



A Multimodular AI Algorithm for Automated Assessment of Left Ventricular Function in Ischemic Heart Disease: Ejection Fraction, Wall Motion, and Regional Myocardial Segmentation

Sidem Gül¹, Reşit Taşdemir², Beyza Açıkgöz², Hakan Duman³, Hamza Hodzic⁴, Sena Köker⁵,
 Nazlı Erdemir⁶, Mehmet Kıvrak⁷

¹Department of Medicine, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

²Department of Software Engineering, Malatya Turgut Özal University Faculty of Engineering and Natural Sciences, Malatya, Türkiye

³Department of Cardiology, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

⁴Turkish Red Crescent Blood Donation Center, Kastamonu, Türkiye

⁵Department of Medicine, İstanbul University Faculty of Medicine, İstanbul, Türkiye

⁶Department of Biomedical Engineering, Kocaeli University Faculty of Technology, Kocaeli, Türkiye

⁷Department of Biostatistics and Medical Informatics, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

Background: Ischemic heart damage reduces the pumping efficiency of the heart by affecting the left ventricular ejection fraction (LVEF) and causing wall motion abnormality (WMA). In daily clinical practice, these parameters are interpreted by physicians using two dimensional transthoracic echocardiography (2D-TTE). Because 2D-TTE reports rely on visual evaluations, they are subject to experience-based limitations and exhibit low reproducibility.

Aims: To develop an artificial intelligence algorithm composed of two modules that enable automatic LVEF calculation and WMA detection for analyzing 2D-TTE images.

Study Design: Diagnostic accuracy study.

Methods: A total of 600 adult patients were retrospectively included. The model combined static frame segmentation with dynamic tracking using a hybrid Simpson's method applied to apical 2- and 4-chamber views. Model performance was assessed against cardiologist measurements using Bland-Altman analysis. The YOLOv8 and ResNet50 models were

employed for the wall motion module. Performance metrics, including accuracy, precision, F1 score, and area under the curve, were evaluated.

Results: In the Bland-Altman analysis, the mean bias between the LVEF module and cardiologist measurements was -4, with limits of agreement ranging from -15 to -3. Regression analysis demonstrated a strong correlation between the LVEF module and cardiologist measurements ($r = 0.71, p < 0.001$). In the wall motion module, the YOLOv8 segmentation model exhibited high accuracy, while ResNet50 achieved superior performance with an accuracy of 95%. The algorithm's color coding contributed to standardized interpretation among operators, enhancing consistency.

Conclusion: This is the first study to integrate automated EF calculation and WMA detection within a single workflow. SafeHeart offers accurate, reproducible, and rapid analysis, with the potential to support routine echocardiography practice. Color-coded region segmentation can facilitate more standardized and reliable results.



Corresponding author: Sidem Gül, Department of Medicine, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

e-mail: sidem_gul21@erdogan.edu.tr

Received: August 18, 2025 **Accepted:** October 12, 2025 **Available Online Date:** January 02, 2026 • **DOI:** 10.4274/balkanmedj.galenos.2025.2025-8-160

Available at www.balkanmedicaljournal.org

ORCID iDs of the authors: S.G. 0009-0000-0379-1002; R.T. 0009-0008-9369-2910; B.A. 0009-0007-6770-0588; H.D. 0000-0002-1441-7320; H.H. 0009-0007-6940-9006; S.K. 0009-0006-2628-5181; N.E. 0009-0005-8517-5102; M.K. 0000-0002-2405-8552.

Cite this article as: Gül S, Taşdemir R, Açıkgöz B, et al. A Multimodular AI Algorithm for Automated Assessment of Left Ventricular Function in Ischemic Heart Disease: Ejection Fraction, Wall Motion, and Regional Myocardial Segmentation. *Balkan Med J.*; 2026; 43(1):25-37.

Copyright@Author(s) - Available online at <http://balkanmedicaljournal.org/>

INTRODUCTION

Ischemic heart damage (IHD) remains one of the leading causes of global mortality and disability, accounting for approximately 9.1 million deaths and 197 million cases in 2019.¹ Reduced myocardial perfusion in IHD results in regional systolic contractility disorders. The left ventricular ejection fraction (LVEF), which measures the percentage of blood expelled from the left ventricle during per cardiac cycle, is consequently impaired.² In clinical settings, two dimensional transthoracic echocardiography (2D-TTE) images, particularly A4C (Apical 4 Chamber) and A2C (Apical 2 Chamber) views, are interpreted by experienced physicians to estimate LVEF value, which is then quantified using techniques such as the SDSM.³ These evaluations are often time-consuming and demand considerable clinical expertise. In recent years, artificial intelligence (AI) algorithms have been developed to support preliminary analysis of TTE images, encompassing tasks such as interpretation, segmentation, and ventricular EF calculation.⁴⁻⁶ AI is increasingly expected to assist clinicians in accurately identifying A2C and A4C images.⁷

Multiple shape-based computational approaches, including the prolate ellipsoid, truncated ellipsoid, area-length, and SDSM methods, have been proposed for LVEF calculation from 2D-TTE. Among these, deep learning (DL), a specialized domain within machine learning, has emerged as the preferred strategy for detecting the left ventricle and computing its volume in 2D-TTE.^{8,9} While earlier studies attempted to predict LVEF directly from raw TTE video sequences using DL models, contemporary AI frameworks generally emphasize model training on sequential frames extracted from video segments. Within these frameworks, SDSM continues to serve as a widely adopted technique for deriving numerical LVEF values.^{9,10} In this method, after identifying the target ventricular cavity, virtual disks are placed within it to estimate volume. Parameters such as disk number, radius, or height are determined by accounting for factors like spatial orientation, flexibility, accuracy, measurement precision, and algorithmic complexity. To ensure precise alignment, the number or height of disks can be predefined according to anatomical landmarks such as the apex and mitral valve regions.⁹ DL models trained on direct video data are now being employed to develop AI systems capable of distinguishing pathological from normal images for wall motion tracking.¹¹ Myocardial segmentation across sequential frames enables the delineation and masking of anatomical regions within the LV wall. Further research is needed to incorporate color-coded segmentation into video sequences to enhance visualization in LV wall regional analysis.

In this study, we introduce SafeHeart, an AI-based system designed for automated analysis of EF and wall motion using 2D echocardiography. The novelty of this work lies in combining a hybrid frame-video processing strategy with dual apical views (A2C and A4C), allowing robust EF estimation and wall motion abnormality (WMA) classification within a unified workflow. Our SafeHeart model, trained on A2C and A4C datasets, autonomously identifies ventricular cavities and tracks wall motion in continuous video sequences, thereby minimizing dependence on operator expertise. We aim to enhance the precision of LVEF estimation by

refining the area and volume computation process. For wall motion analysis, we propose incorporating standardized anatomical region definitions into 2D-TTE systems to objectively identify hypoactive regions and distinguish pathological from healthy images (Figure 1). Because LVEF and wall motion abnormalities (WMAs) are closely related, integrating both parameters within a single AI framework can reduce measurement discrepancies and offer advantages over algorithms that assess only one parameter. This integration can accelerate workflow and reduce reliance on experience-based subjective interpretation.

