of various actors and emphasise decision points, such as cardiologists' segmentation modifications, and includes verification steps prior to data storage to ensure accuracy.

Overall, the diagrams together enhance workflow traceability and system efficiency.

3.5.2. The design of graphical user interfaces

The aforementioned design activities resulted in an intuitive graphical interface that clinical professionals can utilise with ease. A user-centred design approach seeks to enhance diagnostic precision and efficiency while aligning with clinical workflows.

The interface prototyping utilised PyQt and Qt Designer to create an interactive system that met clinical requirements through iterative review by medical professionals. Users can log in to upload echocardiographic videos, enter patient information, and analyse cardiac cycles.

The primary EF dashboard, reported in Fig. 6, displays calculated EF values and clinical conditions, providing direct access to regulatory standards and frame-by-frame analysis of the cardiac cycle. Visualisations of left ventricular area over time help users determine the cardiac cycle phase and examine systolic and diastolic frames, as represented in Section 3.3 of the Supplementary Material. Clinicians can perform manual re-segmentation to enhance accuracy, similar to that achieved with conventional echocardiographic devices. The instrument allows exclusion of cycles affected by ventricular extrasystole for recalibrated ejection fraction calculations. A pixel-to-millimetre calibration ensures accuracy in ventricular measurements, while anomaly-detection capabilities alert physicians to values outside the standard range, accompanied by text descriptions.

4. Results and discussion

This section thoroughly evaluates the proposed solution, focusing on the performance of the DL module, the accuracy of the segmentation methodology, and the clinical reliability and impact of the results.

4.1. Benchmarking MultiResUNet, YOLO, and SAM

To facilitate a more precise analysis, the following subsections address the training process, fine-tuning strategy, and the impact of data augmentation on segmentation performance.

4.1.1. Training on EchoNet dataset

Firstly, the performance of the model trained only on the EchoNet-Dynamic Dataset has been analysed, comparing the DSC and IoU scores on the EchoNet-Dynamic test set, which comprises 1998 images. While the DSC and Io scores of SAM v2 and YOLO v11 are not significantly lower than those of MultiResUNet in clinical settings, as demonstrated in Table 2, discussions with cardiologists suggest that the more anatomically plausible segmentations should be taken into account.

As concerns the MultiResUNet model, the mean DSC reached 0.9328, surpassing the state of the art of 0.92 [2]. The higher DSC achieved in our study is mainly attributable to the use of MultiResUNet, which differs from the segmentation architecture initially employed in EchoNet-Dynamic, a CNN model with residual connections and spatiotemporal convolutions [2]. In particular, MultiResUNet enhances

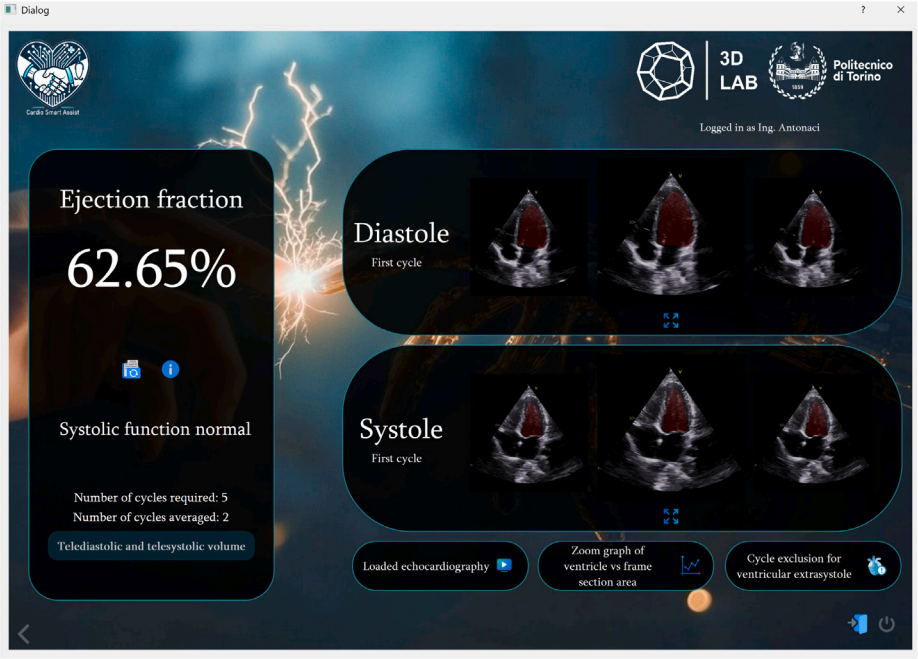

Fig. 6. Principal CardioSmartAssist dashboard: Ejection Fraction Dashboard.

Table 2
Performance comparison of different segmentation models after the training on the EchoNet-Dynamic dataset.

Model	Mean DSC	Mean IoU
MultiResUNet	0.9328	0.8957
YOLO v11	0.9141	0.8432
SAM v2	0.9279	0.8658

the classical U-Net by improving multi-resolution analysis and reducing the semantic gap between encoder and decoder layers through two key architectural components [22]. First, MultiRes Blocks replace the standard convolutional layers of the U-Net by integrating spatial features across different receptive fields. They use cascaded 3×3 convolutions to efficiently approximate larger 5×5 and 7×7 kernels, enabling more precise segmentation, important for delineating left-ventricular boundaries [22]. Second, Res Paths replace the original skip connections more effectively, better aligning low-level encoder features with high-level decoder representations. By adding non-linear convolutional layers with residual connections, Res Paths produce more homogeneous fused feature maps, improving robustness to perturbations and accelerating convergence [22]. This result indicates a high degree of overlap between predicted and ground-truth masks, with an IoU of 0.8957, as shown in Table 2. Precision and recall were balanced at 0.9612 and 0.9293, respectively, yielding an F_1 score of 0.9450. These results underscore the method's robustness and precision, as evidenced by an overall accuracy of 99,10% and a low area estimation error of 0.0138.

MultiResUNet was selected based on a quantitative performance comparison with the YOLO and SAM models, which demonstrated superior DSC, as shown in Table 2.

4.1.2. Fine tuning on CardioSmartSet dataset

Beyond numerical metrics, a qualitative assessment of the inference results across all frames of the echocardiographic video confirmed that MultiResUNet produced the most consistent and anatomically coherent segmentations throughout the cardiac cycle. Fig. 7 offers an assessment of the segmentation performance of the three models, emphasising their accuracy in delineating the LV structures of an example image test of the CardioSmartSet Dataset.

In particular, the main issue with the masks produced by YOLO is that, because of its bounding-box-based approach, they tend to generate overly straight, sharp edges that do not accurately reflect the LV's natural contours. On the other hand, the segmentations produced by SAM tend to exhibit excessively jagged edges, which may compromise their clinical reliability.