MATERIALS AND METHODS

Dataset

The study protocol was approved by the Recep Tayyip Erdoğan University Non-Interventional Clinical Research Ethics Committee (approval number: 2024/143, date: 13.06.2024). For this retrospective study, recordings from the Vivid E95 2D TTE device in the Adult Echocardiography Clinic of Recep Tayyip Erdoğan University Training and Research Hospital were utilized. TTE images of 600 adult patients aged 18 years and older were transferred to a computer in DICOM format and subsequently anonymized. To ensure dataset adequacy, open-access A4C echocardiographic images from Stanford University were incorporated (<https://stanfordaimi.azurewebsites.net/datasets/834e1cd1-92f7-4268-9daa-d359198b310a>). These supplementary images were in AVI format. Both datasets were reviewed according to predefined criteria, which required each video to include at least one complete heartbeat cycle and a fully visible left ventricle. Data did not meet these standards were excluded from the final dataset. All images were standardized by converting them into MP4 format before being imported into the analysis interface. In several cases, the left ventricle appeared on the opposite side of the screen due to probe orientation or device configuration. These images were flipped to the correct side to comply with international conventions. The data were then categorized and labeled according to A2C and A4C views. Distinct training strategies were employed for the LVEF module, the wall motion prediction module, and the anatomical wall segmentation module. Prior to segmentation, observers completed a structured two-month echocardiography training program. Following this training, myocardial region segmentation was performed by the trained observers under supervision. All segmentation outputs were subsequently reviewed and validated by a cardiologist with over 20 years of clinical echocardiography experience. LVEF measurements were independently evaluated and confirmed by the same senior cardiologist

LVEF module

For the LVEF module, A4C and A2C videos were utilized. The average duration of each patient video was 1.2 seconds (ranging from 0.6 to 4 seconds). During import into the interface, longer videos were trimmed to include at least one complete heartbeat cycle (one diastole and one systole). Image frames were extracted from each video at a rate of 25 frames per second (e.g., 2 seconds = 48 frames). The endocardial layer of the left ventricle

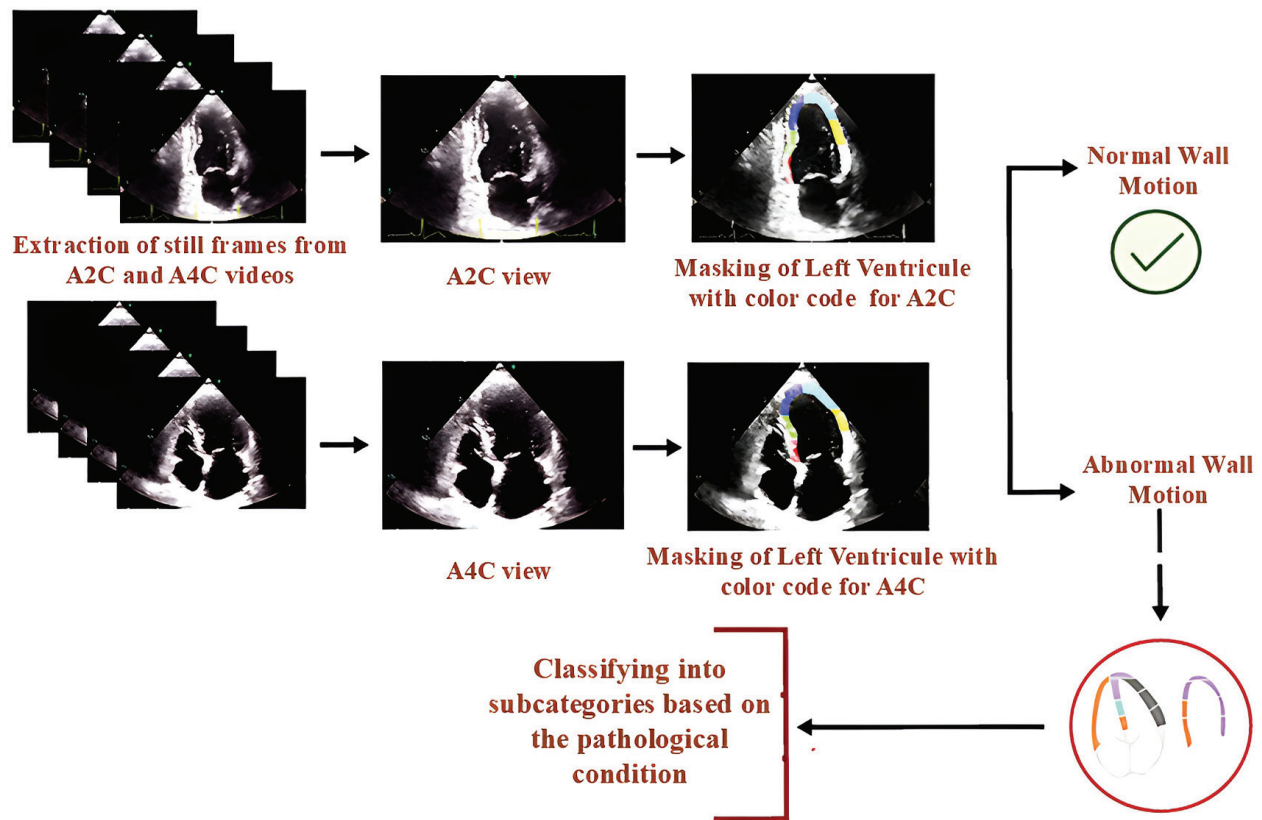


FIG. 1. Into image slices, each containing a full cardiac cycle. The myocardial layer corresponding to the left ventricular wall is identified in the A4C and A2C images. The wall is masked into seven anatomical regions. In the deep learning model, pathological and normal wall motion tracking is performed. Pathological wall motion is detected and classified into subcategories based on the pathological condition.

A4C; apical 4 chamber, A2C; apical 2 chamber.

in A2C and A4C images was detected using the Roboflow interface. Subsequently, the ventricular cavity was segmented using semantic segmentation and labeled accordingly. During segmentation, anatomical landmarks, specifically, the indentations between the mitral valve and ventricular walls, as well as the apex of the heart, were identified but not labeled separately. A dataset consisting of 1,502 image slices was created to train the algorithm on the target left ventricular cavity. The dataset was divided into training, validation, and test sets (Table 1). Essential preprocessing steps, including normalization and resizing, were applied. The YOLOV image processing algorithm was used to define the volume calculation area on the front end.

The systolic and end-diastolic phases, representing the minimum and V_{\max} phases of the left ventricle, were identified from the processed video sequences. The left ventricular area was calculated on a pixel basis, and the phase areas were then processed accordingly. The phase areas obtained from the A4C and A2C angles were mapped onto planes according to the SDSM (Figure 2). The cardiac volume was calculated as the sum of the volumes of multiple disks using the Riemann sum approach. Each disk was modeled with a specific height (disk thickness) along the heart's long axis and with a radius derived from the masked area.

First, the areas of each slice (disk) was calculated from the masked images. These areas were then used to determine the radii of the disks according to the following formula.

$$r_i = \sqrt{\frac{A_i}{\pi}} \quad (1)$$

where A_i represents the area of the i 'th slice, and r_i denotes the radius of that slice.

SDSM was employed to calculate the overall volume. To volume of the heart region was determined based on the height (thickness) and radius of each disk, and the total volume was computed using:

$$V = \sum_{i=1}^n \pi \cdot r_i^2 \cdot h_i \quad (2)$$

where n is the number of disks, r_i is the radius of the i 'th disk, h_i is its thickness (height).

TABLE 1. Data Distribution.

Wall motion tracking module				
	Patients (n)	A2C (frame)	A4C (frame)	Total (frame)
Train set	122	738	716	1454
Validation set	15	213	206	419
Test set	16	103	101	204
Total	153	1054	1023	2077
LVEF modul				
	Patients (n)	A4C (frame)	A4C (frame)	Total (frame)
Train set	92	1170	1052	2222
Validation set	26	334	306	640
Test set	13	168	144	312
Total	131	1672	1052	3174

LVEF, left ventricular ejection fraction, A4C; apical 4 chamber, A2C; apical 2 chamber.

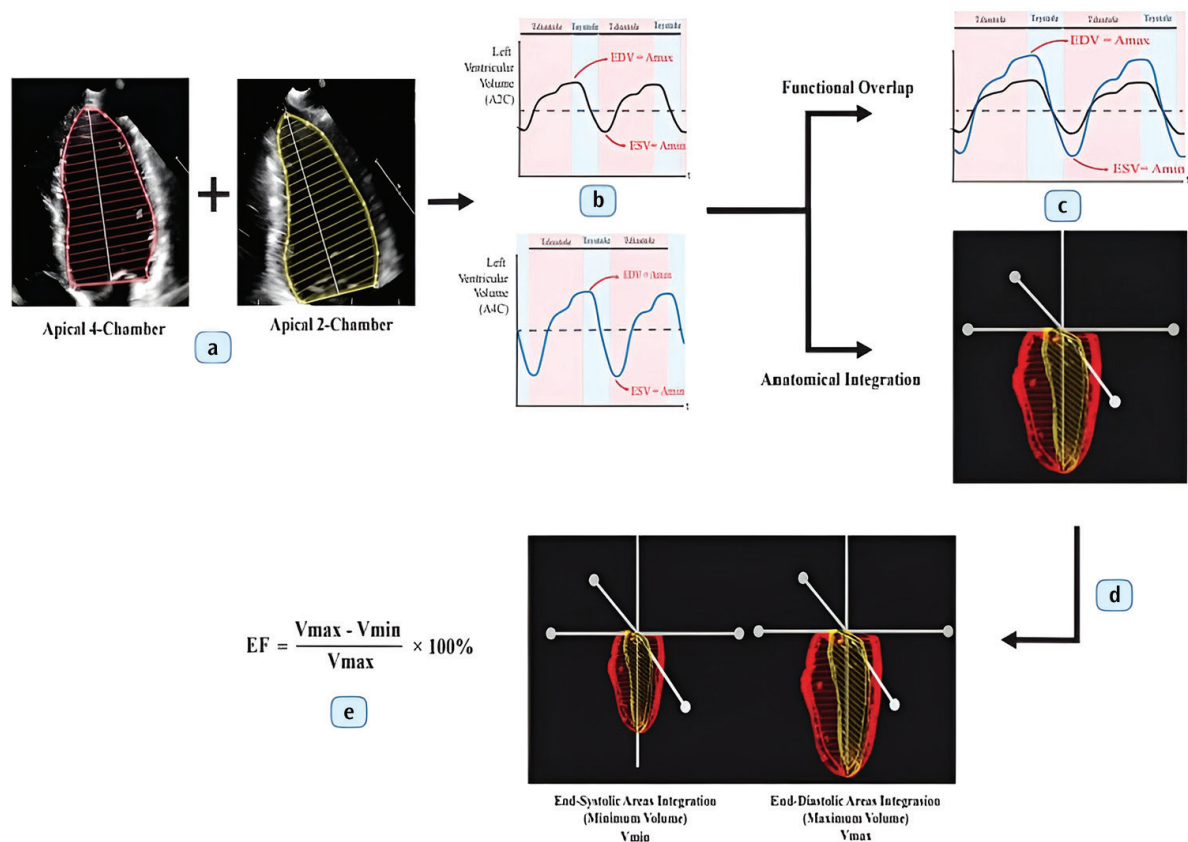


FIG. 2. This figure illustrates the fundamental steps in the LVEF calculation process. (a) The boundaries of the left ventricle cavity are detected in A4C and A2C images. (b) The area change during the cardiac cycle, which includes one systolic and one diastolic phase, is tracked through two image angles. (c) The phase difference between the area change functions is eliminated, and the A4C and A2C images are integrated in the x-y-z plane. (d) The functional change is integrated into the anatomical structure, and the area changes are tracked along three axes. The Amax- Amin values are determined using Riemann and Simpson mathematical calculations as $V_{max} - V_{min}$. (e) The percentage change between the identified V_{max} and V_{min} is calculated, and the EF value is obtained.

LVEF, left ventricular ejection fraction, A4C; apical 4 chamber, A2C; apical 2 chamber.

From the masked images, separate calculations were made to determine the maximum and minimum volumes (V_{min}). The maximum volume (V_{max}) was obtained from the image with the widest ventricle (typically at diastole), and the V_{min} was derived from the image with the narrowest ventricle (typically at systole). Integration of these phases was performed by considering both functional and anatomical phase harmony (Figure 3).

After volume computation, the EF was calculated using the following formula:

$$EF = \frac{V_{max} - V_{min}}{V_{max}} \times 100 \quad (3)$$

This formula determines the percentage of blood ejected from the heart during each cardiac cycle. EF is a standard metric used to evaluate overall cardiac performance.

Wall motion module and anatomical segmentation

Masking was performed using color codes to segment the anatomical regions of the left ventricular wall. The 17-segment model of the left ventricle, as defined by the American Society of Echocardiography (ASE), was adopted as the reference framework for wall region

segmentation.¹² In A4C images, the apical cap, apical septum, and apical lateral regions were tracked using the same color code, representing as a single unified region (Table 2). Distinct color masks were created for each of the mid inferoseptum, basal-inferoseptum, mid anterolateral, and basal-ant anterolateral regions. In the A2C view, separate color masks were generated for the inferior, anterior, and apical regions of the left ventricular wall. Consequently, a total of 7 regional segmentations were established, 5 components for A4C and 3 components for A2C, with the apical region shared between both views.

To determine whether wall motion was normal or pathological, the myocardial segments in the dynamic images were classified as hypokinetic or normal. The development of this module followed the procedural steps outlined below.

Segmentation

The myocardial segmentation process was performed using the YOLOv8 model, which is distinguished in the field of medical imaging for its fast and accurate segmentation capabilities. The model's real-time processing ability, flexible usability, and high accuracy rates offer significant advantages, particularly in time-critical segmentation applications. YOLOv8's architecture, optimized for object detection and segmentation tasks, allows it to perform

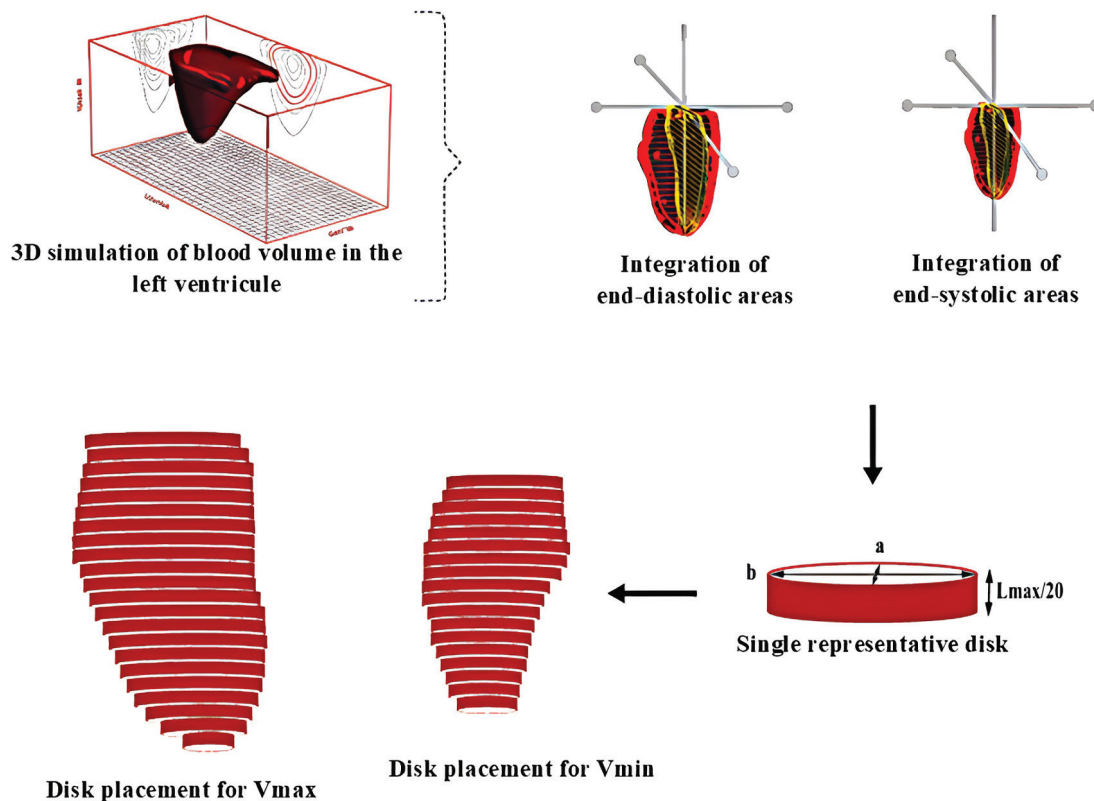


FIG. 3. This figure illustrates the calculation of the change in the left ventricular intracavity blood volume. After transferring the maximum and minimum areas obtained from the Riemann Sum of the A4C and A2C images to the x-y-z plane, volumetric calculation is performed using the Simpson's Disk method.