The combination of robust performance, stability across frames, and computational feasibility made the MultiResUNet the most suitable choice for integration into the proposed framework and was therefore selected as the reference model for all subsequent analyses, while preliminary experiments with YOLO and SAM produced segmentation outputs that were not sufficiently clinically acceptable for downstream EF computation, according to feedback from the cardiologists involved in the focus group.

4.1.3. The influence of fine-tuning and data augmentation on MultiResUnet performance

In developing the proposed framework, extensive evaluations and multiple experimental trials were conducted for each of the three DL architectures presented. The study evaluated different loss functions, learning rates, batch sizes, training epochs, augmentation strategies, schedulers, and other hyperparameters to optimise the pretrained model on the proprietary CardioSmartSet dataset. These variations enabled an analysis of how different configurations affect segmentation performance.

This evaluation compared models trained with and without fine-tuning on the CardioSmartSet dataset and with and without data augmentation. This comparative analysis examined how domain-specific clinical data fine-tuning improves segmentation accuracy and how data augmentation improves model generalisability.

As concerns data augmentation, rotations, translations, and flips were implemented to further test the model's generalisation capability [33]. The augmentation techniques were carefully selected in agreement with experienced cardiologists to ensure they did not introduce distortions or artefacts that could render the images non-realistic or clinically irrelevant [33].

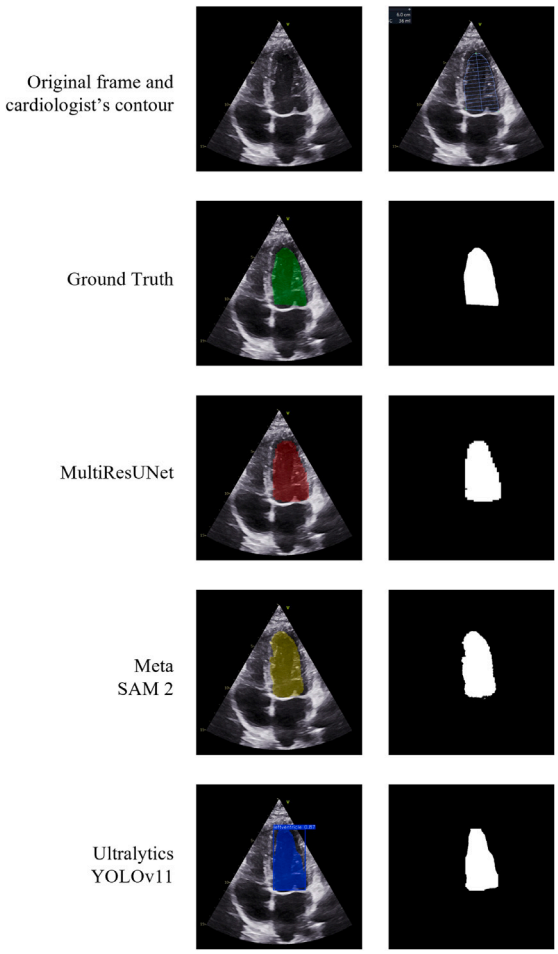

Fig. 7. Example Inference of MultiResUNet, SAM2 and YOLOv11. The first row displays, from left to right, the original frame acquired and saved by the cardiologist directly from the ultrasound system, followed by the same frame annotated with the left ventricular contour manually drawn by the clinician. The second row, labelled Ground Truth, shows the clinician-defined contour overlaid as a semi-transparent green mask and the corresponding binary segmentation mask. The third, fourth, and fifth rows present the segmentation outputs of MultiResUNet, Meta SAM 2, and Ultralytics YOLO 11, respectively. In each case, the predicted masks are shown as overlays on the original frame (in red, yellow, and blue), alongside their corresponding binary masks. The architecture logos were sourced from their respective official repositories or documentation [31,32].

Based on the DSC and IoU metrics in Table 2, MultiResUNet was ultimately chosen as the most suitable model.

Table 3 demonstrates that thanks to fine-tuning on the high-resolution CardioSmartSet dataset, the performance of the MultiResUNet model improves, achieving a DSC of 0.9189 and an IoU of 0.8586 compared to the DSC of 0.8973 and an IoU of 0.8131 attained by the model without fine-tuning when performing inference on the same images.

After training on a larger dataset, it is recommended to fine-tune on images acquired under the same conditions as the target inference dataset, because medical ultrasound images may vary widely depending on acquisition conditions, scanner, and patient [34]. Transfer learning and domain adaptation are required to improve segmentation performance, accurately capture the distinctive visual features of clinical images, and refine features for specific tasks more effectively [34,35]. The procedure facilitates edge detection, maintains segmentation stability while working with real-world data [35], and enhances the model's accuracy in drawing tighter endocardial outlines and its stability in the presence of noise, speckle fluctuation, and low-contrast boundaries [34].

However, after fine-tuning the CardioSmartSet dataset without data augmentation, the model's performance improved, achieving an average DSC of 0.9189 and an average IoU of 0.8586. Further fine-tuning with data augmentation yielded a similar DSC of 0.918, suggesting that while augmentation contributed to better generalisation, it introduced slight variability in segmentation performance, making it preferable to adopt the last-described model configuration without data augmentation.

Combined analyses support a structured validation of the proposed segmentation model. A multi-level validation strategy included geometric, architectural, and clinical evaluations. Geometric segmentation accuracy was assessed using DSC and IoU, and ablation experiments analysed the contributions of the proposed training pipeline, focusing on data augmentation and fine-tuning on the CardioSmartSet clinical dataset. Clinical performance was evaluated by comparing model outputs (end-diastolic and end-systolic volumes, ejection fraction) to verified clinical measurements, and model error was assessed against inter-observer variability.

4.2. Statistical validation on EchoNet-dynamic dataset of the area-length method against simpson's method

A review of the literature led to the selection of the area-length method for volume estimation in two-dimensional echocardiography. This method uses the mitral valve's apex and midpoint to estimate volume, differentiating it from Simpson's biplane method, which requires multiple apical views and disc summation.

The requirement to disclose the results of both methods (area–length and Simpson's Biplane) stems from their inherently distinct characteristics and their recognition as established clinical standards for estimating LV volume. According to the guidelines established by the American Society of Echocardiography [36], the biplane Simpson method (also known as the method of discs) is considered the most accurate approach for determining left ventricular volumes. This is because the Simpson Biplane method uses data from two orthogonal planes (the apical 4- and 2-chamber views) to correct shape distortions and requires fewer geometric assumptions than other 2D techniques. Nevertheless, the application of alternative two-dimensional techniques, such as the area–length method, remains a viable option referenced in the ASE guidelines, despite the Biplane method being the preferred approach [36]. The area–length method differs from Simpson's method in that it relies on geometric assumptions about the ventricle that are often derived from a single dimension or plane. This may result in anticipated variations in absolute volumetric measurements [30].