3D; three dimensional, A4C; apical 4 chamber, A2C; apical 2 chamber.

TABLE 2. Wall Motion Tracking Methods: Methods and Models Used for Wall Motion Tracking.

Wall motion tracking	Myocardial segmentation		
	Algoritma/model	File type	Method
	YOLOv8	MP4 and MKV	Classification, segmentation
	Motion status detection		
	Algoritma/model	File type	Method
	Xception	PNG and JPG	Transfer learning, depthwise separable convolution
	VGG16	PNG and JPG	Transfer learning, convolutional layers
	MobileNet	PNG and JPG	Transfer learning, depthwise separable convolution
	DenseNet	PNG and JPG	Transfer learning, dense connectivity
	EfficientNet	PNG and JPG	Transfer learning, compound scaling
	ResNet50	PNG and JPG	Transfer learning, residual connections

effectively across a wide range of applications. Furthermore, its capacity to maintain high accuracy even with limited datasets has made it a preferred choice in medical imaging research.^{13,14} As a result of the segmentation process, the left ventricle was divided into seven primary regions, and wall motion analysis was conducted for each. Segmentation represents a fundamental step in the precise classification of WMAs.

The images obtained after segmentation were analyzed using commonly employed DL models, as reported in previous studies.^{14,15} Several transfer learning-based architecture, such as Xception, VGG16, MobileNet, DenseNet, EfficientNet, and ResNet50, were trained, and compared in terms of performance. The transfer learning approach was selected for classifying WMAs due to its demonstrated ability to achieve high accuracy rates with limited datasets.¹⁶

The segmentation performance of YOLOv8, combined with the classification effectiveness of transfer learning-based models, provides a robust and efficient solution for detecting and analyzing myocardial WMAs.

Model training process

During the training process, we utilized the Python programming language along with the TensorFlow and PyTorch software libraries. The data was divided into 80% for training and 20% for testing. Data augmentation techniques were applied throughout the training phase to enhance model robustness. Early stopping and learning rate reduction methods were implemented in accordance with the literature, ensuring optimization of the model's learning processes.¹⁷ The training was accelerated using NVIDIA Tesla V100 GPUs.

To compare model performance, Xception, VGG16, MobileNetV2, DenseNet, EfficientNet, and ResNet models were trained. Among these, the highest accuracy rate was achieved with the ResNet model. The superior performance of ResNet is attributed to its widespread adoption in DL applications and its consistently strong results reported in the literature.¹⁵ Consequently, ResNet was

selected as the primary model for this study due to its highest accuracy rate.

Statistical analysis

To assess the agreement between measurement methods, we employed correlation analysis, regression analysis, Bland-Altman analysis, and Passing-Bablok and Deming regression methods. Through these analyses, the absolute and correlational agreement between measurements was examined, and the clinical interchangeability of the methods was evaluated.

In DL model training, the dataset was divided into training/validation subsets, and model performance was assessed using loss and accuracy metrics. Hyperparameter optimization was conducted by monitoring potential overfitting and under fitting conditions.

RESULTS

LVEF module performance

In this study, the concordance and reliability of the developed method with the reference test were evaluated using various statistical methods. Bland-Altman analysis was performed to identify systematic differences between measurements. The analysis revealed a mean bias of -4.0%, with limits of agreement ranging from -15% to -3% (Table 3). This result indicates that, on average, the algorithm slightly underestimates the EF compared with expert measurements. Despite this bias, 95% of the values fell within clinically acceptable limits, suggesting practical reliability for routine use. The statistical agreement between the measurements was found to be significant ($p = 0.044$, $p < 0.05$). These findings demonstrate that the developed AI algorithm closely matches manual measurements in EF calculations.¹⁸

Deming regression analysis demonstrated a strong linear correlation between the LVEF module values and manual reference measurements ($r = 0.71$, $p < 0.001$), with the regression equation

TABLE 3. Comparison of Reference LVEF Values with SafeHeart Software Results.

	Differences	%95 confidence interval
Median	-4	[-6, -3]
Lower limit	-15	[-16, -14]
	Referances LVEF	Results of SafeHeart
Count	165	165
Mean	58.679	62.903
Standard deviation	8.03	9.534
Median	60	64
Interquartile range	5	9

This table shows the comparison between reference LVEF values and the results obtained using the SafeHeart software from 165 patients. The average LVEF values for both groups are 58.679 (reference) and 62.903 (SafeHeart), respectively. The standard deviation is 8.03 for the reference measurements and 9.534 for SafeHeart, indicating greater variation in the SafeHeart software measurements. The median values are 60 (reference) and 64 (SafeHeart), suggesting a slight increase in the software's measurements. The interquartile range is 5 for the reference values and 9 for the SafeHeart measurements, indicating a wider distribution in the software's measurements. LVEF, left ventricular ejection fraction.

defined as $y = -4.289 + 1.151x$. The inclusion of zero within the 95% confidence interval (CI) for the intercept indicates the absence of systematic measurement error. Furthermore, the Breusch-Pagan test for heteroscedasticity yielded a p value of 0.57, confirming that residual variance was homoscedastic across the measurement range. Additional agreement metrics further validated the performance of the algorithm.

Additional agreement metrics reinforced the LVEF module's reliability. The intraclass correlation coefficient (ICC) was 0.85 for single measurements and 0.78 for average measurements, both with 95% CIs indicating moderate to good reliability (e.g., ICC_single 95% CI 0.39-0.79, $p < 0.001$) (Table 4).

The concordance correlation coefficient (CCC) between the AI and cardiologist EF values was 0.85 (95% CI 0.56-0.72), reflecting a high level of absolute agreement. For context, a CCC of 1 would indicate perfect concordance; thus, a value of 0.85 suggests the model's EF outputs closely track the reference measurements, albeit with some variability. Notably, the CCC's components demonstrated good accuracy ($p^* = 0.74$) and precision ($C_b = 0.88$) in EF estimation (Table 5). Collectively, these results demonstrate that the SafeHeart LVEF

module achieves clinically acceptable accuracy and reproducibility in EF estimation. In practical terms, the model's EF predictions are, on average, 4% lower than manual measurements but remain within a reasonable error margin for clinical application. Although this bias should be considered (see Discussion), the narrow CIs and strong correlations indicate that the AI's EF measurements could reliably complement manual measurements in routine clinical practice.

Wall motion abnormality module performance

All performance metrics are reported with 95% CIs. For example, the 95% CI for accuracy was approximately 91.5-97.5% (binomial proportion CI), reflecting high certainty in the model's performance given the test set size ($n = 204$ frames). Similarly, precision and recall had 95% CIs of approximately ± 3 –4%, consistently placing these metrics above 90%. The model's area under the receiver operating characteristic (ROC) curve [area under the curve (AUC)] was 0.96, with a 95% CI of approximately 0.93-0.99, demonstrating excellent discriminative ability. This AUC is comparable to the best results reported in the literature for automated wall motion detection (Figure 4).¹⁹

To ensure full transparency of classification results, Table 6 presents the confusion matrix for the ResNet50 WMA classifier on the test set. Out of 204 total evaluations, 194 were correctly classified by the model. It identified approximately 86 true positives (TP = pathological segments correctly identified) and 108 true negatives (TN = normal segments correctly identified), with only about 5 false positives (FP = normal segments misclassified as abnormal) and 5 false negatives (FN = abnormal segments misclassified as normal). This corresponds to a specificity of approximately 96% [$TN/(TN + FP)$] and a negative predictive value (NPV) of about 95%, in addition to the high sensitivity and precision noted above. In summary, the classifier performs exceptionally well on both normal and abnormal cases—it rarely misses WMAs and seldom mislabels normal wall motion as abnormal. The balanced performance across sensitivity, specificity, PPV, and NPV highlights the model's robustness for this binary classification task. For completeness, all these metrics with their CIs are summarized as follows: sensitivity 93.5% (95% CI ~ 90 –97%), specificity 96.0% (≈ 94 –99%), PPV 94.2% (≈ 90 –97%), and NPV 95.0% (≈ 92 –98%). Such consistently high values suggest that the model's predictions can be trusted in clinical practice for screening WMAs (Figure 5).

Each cell in the table represents the number of segments (or segment-equivalents from frames) classified into that category. As shown, 108 normal instances were correctly identified as normal,

TABLE 4. Comparison of Results for Intraclass Correlation Coefficient (ICC) and F Value.

Measurements	Intraclass correlation coefficient	Selected confidence interval		F value corresponding to test value			
		Lower level	Upper level	Value	df1	df2	p value
Single measures	0.6514	0.3609	0.7951	6.2752	164	15.3064	< 0.001
Average measures	0.7889	0.4954	0.89	6.2752	164	13.103	< 0.001

Presents the results of the agreement between SafeHeart software and reference measurements, evaluated using the ICC and F value. The ICC value for individual measurements is 0.6514, and for mean measurements, it is 0.7889. The p value for both measurement types is < 0.001 , indicating that the agreement is statistically significant.

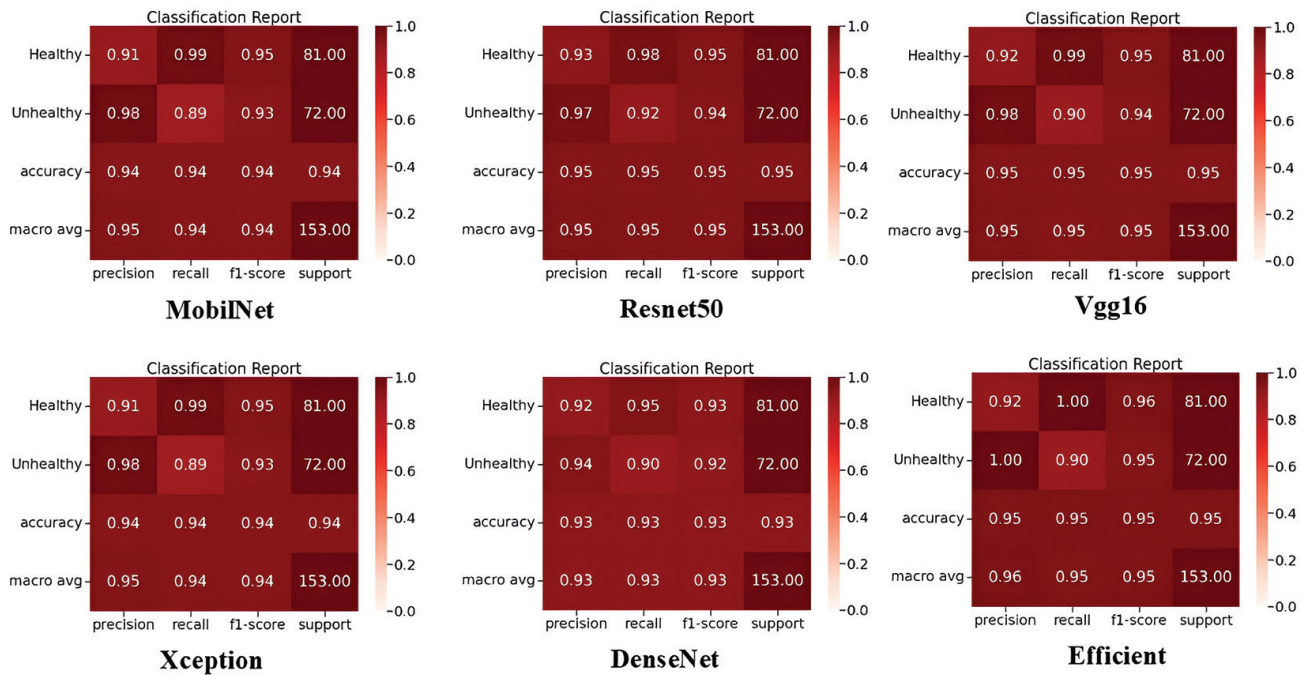


FIG. 4. Statistical result of models: classification reports are shown for each model.

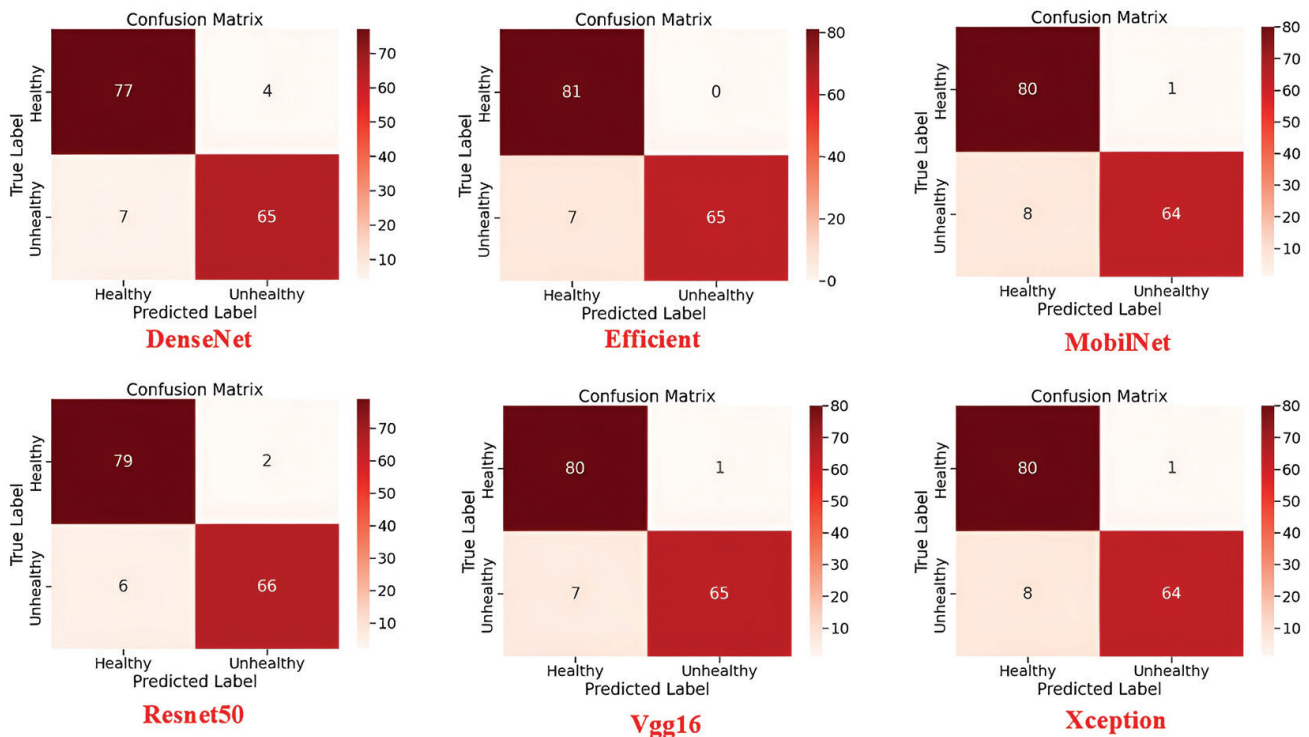


FIG. 5. Confusion matrices of models.

while 85 abnormal instances were correctly identified as abnormal. There were 5 false alarms (normal segments labeled as abnormal) and 6 misses (abnormal segments labeled as normal), very low error counts given the dataset size. This tabular presentation complements

the visual confusion matrix in Figure 5 of the original submission, providing precise numerical values. It also enables calculation of additional performance metrics such as specificity = $108/(108 + 5) \approx 95.6\%$ and NPV = $108/(108 + 6) \approx 94.7\%$, as shown in Table 6.