A comprehensive statistical comparison between the area–length method and the Simpson method, used as the reference in EchoNet-Dynamics, shows that LVEF estimates from the former closely match those of the latter, confirming its consistency with the gold-standard Simpson's method. Using the 20060 EchoNet-Dynamics frame values as ground truth, EF was computed by deriving end-diastolic and end-systolic volumes using the area–length approach. The mean EF obtained from EchoNet-Dynamics was 55.71%, whereas the area–length method yielded 56.81%, a negligible difference of 1.10%. Variability, as measured by the standard deviation, was similarly close (12.34% versus 12.90%). Error metrics further corroborate the method's reliability: the MAE was 2.61%, indicating minimal average deviation from the reference, and the RMSE was 3.51 percentage points, suggesting only a few larger deviations without substantially undermining overall performance.

The mean, standard deviation, MAE, and RMSE indicate that the Area-Length method accurately estimates EF with minimal error and high stability, suggesting that it may offer advantages in computational simplicity and automation.

Table 3
Performance of the MultiResUNet model before and after fine-tuning on CardioSmartSet dataset and tested on CardioSmartSet Test Dataset.

MultiResUnet	Test size	DSC	IoU	Precision	Recall	F1	Accuracy	MAE
Pre Fine Tune	12	0.8973	0.8131	0.9264	0.8710	0.8967	0.9869	0.0131
Post Fine Tune	12	0.9189	0.8586	0.8982	0.9533	0.9239	0.9900	0.0100

4.3. Comparison of EF and volumes calculated by CardioSmartAssist vs. clinical values

Each video in the CardioSmartSet dataset was analysed by CardioSmartAssist to detect end-diastolic and end-systolic frames, compute the segmented areas, convert these to volumes via the area–length method, and calculate the corresponding EF. The mean diastolic and systolic volumes ($\overline{V}_{ED}$, $\overline{V}_{ES}$) and EF values produced by the new model were then compared against the single-cycle measurements reported by expert cardiologists using standard echocardiographic practice.

Beyond segmentation accuracy, the clinical impact of the proposed workflow was evaluated through complementary quantitative metrics Specifically, the agreement between automated and clinician-derived ejection fraction values was assessed, together with the accuracy of ventricular volume estimation and the reproducibility of measurements across multiple cardiac cycles. The comparison with clinician measurements was quantified using mean absolute error and additional distribution statistics, while the multi-cycle aggregation analysis was used to evaluate the robustness of the system.

As demonstrated in Supplementary Table S2, the mean absolute differences in EF are equivalent to 10%, while the mean discrepancies in volume are 28%.

To provide a visual summary of these results, Fig. 8 illustrates the agreement between clinician-measured EF and the EF estimated by CardioSmartAssist across the study cohort.

While the scatter plot provides a visual indication of the correlation between the two measurements, the agreement between the automated and clinical EF estimates was further evaluated using a Bland–Altman analysis (Fig. 9).

The Bland–Altman plot assesses the agreement between the two methods by representing the difference between CardioSmartAssist and clinician EF measurements as a function of their mean value. In this analysis, the mean difference (bias) between the two methods was 0.91%, indicating a minimal systematic deviation of the automated estimates relative to the clinical reference. The standard deviation of the differences was 6.87%, resulting in 95% limits of agreement ranging from −12.55%to 14.37% (bias ±1.96 SD).

Most observations fall within these limits, suggesting that the EF estimates produced by CardioSmartAssist show good agreement with clinician measurements across the observed EF range. These findings support the reliability of the automated approach for estimating ventricular ejection fraction in comparison with standard clinical assessment.

The observed 10% mean absolute EF difference lies below the inter-observer variability among cardiologists reported in the literature, equal to 13.9% [2]; cardiologists participating in the focus group agreed that the pipeline functions within the range of human variation.

The observed discrepancies in volume assessment can be attributed to inherent variability in operator methodology during manual calibration. The application prompts "pixel-cm" calibration through its screen, which contributes to the observed variability. Moreover, the clinical significance of absolute volume differences is often context-dependent. In routine practice, cardiologists primarily evaluate trends in ventricular volume over time rather than relying on absolute measurements. This perspective aligns with the notion that monitoring a patient's cardiac function over time is more clinically relevant than single-point measurements, provided that the methodology remains consistent. Furthermore, discussions with clinicians confirm that absolute volume measurements obtained via echocardiography are generally not considered highly reliable in practice. In cases where precise volumetric assessment is clinically necessary, Cardiac Magnetic Resonance (CMR) remains the preferred standard of Ref. [37].

Typically, LVEF ranges from 50% to 70% in healthy individuals, with slight gender-specific variations of 52% to 72% in males and 54% to 74% in females. Values below these thresholds indicate systolic dysfunction, which is essential for classifying heart failure. Specifically, an LVEF below 40% is indicative of heart failure with reduced ejection fraction (HFrEF), an LVEF between 41% and 49% suggests heart failure with mildly reduced ejection fraction (HFmrEF), and an EF of 50% or higher corresponds to heart failure with preserved ejection fraction (HFpEF) [37,38].

As already cited, the variability among operators is up to 13.9% [2]. In the CardioSmartSet test cohort of 27 patients, only five cases exhibited an absolute EF discrepancy greater than 15% between CardioSmartAssist and expert cardiologist measurements (Table 4).

Patient 4 represented the only case in which the automated EF estimation could potentially lead to misclassification around a clinically relevant threshold. Specifically, the algorithm underestimated EF (49.8% vs. 66%, δEF = 25%), which could have resulted in an incorrect classification around the HFmrEF boundary (41%–49%).

Patient 16's EF was overestimated from 68% to 79.1%, but remained within the preserved-function range and thus did not affect clinical decision-making. Patients 20, 21, and 22 presented large EF deviations but were already clinically classified as high-risk cases and were therefore subject to close medical surveillance. Consequently, the observed discrepancies did not alter the clinical management pathway.

It can be concluded that the robust nature of the CardioSmartAssist automated EF assessment is supported by two factors. Firstly, the incidence of substantial discrepancies is minimal. Secondly, the correct clinical status of patients who require further investigation is preserved.

A well-known challenge in deploying deep learning models in real-world medical imaging is the domain shift caused by variations in ultrasound equipment vendors, acquisition protocols, and operator techniques [19]. This variability can affect model performance when algorithms trained on one dataset are applied to data acquired under different clinical conditions.