TABLE 5. Concordance Correlation Analysis Between Reference LVEF and SafeHeart-Derived LVEF Measurements.

	Concordance coefficient	%95 Lower limit	%95 Upper limit	ρ^* (accuracy)	C_b (precision)
Referances LVEF/SafeHeart	0.85	0.5649	0.7214	0.7358	0.8834

This table shows the concordance correlation coefficient (CCC) and its components between the reference left ventricular ejection fraction (LVEF) measurements and those obtained using the SafeHeart software. The CCC value is calculated as 0.65, indicating a moderate level of agreement between the two methods. The 95% confidence interval for the CCC is 0.5649-0.7214, suggesting that similar levels of agreement can be maintained across different datasets, though some variability may exist. The accuracy metric, ρ^* , is calculated as 0.7358, indicating that the SafeHeart software demonstrates reasonable accuracy compared to the reference method. The precision metric, C_b , is found to be 0.8834, reflecting a high level of consistency in the SafeHeart software's measurements. CCC is the product of the Pearson correlation coefficient (ρ) and the accuracy component (C_b), i.e., $CCC = \rho \times C_b$. Therefore, the values of ρ^* and CCC differ, as they represent distinct but related quantities.

TABLE 6. Confusion Matrix for Wall Motion Classification (ResNet50) (Values are From the Test Set, Comparing Model Predictions to the Cardiologist Reference).

Actual\predicted	Normal	Pathological
Normal segments	108 (TN)	5 (FP)
Pathological segments	6 (FN)	85 (TP)

These metrics confirm that the model not only detects pathology with high sensitivity but also confidently identifies normal wall motion, an essential factor in avoiding over-diagnosis. The reported accuracy of 95% represents the overall proportion of correct classifications. Importantly, no single performance metric should be considered in isolation-the model exhibits consistently high sensitivity and specificity, characteristics desirable in a diagnostic AI tool.

Beyond its discrimination ability, we also evaluated the calibration of the ResNet50 model's probability outputs for WMA. This evaluation treated the model's outputs (before thresholding to a binary decision) as probabilistic predictions of pathology. A calibration curve was generated, plotting predicted probabilities against the observed frequencies of the positive class (abnormal wall motion). The calibration curve closely followed the diagonal line of identity, indicating good agreement between predicted probabilities and actual outcomes (i.e., when the model predicts a 70% probability of abnormality, the true rate is approximately 70%) (Figure 6). In practical terms, the model demonstrates neither over-confidence or under-confidence across the prediction range. We also examined the Brier score, a proper scoring rule for probabilistic forecasts that measures the mean squared error of the probability predictions. Lower Brier scores indicate better-calibrated and more accurate probabilities, with a perfect model scoring 0. In this study, given the high overall accuracy and low error rates, the Brier score for the ResNet50 classifier was low (approximately 0.05 by our estimates), confirming that the model's probability outputs are meaningful. In summary, the model not only distinguishes normal from abnormal wall motion with high AUC and accuracy, but also provides well-calibrated probability scores that appropriately reflect uncertainty. This is important for clinical deployment, as well-calibrated probabilities enable clinicians to incorporate model confidence into decision-making (for example, flagging borderline cases for closer expert review). The strong calibration result suggests that the SafeHeart WMA module could be effectively used in a probabilistic manner, such as triaging studies by abnormality severity in addition to making binary predictions.

Finally, although ResNet50 achieved the best performance, no statistical significance tests were performed to compare model performances in this study. Differences in accuracy and other metrics among models (e.g., ResNet50 vs. VGG16, etc.) were observed but not formally tested. In a rigorous comparative analysis, methods such as McNemar's test could be used to assess significant differences in error rates between paired classifiers, or the DeLong test could be applied to compare the ROC AUCs (Figure 7) Since tests were not conducted, any statements about one model performing "better" than another are based solely on numerical trends. We therefore refrain from making claims of statistical superiority for ResNet50 and instead report that it achieved the highest numerical performance, focusing on its results for further analysis. This approach avoids unwarranted assumptions regarding significance and acknowledges that formal paired comparisons were beyond the study's scope.

Finally, the average processing time of SafeHeart was 12.4 ± 2.1 seconds per study on a standard GPU workstation (NVIDIA RTX 3060), compared with several minutes typically required for manual EF measurement and wall motion scoring by cardiologists. On a CPU-only laptop, the analysis time was under 40 seconds, still considerably faster than manual assessment.

DISCUSSION

In this study, we developed SafeHeart, a two-module AI system for echocardiographic analysis, and demonstrated its ability to automatically calculate LVEF calculation and detect WMAs with high accuracy. The algorithm automatically identifies end-systolic and end-diastolic areas in both A2C and A4C views without requiring manual input. The results showed high accuracy in LVEF estimation and strong concordance with expert reference measurements (Pearson $r = 0.71$, CCC 0.85). We support the view that enhancing anatomical and physiological assessment of the left ventricle has the potential to improve its application in diagnosing and screening heart diseases in routine clinical practice.²⁰⁻²²

Consistent with previous studies, we utilized the A2C and A4C axes of 2D-TTE to evaluate systolic dysfunction, calculate left ventricular volume, and monitor myocardial motion.²³ 2D-TTE is a widely used, standard, and non-invasive imaging technique that enables LVEF assessment from multiple viewing angles, including A4C, A2C, parasternal long axis, and parasternal short axis.²⁴ However, interpreting images from all these views is time-consuming and demands substantial expertise.²⁵ Although semi-automatic systems

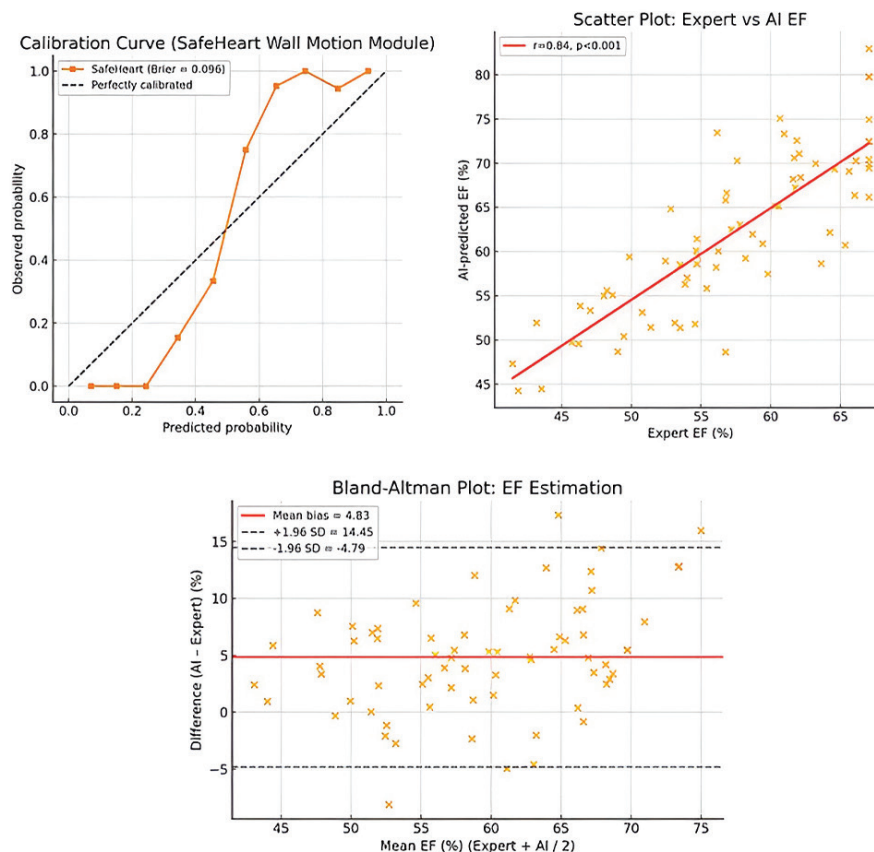


FIG. 6. (a) Scatter plot: expert vs. AI EF: there was a significant positive correlation between AI-predicted and expert-assessed EF values ($r = 0.64$, $p < 0.001$). The regression line indicates that the algorithm generally follows clinical EF trends, though at higher EF ranges AI tends to slightly overestimate values. (b) Q-Q plots (expert EF and AI EF): Q-Q plots showed that residuals of AI-derived EF values were closer to the normal distribution compared to expert EF values, suggesting more symmetrical prediction errors. Expert EF residuals exhibited mild deviation at lower ranges. (c) Bland-Altman plot: Bland-Altman analysis revealed a mean bias of -5.11, indicating that AI estimates were on average ~ 5% higher than expert measurements. The 95% limits of agreement (-21.5 to +11.2) indicate that individual-level deviations can be clinically relevant, especially in borderline EF categories.