Rather than relying solely on static algorithmic generalisation, the CardioSmartAssist framework incorporates a human-in-the-loop workflow. In cases where domain shift leads to segmentation inaccuracies, clinicians can review and correct the generated contours through the graphical interface [19,39]. These corrections can then be incorporated into the continuous learning pipeline, potentially enabling the model to progressively adapt to the data distribution of the deploying clinical environment.

4.3.1. Reproducibility and robustness across multiple cardiac cycles

To evaluate the added benefit of multi-cycle analysis, CardioSmartAssist computed EF for every detected cardiac cycle and averaged these to provide a multi-cycle EF ($\overline{EF}$) along with mean diastolic and systolic volumes. These metrics were then compared to the single-cycle EF values provided by clinicians reported in Supplementary Table S3. While single-cycle estimates (based on one to three cycles) are common in clinical practice, averaging across all available cycles reduces beat-to-beat variability, motion artefacts, and transient physiological fluctuations. In this particular instance, the number of cycles was constrained by the duration of the input video, which comprised only a limited number of cardiac cycles. It is noteworthy, however, that the software can operate for a greater number of cycles in accordance with current clinical guidelines.

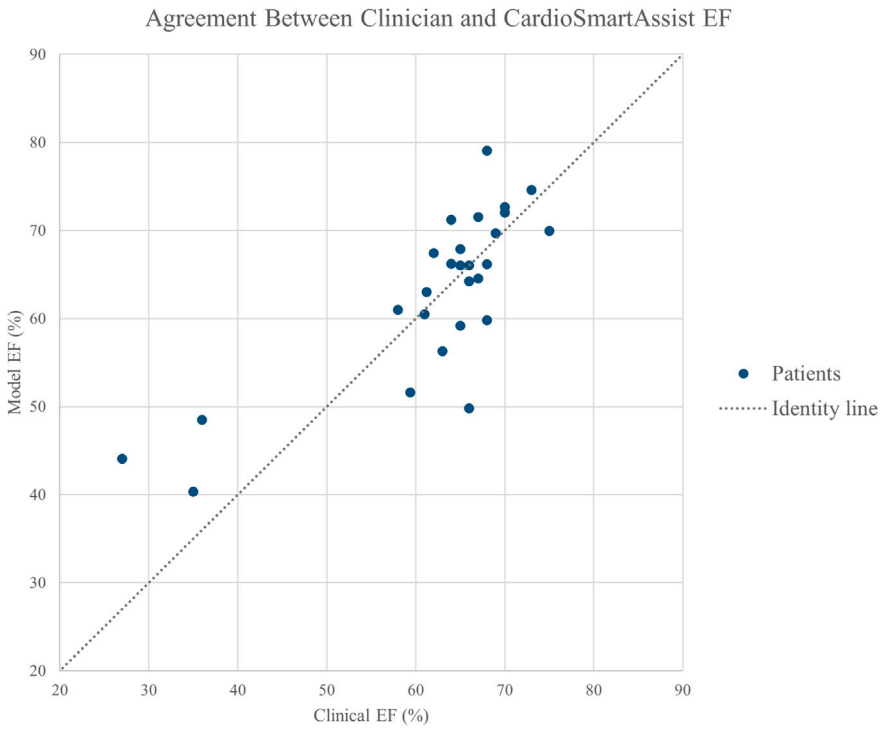

Fig. 8. Agreement between clinician-measured ejection fraction and CardioSmartAssist EF estimates across the study cohort. Each point represents an individual patient. The dotted diagonal line corresponds to the identity line ($y = x$), indicating perfect agreement between clinician and model measurements. Points closer to the identity line indicate higher agreement, while deviations highlight cases with larger estimation differences.

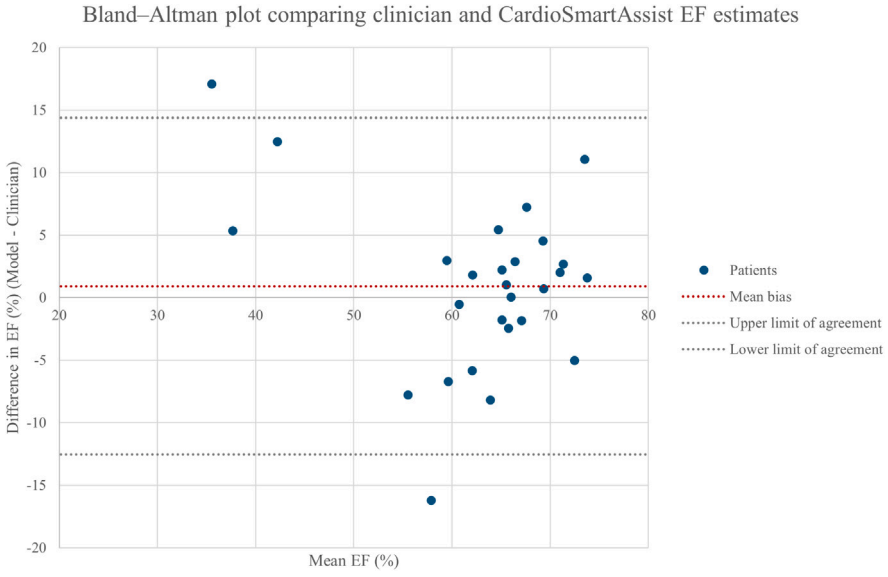

Fig. 9. Bland–Altman plot comparing clinician-measured ejection fraction and CardioSmartAssist EF estimates across the study cohort. The x-axis represents the mean EF of the two measurements, while the y-axis represents their difference. The central dashed line indicates the mean bias, and the upper and lower dashed lines represent the 95% limits of agreement (bias ±1.96 SD).

Table 4
Patients with $|\Delta EF| > 15\%$ in the CardioSmartSet test cohort ($N = 27$).

Patient ID	Filename	Phase	Clin. EF (%)	Model EF (%)	$\lvert\Delta EF\rvert$
4	p4_d	Diastole	66	49.79	25%
16	p16_d	Diastole	68	79.06	16%
20	p20_d	Diastole	35	40.33	15%
21	p21_d	Diastole	36	48.49	35%
22	p22_d	Diastole	27	44.08	63%

Table 5
Summary comparison of EF estimation errors between single-cycle and multi-cycle aggregation strategies. Metrics are derived from the cycle-level analyses reported in the Supplementary Material.

Metric	Single-cycle	Multi-cycle
Mean EF error (%)	1.09	0.92
MAE EF (%)	6.26	5.09
Median EF error (%)	1.80	1.58
SD EF error (%)	8.46	6.87

By averaging EF values across several cycles, the proposed approach mitigates inconsistencies, resulting in more stable and physiologically coherent EF estimates.

To further quantify the effect of this multi-cycle aggregation strategy, the complete cycle-level measurements and patient-level EF estimates are reported in the Supplementary Material (Tables S1–S3). These tables provide the full set of ventricular measurements extracted from the echocardiographic videos, including EF values obtained from individual cardiac cycles and the corresponding errors relative to the clinician-reported EF measurements.