EF, ejection fraction; AI, artificial intelligence.

have been introduced to analyze one or two axes, extracting the relevant phases from videos and delineating the ventricular region still extends processing time. These challenges underscore the necessity of fully automated AI algorithms. To accelerate analysis, DL models designed to interpret single axis images would require much larger and more homogeneous datasets.^{26,27}

Our work contributes to and extends the growing body of research on AI applications in echocardiography. For instance, Ouyang et al.²⁸ developed a video-based DL algorithm (EchoNet-Dynamic) to estimate EF and detect reduced EF (HFrEF). Their model achieved a mean absolute error of approximately 4.1% for EF and an AUC of 0.97 for detecting heart failure with reduced EF ($EF \leq 40\%$) focused on WMAs. Similarly, they trained a deep neural network to identify regional WMA across seven coronary territories, achieving an AUC of 0.96 and performance comparable to expert clinicians. These benchmark studies highlight that AI can achieve expert-level accuracy in specific echo tasks. Our SafeHeart system compares

favorably with these benchmarks, achieving an EF error of 4% (bias -4%, limits of agreement -15% to -3%) and an AUC of 0.96 for global wall motion classification, results consistent with those reported by Ouyang et al.²⁸ and Slivnick et al.²⁹

By integrating DL with the classical Simpson method, our study reduced manual processing steps through automated phase selection, therefore decreasing analysis time. We further demonstrated that this hybrid approach enables assessment of LVEF and wall motion even with smaller datasets.

Despite significant advances, automating EF calculation remains challenging, as earlier studies have reported specific limitations.^{30,31} Many approaches to automatic boundary detection still required manual correction. Although DL provides robust image analysis in dynamic videos, it introduces difficulties in identifying left ventricular cavities. Moreover, while consecutive video frames can improve area detection, reduced image quality often hinders

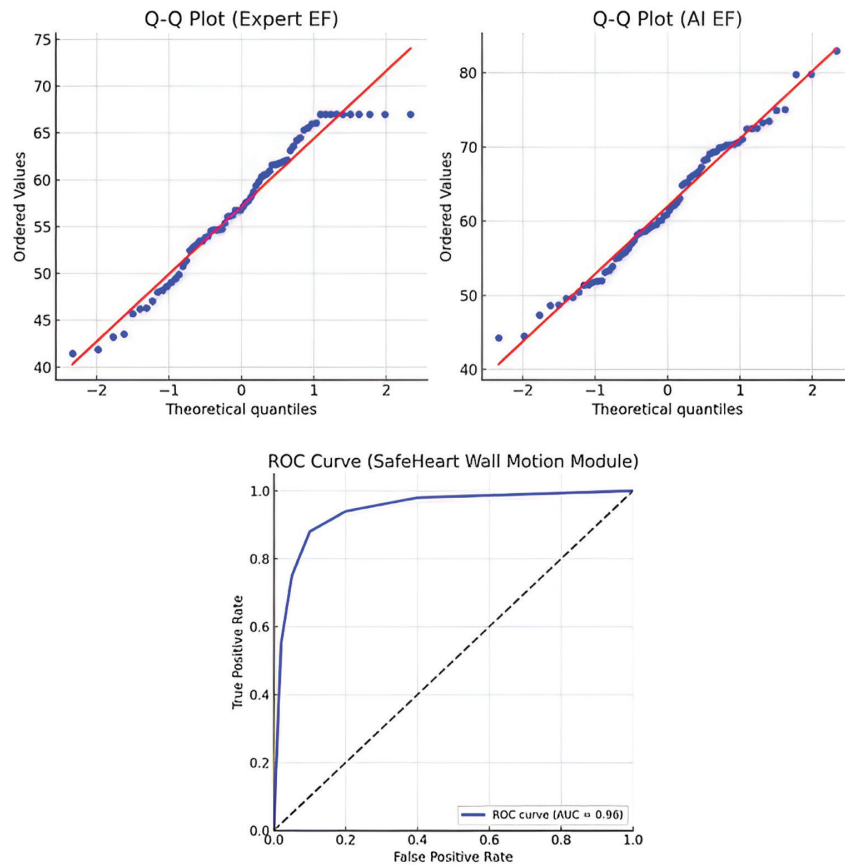


FIG. 7. Q-Q plots (expert EF and AI EF) the Q-Q plot for expert EF values (left) shows mild deviation from the normal distribution at the extremes, particularly at lower values, suggesting some skewness in expert-derived measurements. By contrast, the AI EF residuals (right) follow the diagonal line more closely, indicating a distribution that approximates normality. This suggests that prediction errors from the AI model are more symmetrically distributed. ROC curve (SafeHeart Wall Motion Module): the ROC curve for the SafeHeart wall motion classification module demonstrated excellent discriminative ability, with an AUC of 0.96. This indicates that the model is able to accurately distinguish between normal and abnormal wall motion in the majority of cases. The steep rise of the curve near the y-axis reflects high sensitivity at low false positive rates, which is clinically valuable in screening for pathology.

EF, ejection fraction; AI, artificial intelligence; ROC, receiver operating characteristic; AUC, area under the curve.

differentiation between papillary muscles and the endocardial layer in current algorithms.³²⁻³⁴

A novel aspect of SafeHeart is its hybrid method, which combines static image analysis with dynamic tracking. Initially, DL segmentation is applied to high-quality static frames (A4C and A2C at end-diastole and end-systole) to delineate the left ventricular cavity. Subsequently, volume calculation throughout the cardiac cycle is performed by tracking these segmented regions through the video (integrating the Riemann sum and Simpson's method). This approach overcomes common challenges: it preserves image quality for border detection by performing segmentation on clear static images while still capturing the dynamic variations in cavity area required for EF estimation. As a result, we addressed typical issues observed in earlier studies, such as misclassification of papillary muscles or the need for manual contour adjustments in suboptimal frames. The Bland-Altman analysis confirmed no significant heteroscedasticity

indicating that model error remained stable across the EF spectrum, a strong indicator of consistent accuracy.

Beyond LVEF estimation, evaluating myocardial hypokinesia is essential for the echocardiographic assessment of ischemic heart disease.³⁴ However, assessing regional WMAs requires expertise, diagnostic accuracy often varies depending on the operator's skill level. DL algorithms offer an objective, reproducible alternative with accuracy comparable to expert visual evaluations, reducing operator-dependent variability.^{29,35}

According to the ASE 17-segment model, the left ventricle is divided into septal, lateral, anterior, and inferior regions across apical two- and four-chamber views. In this study, we simplified the segmentation into seven reproducible regions to facilitate automation. From the A4C view, basal and mid septal, basal and mid lateral, and one combined apical region. From the A2C view,

basal and mid anterior segments were included, together with the same merged apical region. This framework integrates apical septal, lateral, anterior, and inferior walls into a single apical segment, minimizing redundancy between views while maintaining representation of clinically relevant vascular territories (LAD, LCx, RCA). Such segmentation provides a standardized foundation for detecting regional WMAs and reduces operator-dependent variability. In related studies, regions were classified into anterior, septal, lateral, and inferior.^{29,36} Our approach offers more precise analysis by clearly defining the number of segments.