Based on these detailed analyses, summary error metrics were computed to compare single-cycle EF estimation with the proposed multi-cycle aggregation strategy. The resulting values are reported in Table 5.

Overall, the multi-cycle approach reduces both the magnitude and variability of the estimation error. In particular, the Mean Absolute Error decreases from 6.26% in the single-cycle configuration to 5.09% with multi-cycle aggregation. Similarly, the standard deviation of the EF error decreases from 8.46% to 6.87%, indicating improved stability of the EF estimates. A reduction is also observed for both the mean EF error (from 1.09% to 0.92%) and the median EF error (from 1.80% to 1.58%).

These results demonstrate that aggregating information across multiple cardiac cycles mitigates the variability, leading to more reliable EF estimates compared with single-cycle evaluation.

In clinical echocardiography, automated systems should ideally achieve estimation errors within the variability range observed between human operators in order to be considered clinically acceptable. The EF estimation errors observed in our experiments (5.09% for the multi-cycle configuration) fall within the variability range typically reported for manual echocardiographic measurements (7.6%–13.9%) [2]. These findings indicate that the automated estimates generated by CardioSmartAssist are comparable to those obtained during routine clinical assessment.

In addition to continuous EF estimation accuracy, the clinical relevance of automated measurements also depends on their ability to correctly identify clinically meaningful EF thresholds. In contemporary cardiology practice, thresholds such as LVEF < 40% (heart failure with reduced ejection fraction) and LVEF ≤ 35% (criteria for implantable cardioverter-defibrillator eligibility) represent key decision points for patient management.

Although the present study focuses primarily on quantitative agreement with clinician measurements, the observed EF error distribution suggests that the automated estimates are sufficiently accurate to support consistent classification around these clinically relevant thresholds. Future studies on larger datasets will further investigate decision consistency metrics, such as sensitivity and specificity for clinically actionable EF categories.

Finally, while execution speed is not the primary focus of the study, the system was designed to support a more streamlined workflow.

By providing consistent, frame-level quantitative outputs, the pipeline may help operators with varying levels of experience obtain more standardised assessments of cardiac function. Although this has not been formally evaluated, such standardisation could potentially contribute to reducing subjective variability in routine practice, pending validation in future clinical studies.

Within hospital settings, CardioSmartAssist could be integrated into the training regimens of medical specialists, given its portable nature that allows for immediate utilisation on a laptop or workstation. The system has been developed to facilitate the instant upload and analysis of echocardiographic videos extracted from ultrasound machines. This provides an interactive, real-time learning experience, allowing the user to navigate the software's modules, analyse its decision-making process, and compare their decisions with the real uploaded video to gain a critical perspective. This feature is valuable in teaching environments, where trainees can systematically refine their skills in cardiac function assessment, validate their interpretations against automated analysis, and develop a more standardised approach to LVEF estimation. The above-mentioned promising results will be further investigated, focusing on the research directions described in Section 5.

5. Limits and further developments

Despite the advancements of CardioSmartAssist, several limitations warrant further investigation, presenting opportunities for future research and improvements. When the model's segmentation of the endocardial border and the mitral valve plane diverges from that of expert cardiologists, it potentially affects volume estimation. To enhance segmentation accuracy, future work may incorporate advanced deep learning architectures, including spatiotemporal models and adversarial learning frameworks like Actor-Critic networks [40]. Additionally, attention mechanisms and transformer-based approaches could be beneficial in complex scenarios where traditional models struggle.

The disparity between absolute volume estimation and echocardiographic measurements using Simpson's rule indicates that automated volume calculations require further refinement to align with clinical standards. Future enhancements could involve integrating additional echocardiographic views, such as the two-chamber view, to enrich spatial information and improve accuracy and volumetric assessment of all four cardiac chambers, to support supporting a wider range of diagnostic applications.

Moreover, the limited size and heterogeneity of the CardioSmartSet dataset restrict the generalisability of the findings. Although the test cohort is small, the methodological validity is bolstered by evaluation against the EchoNet-Dynamic dataset, which comprises 10,030 annotated echocardiography videos, providing a robust sample for deep learning training and validation [20].

The present study was conceived as an initial feasibility study aimed at demonstrating the clinical viability of an integrated, continuous-learning echocardiographic workflow, rather than providing a definitive large-scale validation. Similar proof-of-concept approaches have been reported in early automated echocardiography studies [41].

Accordingly, the proprietary CardioSmartSet cohort currently represents a preliminary testing phase and includes 27 patients with limited pathological diversity. It serves as an initial independent clinical evaluation set distinct from the public EchoNet-Dynamic dataset [2] used during model development. The expert-in-the-loop continuous learning mechanism, while designed to enhance model performance, has to be validated on a sufficiently large number of clinician-corrected masks to assess its real-world impact, as continuous clinician interaction is increasingly recognised as a prerequisite for adapting AI models to clinical environments [39]. Consequently, this work should be viewed as a technical feasibility study rather than a clinically validated tool.

The current work did not include a formal evaluation of reporting time or workflow efficiency. Future research should therefore investigate the impact of CardioSmartAssist on clinical workflow performance, particularly by distinguishing between total processing time and clinician active time. Previous studies have shown that automated echocardiographic analysis may require a slightly longer total processing time than rapid manual measurements, but substantially reduces the active time required from clinicians because computational steps such as view

selection and contour extraction are performed automatically in the background [42]. Future studies should therefore quantify the potential reduction in clinician active time and evaluate how the integration of CardioSmartAssist affects reporting efficiency and diagnostic reproducibility across operators with different levels of experience [43]. Furthermore, prospective clinical evaluations should investigate the reliability of the system in supporting downstream clinical decisions, such as the classification of patients around clinically relevant ejection fraction thresholds (e.g., LVEF $\leq 35\%$) [41].

Future research should involve larger, multi-centre, and pathology-balanced cohorts to establish the clinical reliability and generalisability of CardioSmartAssist, including vendor-stratified analyses to fully assess the robustness of the framework across heterogeneous acquisition environments. Such studies should also include patients with irregular cardiac rhythms, atypical ventricular morphologies, and heterogeneous imaging quality in order to evaluate the robustness of the proposed multi-cycle analysis strategy under real-world conditions [2,44].

In addition, further investigations should systematically analyse possible failures the EF computation pipeline, including segmentation flicker, missing frames, papillary muscle inclusion, trabeculations, and poor acoustic windows. Future work should therefore quantify how frequently cardiac cycles are excluded due to detected anomalies and evaluate the effectiveness of the anomaly detection and cycle filtering mechanisms currently implemented in the framework.

Moreover, the development of explicit reliability indicators or quality-control scores that trigger clinician review when automated measurements are uncertain could further enhance the clinical credibility and safety of the proposed approach.

The next phase involves designing a clinical study to assess CardioSmartAssist's impact on medical workflows, particularly on the educational benefits for medical residents regarding diagnostic accuracy.

Addressing ethical challenges, such as algorithmic bias, is crucial and requires diverse training datasets to ensure accurate diagnoses across demographic groups. Enhancing the representation of age groups, genders, and pathological cases will promote a more generalisable model, mitigate class-imbalance risks, and improve robustness [18].

Additionally, systematic assessments of network retraining using clinician-gathered batch images will be essential. This includes analysing learning curves and performance improvements with each retraining iteration, as well as comparing the accuracy of automatically identified end-systolic and end-diastolic frames with those labelled by expert cardiologists (previously excluded from evaluation due to clinical time constraints, which may provide valuable insights into the algorithm's temporal precision).

A further contribution of this study is the potential integration of CardioSmartAssist into out-of-hospital settings (e.g., home visits). Future investigations may examine the potential integration of portable ultrasound devices utilised by healthcare providers, nurses, or general practitioners during routine home examinations into existing workflows. However, this integration necessitates dedicated studies on usability and diagnostic reliability to determine its feasibility and impact on accessibility for patients with limited mobility.

Finally, a new data-acquisition phase is planned to improve diversity in patient demographics, clinical conditions, and imaging characteristics for follow-up studies. By increasing representation across various groups, the study aims to develop a more generalisable model, reduce dataset-induced biases, and enhance the model's robustness and reliability across diverse patient populations and clinical scenarios.

6. Conclusions

The global health challenge posed by CVDs necessitates efficient, scalable, and accessible tools for evaluating cardiac function. Echocardiography, despite its safety and real-time diagnostic capabilities, remains operator-dependent, underscoring the need for automated and effective solutions.

The study introduces CardioSmartAssist, a DL-based framework for cardiac diagnosis that automates EF calculation across multiple cycles and improves cardiac function analysis by attenuating beat-to-beat variability and transient noise compared to traditional single-cycle assessment.

Importantly, CardioSmartAssist's EF calculations demonstrated a 10% deviation from clinical single-cycle estimates, which is lower than the reported inter-observer variability of 13.9% [2]. However, average discrepancies in volume estimation reached 28%, partly attributable to inter-operator variability during the calibration phase.

The segmentation process was based on MultiResUNet, which demonstrated superior clinical reliability and robustness, and was further enhanced through fine-tuning.

The system also includes a continuous learning mechanism, which iteratively enhances model performance by incorporating clinicians' corrected segmentations.

Moreover, no direct evaluation in extra-hospital environments was performed; however, the modular design suggests, pending validation, potential adaptability in home visits, potentially broadening access to cardiac diagnostics and supporting decentralised healthcare delivery.

Future developments will enhance dataset expansion, segmentation accuracy, volume estimation, integration of echocardiographic views, and system validation through clinical, multicentre, and cost-effectiveness studies.

Additional efforts will aim to improve the model's long-term performance and generalisability by mitigating algorithmic bias through dataset diversification and adaptive retraining strategies using expert-corrected feedback.

CardioSmartAssist, developed in collaboration with clinicians, offers a modular, user-centred design with a customisable graphical interface, utilising UML prior to GUI development to optimise complex clinical processes [45]. As a systematic methodology for visualising system architecture, processes, and interactions, UML facilitated compliance with user requirements and the identification of bottlenecks in healthcare processes [46].

CRediT authorship contribution statement

Francesca Giada Antonaci: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Piera Ciaramella:** Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Giorgia Marullo:** Writing – review & editing, Methodology, Formal analysis. **Luca Ulrich:** Writing – original draft, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Vincenza Papa:** Writing – review & editing, Validation, Methodology, Formal analysis. **Walter Grosso Marra:** Writing – review & editing, Validation, Formal analysis, Data curation, Conceptualization. **Alessandro Depaoli:** Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Riccardo Miraglia:** Writing – original draft, Visualization, Investigation, Formal analysis. **Sandro Moos:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Enrico Vezzetti:** Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization.

Code availability

To ensure reproducibility, the full codebase will be provided to researchers upon reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material, including more detailed information about patients, hardware/software configuration, and results, can be found online at https://doi.org/10.1016/j.bspc.2026.110363.

Data availability

The EchoNet dataset is public. For what concerns the proprietary dataset, it is currently not possible to share it due to restrictions.

References

[1] C.U. Eriksen, O. Rotar, U. Toft, T. Jørgensen, What is the effectiveness of systematic population-level screening programmes for reducing the burden of cardiovascular diseases?, in: WHO Health Evidence Network Synthesis Report, (71) WHO Regional Office for Europe, Health Evidence Network, 2021.

[2] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C.P. Langlotz, P.A. Heidenreich, R.A. Harrington, D.H. Liang, E.A. Ashley, J.Y. Zou, Video-based AI for beat-to-beat assessment of cardiac function, Nature 580 (7802) (2020) 252–256, http://dx.doi.org/10.1038/s41586-020-2145-8, Publisher: Nature Publishing Group. URL https://www.nature.com/articles/s41586-020-2145-8.

[3] X. Deng, H. Wu, R. Zeng, J. Qin, MemSAM: Taming segment anything model for echocardiography video segmentation, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2024, pp. 9622–9631, http://dx.doi.org/10.1109/CVPR52733.2024.00919, URL https://ieeexplore.ieee.org/document/10656025/.

[4] A. Madani, J.R. Ong, A. Tibrewal, M.R.K. Mofrad, Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease, Npj Digit. Med. 1 (1) (2018) 59, http://dx.doi.org/10.1038/s41746-018-0065-x, URL https://www.nature.com/articles/s41746-018-0065-x.

[5] P. Thavendiranathan, A.D. Grant, T. Negishi, J.C. Plana, Z.B. Popović, T.H. Marwick, Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes, J. Am. Coll. Cardiol. 61 (1) (2013) 77–84, http://dx.doi.org/10.1016/j.jacc.2012.09.035, URL https://linkinghub.elsevier.com/retrieve/pii/S0735109712052862.

[6] S. Adabag, K.K. Patton, A.E. Buxton, T.S. Rector, K.E. Ensrud, K. Vakil, W.C. Levy, J.E. Poole, Association of implantable cardioverter defibrillators with survival in patients with and without improved ejection fraction: Secondary analysis of the sudden cardiac death in heart failure trial, JAMA Cardiol. 2 (7) (2017) 767–774, http://dx.doi.org/10.1001/jamacardio.2017.1413.

[7] K. Kusunose, R. Zheng, H. Yamada, M. Sata, How to standardize the measurement of left ventricular ejection fraction, J. Med. Ultrason. 49 (1) (2022) 35–43, http://dx.doi.org/10.1007/s10396-021-01116-z, URL https://link.springer.com/10.1007/s10396-021-01116-z.

[8] G. Sanjeevi, U. Gopalakrishnan, R.K. Parthinarupothi, T. Madathil, Deep learning supported echocardiogram analysis: A comprehensive review, Artif. Intell. Med. 151 (2024) 102866, http://dx.doi.org/10.1016/j.artmed.2024.102866, URL https://linkinghub.elsevier.com/retrieve/pii/S0933365724001088.

[9] M. Alsharqi, W.J. Woodward, J.A. Mumith, D.C. Markham, R. Upton, P. Leeson, Artificial intelligence and echocardiography, Echo Res. Pr. 5 (4) (2018) R115–R125, http://dx.doi.org/10.1530/erp-18-0056, Publisher: Springer Science and Business Media LLC. URL https://echo.biomedcentral.com/articles/10.1530/ERP-18-0056.

[10] J. Zhang, S. Gajjala, P. Agrawal, G.H. Tison, L.A. Hallock, L. Beussink-Nelson, M.H. Lassen, E. Fan, M.A. Aras, C. Jordan, K.E. Fleischmann, M. Melisko, A. Qasim, S.J. Shah, R. Bajcsy, R.C. Deo, Fully automated echocardiogram interpretation in clinical practice: Feasibility and diagnostic accuracy, Circulation 138 (16) (2018) 1623–1635, http://dx.doi.org/10.1161/CIRCULATIONAHA.118.034338, URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.034338.

[11] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A.R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. Dhooge, L. Lovstakken, O. Bernard, Deep learning for segmentation using an open large-scale dataset in 2D echocardiography, IEEE Trans. Med. Imaging 38 (9) (2019) 2198–2210, http://dx.doi.org/10.1109/TMI.2019.2900516, URL https://ieeexplore.ieee.org/document/8649738/.

[12] J. Tromp, P.J. Seekings, C.-L. Hung, M.B. Iversen, M.J. Frost, W. Ouwerkerk, Z. Jiang, F. Eisenhaber, R.S.M. Goh, H. Zhao, W. Huang, L.-H. Ling, D. Sim, P. Cozzone, A.M. Richards, H.K. Lee, S.D. Solomon, C.S.P. Lam, J.A. Ezekowitz, Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study, Lancet Digit. Health 4 (1) (2022) e46–e54, http://dx.doi.org/10.1016/S2589-7500(21)00235-1, URL https://www.sciencedirect.com/science/article/pii/S2589750021002351.

[13] H. Reynaud, A. Vlontzos, B. Hou, A. Beqiri, P. Leeson, B. Kainz, Ultrasound video transformers for cardiac ejection fraction estimation, 2021, http://dx.doi.org/10.48550/arXiv.2107.00977, arXiv:2107.00977[cs]. URL http://arxiv.org/abs/2107.00977.

[14] F.M. Asch, N. Poilvert, T. Abraham, M. Jankowski, J. Cleve, M. Adams, N. Romano, H. Hong, V. Mor-Avi, R.P. Martin, R.M. Lang, Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert, Circ.: Cardiovasc. Imaging 12 (9) (2019) http://dx.doi.org/10.1161/circimaging.119.009303, Publisher: Ovid Technologies (Wolters Kluwer Health). URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.119.009303.

[15] I. Iqbal, K. Walayat, M.U. Kakar, J. Ma, Automated identification of human gastrointestinal tract abnormalities based on deep convolutional neural network with endoscopic images, Intell. Syst. Appl. 16 (2022) 200149, http://dx.doi.org/10.1016/j.iswa.2022.200149, URL https://linkinghub.elsevier.com/retrieve/pii/S2667305322000862.

[16] I. Iqbal, I. Ullah, T. Peng, W. Wang, N. Ma, An end-to-end deep convolutional neural network-based data-driven fusion framework for identification of human induced pluripotent stem cell-derived endothelial cells in photomicrographs, Eng. Appl. Artif. Intell. 139 (2025) 109573, http://dx.doi.org/10.1016/j.engappai.2024.109573, URL https://linkinghub.elsevier.com/retrieve/pii/S0952197624017317.

[17] A. Narang, R. Bae, H. Hong, Y. Thomas, S. Surette, C. Cadieu, A. Chaudhry, R.P. Martin, P.M. McCarthy, D.S. Rubenson, S. Goldstein, S.H. Little, R.M. Lang, N.J. Weissman, J.D. Thomas, Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use, JAMA Cardiol. 6 (6) (2021) 624–632, http://dx.doi.org/10.1001/jamacardio.2021.0185.

[18] B. Maturi, S. Dulal, S.B. Sayana, A. Ibrahim, M. Ramakrishna, V. Chinta, A. Sharma, H. Ravipati, Revolutionizing cardiology: The role of artificial intelligence in echocardiography, J. Clin. Med. 14 (2) (2025) 625, http://dx.doi.org/10.3390/jcm14020625, URL https://www.mdpi.com/2077-0383/14/2/625.

[19] W. Wang, J. Zhou, C. Zhang, W. Xing, S. Fan, X. Qu, Prototype bank-driven test-time adaptation for medical ultrasound image segmentation, Med. Phys. 53 (1) (2026) e70280, http://dx.doi.org/10.1002/mp.70280, _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.70280. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.70280.

[20] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C.P. Langlotz, P.A. Heidenreich, R.A. Harrington, D.H. Liang, E.A. Ashley, J.Y. Zou, EchoNet-dynamic: A large new cardiac motion dataset, 2020, Published in *Nature* (2020). Accessed: 2025-02-28. License: Publicly available for research use. URL https://echonet.github.io/dynamic/index.html#dataset.

[21] P. Maheshwari, et al., Echonet-frames-masks-dataset, 2024, License: Community Data License Agreement - Sharing - Version 1.0. URL https://www.kaggle.com/datasets/foghorn/echonet-frames-masks-dataset.

[22] N. Ibtehaz, M.S. Rahman, MultiResUNet : Rethinking the U-net architecture for multimodal biomedical image segmentation, Neural Netw. 121 (2020) 74–87, http://dx.doi.org/10.1016/j.neunet.2019.08.025, arXiv:1902.04049[cs]. URL http://arxiv.org/abs/1902.04049.

[23] J. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Trans. Med. Imaging 25 (8) (2006) 987–1010, http://dx.doi.org/10.1109/TMI.2006.877092, URL http://ieeexplore.ieee.org/document/1661695/.

[24] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, 2015, http://dx.doi.org/10.48550/arXiv.1505.04597, arXiv:1505.04597[cs]. URL http://arxiv.org/abs/1505.04597.

[25] Ultralytics, Ultralytics YOLO, 2024, URL https://github.com/ultralytics/ultralytics. (Accessed: 21 January 2024).

[26] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, C. Feichtenhofer, SAM 2: Segment anything in images and videos, 2024.

[27] D.W. Wahr, Y.S. Wang, N.B. Schiller, Left ventricular volumes determined by two-dimensional echocardiography in a normal adult population, J. Am. Coll. Cardiol. 1 (3) (1983) 863–868, http://dx.doi.org/10.1016/S0735-1097(83)80200-9, URL https://linkinghub.elsevier.com/retrieve/pii/S0735109783802009.

[28] R.B. Stamm, P. Martin, M.D. Charlottesuille, Two-dimensional echocardiugraphic measurement of left ventricular ejection fraction: Prospective analysis of what constitutes an adequate determination, Am. Heart J. (1982).

[29] J. Li, T.S. Denney, Left ventricular motion reconstruction with a prolate spheroidal B-spline model, Phys. Med. Biol. 51 (3) (2006) 517, http://dx.doi.org/10.1088/0031-9155/51/3/004.

[30] D.W. Wahr, Y.S. Wang, N.B. Schiller, Left ventricular volumes determined by two-dimensional echocardiography in a normal adult population, J. Am. Coll. Cardiol. 1 (3) (1983) 863–868, http://dx.doi.org/10.1016/s0735-1097(83)80200-9.

[31] C.D. Under, Meta's SAM 2: The AI that can segment anything, even video, 2024, URL https://medium.com/@cognidownunder/metas-sam-2-the-ai-that-can-segment-anything-even-video-61b6704b4a4f.

[32] YOLO11: Il futuro della visione AI | ultralytics, 2024, URL https://www.ultralytics.com/it/blog/ultralytics-yolo11-has-arrived-redefine-whats-possible-in-ai.

[33] G. Marullo, L. Ulrich, F.G. Antonaci, A. Audisio, A. Aprato, A. Massè, E. Vezzetti, Classification of AO/OTA 31A/B femur fractures in X-ray images using YOLOv8 and advanced data augmentation techniques, Bone Rep. 22 (2024) 101801, http://dx.doi.org/10.1016/j.bonr.2024.101801, URL https://www.sciencedirect.com/science/article/pii/S2352187224000688.

[34] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, Adv. Neural Inf. Process. Syst. 325 (2019).

[35] G. Wilson, D.J. Cook, A survey of unsupervised deep domain adaptation, ACM Trans. Intell. Syst. Technol. 11 (5) (2020) 1–46, http://dx.doi.org/10.1145/3400066, URL https://dl.acm.org/doi/10.1145/3400066.

[36] R.M. Lang, L.P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F.A. Flachskampf, E. Foster, S.A. Goldstein, T. Kuznetsova, P. Lancellotti, D. Muraru, M.H. Picard, E.R. Rietzschel, L. Rudski, K.T. Spencer, W. Tsang, J.-U. Voigt, Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American society of echocardiography and the European association of cardiovascular imaging, J. Am. Soc. Echocardiogr. 28 (1) (2015) 1–39.e14, http://dx.doi.org/10.1016/j.echo.2014.10.003, URL https://linkinghub.elsevier.com/retrieve/pii/S0894731714007457.

[37] C. Otto, R. Bragato, Ecocardiografia clinica: Sesta edizione, Edra, 2019, URL https://books.google.it/books?id=TAKfDwAAQBAJ.

[38] P. Lancellotti, B. Cosyns, ESC, EACVI, Manuale di Ecocardiografia della EACVI, vol. Unico, Momento Medico, 2018, p. 353.

[39] Y. Li, S. Ercisli, Explainable human-in-the-loop healthcare image information quality assessment and selection, CAAI Trans. Intell. Technol. n/a (2023) http://dx.doi.org/10.1049/cit2.12253, _eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/cit2.12253. URL https://onlinelibrary.wiley.com/doi/abs/10.1049/cit2.12253.

[40] N. Araslanov, C. Rothkopf, S. Roth, Actor-critic instance segmentation, 2019, http://dx.doi.org/10.48550/arXiv.1904.05126, arXiv:1904.05126[cs]. URL http://arxiv.org/abs/1904.05126.

[41] F.M. Asch, N. Poilvert, T. Abraham, M. Jankowski, J. Cleve, M. Adams, N. Romano, H. Hong, V. Mor-Avi, R.P. Martin, R.M. Lang, Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert, Circ. Cardiovasc. Imaging 12 (9) (2019) e009303, http://dx.doi.org/10.1161/CIRCIMAGING.119.009303, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7099856/.

[42] K.M. Sveric, R. Botan, Z. Dindane, A. Winkler, T. Nowack, C. Heitmann, L. Schleußner, A. Linke, Single-site experience with an automated artificial intelligence application for left ventricular ejection fraction measurement in echocardiography, Diagnostics 13 (7) (2023) 1298, http://dx.doi.org/10.3390/diagnostics13071298, URL https://www.mdpi.com/2075-4418/13/7/1298.

[43] S. Akil, J. Castaings, P. Thind, T. Åhlfeldt, M. Akhtar, A.T. Gonon, M. Quintana, K. Bouma, Impact of experience on visual and Simpson's biplane echocardiographic assessment of left ventricular ejection fraction, Clin. Physiol. Funct. Imaging 45 (1) (2025) e12918, http://dx.doi.org/10.1111/cpf.12918, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cpf.12918. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cpf.12918.

[44] K.V. Bunting, K. O'Connor, R.P. Steeds, D. Kotecha, Cardiac imaging to assess left ventricular systolic function in atrial fibrillation, Am. J. Cardiol. 139 (2021) 40–49, http://dx.doi.org/10.1016/j.amjcard.2020.10.012, URL https://www.sciencedirect.com/science/article/pii/S000291492031081X.

[45] F. Pecoraro, D. Luzi, Using unified modeling language to analyze business processes in the delivery of child health services, Int. J. Environ. Res. Public Health 19 (20) (2022) 13456, http://dx.doi.org/10.3390/ijerph192013456, URL https://www.mdpi.com/1660-4601/19/20/13456.

[46] E.C.A.D. Carvalho, M.K. Jayanti, A.P. Batilana, A.M.O. Kozan, M.J. Rodrigues, J. Shah, M.R. Loures, S. Patil, P. Payne, R. Pietrobon, Standardizing clinical trials workflow representation in UML for international site comparison, in: G. Malaga (Ed.), PLoS ONE 5 (11) (2010) e13893, http://dx.doi.org/10.1371/journal.pone.0013893, URL https://dx.plos.org/10.1371/journal.pone.0013893.