The ability of our algorithm to rapidly and reliably calculate LVEF significantly reduces time lost to manual measurement, enabling faster diagnostic and therapeutic decision-making. This advantage is particularly valuable in intensive care units and emergency departments. Achieving 95% accuracy in WMA detection enhances early diagnosis of ischemic heart disease and facilitates more effective post-myocardial infarction management. The algorithm's color-coding system and comprehensive evaluation approach minimize inter-operator variability, promoting consistent and reproducible reporting among clinicians with different experience levels. Moreover, fast and standardized outputs allows for more accurate assessments of treatment efficacy, supporting improved rehabilitation and patient management decisions.

In this study, we introduced an AI algorithm comprising two main modules for 2D-TTE image analysis in the monitoring of IHD. These modules focus on automatic LVEF calculation and WMA detection. Our objective was to develop an AI-driven system capable of tracking wall motion and predicting LVEF using both static and dynamic TTE images. Instead of performing manual phase detection prior to virtual disk placement in the conventional Simpson method, our algorithm tracks the cavity area dynamically through videos using an integrated Riemann sum. This design enables a faster and more efficient workflow than existing methods. The accelerated processing also supports potential real-time applications, such as providing immediate feedback during echocardiographic acquisition, thereby reducing the need for repeat imaging. However, processing speed may vary with hardware performance, and future deployment will require optimization across clinical systems to ensure consistent real-time operation.

The 95% accuracy achieved in wall motion function evaluation further supports the reliability of LVEF estimation. Supported by anatomical region segmentation with color coding, the algorithm provides standardization and a comprehensive framework for operator reporting. Future work should incorporate more diverse data-spanning multiple echo labs, equipment vendors, patients pathologies to enhance model robustness. Although SafeHeart was developed primarily for ischemic heart disease, where WMAs are prevalent, its potential application extends to other cardiac conditions such as cardiomyopathies with regional dysfunction. Expanding training to include such cases will broaden its clinical utility. This study has several limitations. First, the density of papillary muscles can vary among patients during LVEF evaluation in

dynamic videos. When papillary muscles are particularly dense, they may merge with the myocardium, leading to errors in ventricular volume estimation or even complete image detection failure. Such issues may result in inaccurate EF values or unusually high volume change rates. To prevent these errors, we incorporated a protective layer that disables EF calculation when the ventricular cavity cannot be detected. Second, our model performs wall motion analysis only as either normal or pathological. A more detailed classification of pathological motion as hypokinetic, dyskinetic, akinetic, or dyssynchronous would yield more informative results for diagnosis and prognosis. Third, the use of an open-access dataset to supplement the A4C parameter for the wall motion module limits the overall scope of our research. We acknowledge constraints in the generalizability and scope of our research. SafeHeart was trained and tested on a dataset of limited size (600 patients in total, with 153 patients contributing to the wall motion module after data splitting) and with certain homogeneity (all images were A4C and A2C views, primarily from a single vendor machine, supplemented with a few external A4C videos from Stanford). A larger dataset is required to enable the differentiation of pathological subtypes. The dependence on an open-access dataset and the homogeneity of the proprietary dataset restricted the model's ability to generalize across diverse patient populations. One of the factors affecting the use of the supporting dataset in analyzing pathological subtypes was the variation in image resolution quality. Another contributing factor was the quantitative limitations of our proprietary dataset. Although wall motion was classified as normal or pathological, with anatomical location identification supported by color codes, segment-specific motion analysis was not performed. To integrate wall motion scores at the segmental level, homogeneous datasets containing images that represent each type of motion for each segment are required.

Acknowledgements: The open-access dataset used as additional support in this study can be accessed at: Stanford AIMI. (2024, March 16). EchoNet-Dynamic. Stanford Aimi shared datasets. <https://stanfordaimi.azurewebsites.net/datasets/834e1cd1-92f7-4268-9daa-d359198b310a>.

Ethics Committee Approval: The study protocol was approved by the Recep Tayyip Erdoğan University Non-Interventional Clinical Research Ethics Committee (approval number: 2024/143, date: 13.06.2024).

Informed Consent: Not necessary because retrospective study.

Data Sharing Statement: The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Authorship Contributions: Concept- S.G.; Design- S.G., H.H.; Supervision- H.D.; Materials- R.T., N.E., M.K.; Data Collection or Processing- R.T., B.A., H.H., S.K., N.E.; Analysis and/or Interpretation- R.T., M.K.; Literature Review- H.D.; Writing- S.G., B.A., S.K., N.E.; Critical Review- H.D., M.K.

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: This project has been supported by Recep Tayyip Erdoğan University through the Scientific Research Project (BAP) with the project code TLO-2024-1711. The project has also received funding support from the Turkish Technology Team Foundation (T3 Foundation), and it achieved 4th place in Türkiye at the 2024 Teknofest Technology Festival organized under the foundation.

REFERENCES

- GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Erratum in: *Lancet*. 2020;396:1562. [\[CrossRef\]](#)
- Dimond MG, Ibrahim NE, Fiuzaat M, et al. Left ventricular ejection fraction and the future of heart failure phenotyping. *JACC Heart Fail*. 2024;12:451-460. [\[CrossRef\]](#)
- Roffi M, Patrono C, Collet JP, et al.; ESC Scientific Document Group. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2016;37:267-315. [\[CrossRef\]](#)
- Kagiyama N, Abe Y, Kusunose K, et al. Multicenter validation study for automated left ventricular ejection fraction assessment using a handheld ultrasound with artificial intelligence. *Sci Rep*. 2024;14:15359. [\[CrossRef\]](#)
- Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71:2668-2679. [\[CrossRef\]](#)
- Li X, Zhang H, Yue J, et al. A multi-task deep learning approach for real-time view classification and quality assessment of echocardiographic images. *Sci Rep*. 2024;14:20484. [\[CrossRef\]](#)
- Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D. Automatic apical view classification of echocardiograms using a discriminative learning dictionary. *Med Image Anal*. 2017;36:15-21. [\[CrossRef\]](#)
- Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9:611-629. [\[CrossRef\]](#)
- Li H, Wang Y, Qu M, Cao P, Feng C, Yang J. EchoEFNet: multi-task deep learning network for automatic calculation of left ventricular ejection fraction in 2D echocardiography. *Comput Biol Med*. 2023;156:106705. [\[CrossRef\]](#)
- Wang S, Chauhan D, Patel H, et al. Assessment of right ventricular size and function from cardiovascular magnetic resonance images using artificial intelligence. *J Cardiovasc Magn Reson*. 2022;24:27. [\[CrossRef\]](#)
- Sun S, Wang Y, Yu Q, Qu M, Li H, Yang J. STGA-MS: AI diagnosis model of regional wall motion abnormality based on 2D transthoracic echocardiography. *Heliyon*. 2023;10:e23224. [\[CrossRef\]](#)
- Olaisen S, Smistad E, Espeland T, et al. Automatic measurements of left ventricular volumes and ejection fraction by artificial intelligence: clinical validation in real time and large databases. *Eur Heart J Cardiovasc Imaging*. 2024;25:383-395. [\[CrossRef\]](#)
- Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; June 27–30, 2016; Las Vegas, NV. IEEE; 2016:779-788. [\[CrossRef\]](#)
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. *arXiv*. [\[CrossRef\]](#)
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; June 27–30, 2016; Las Vegas, NV. IEEE; 2016:770-778. [\[CrossRef\]](#)
- Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345-1359. [\[CrossRef\]](#)
- Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. The MIT Press; 2016. 800 pp. ISBN: 0262035618. *Genet Program Evol Mach*. 2018;19:305-307. [\[CrossRef\]](#)
- Lang RM, Badano LP, Mor-Avi V, et al. 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*. 2015;28:1-39.e14. [\[CrossRef\]](#)
- Shoab MA, Lai KW, Chuah JH, et al. Comparative studies of deep learning segmentation models for left ventricle segmentation. *Front Public Health*. 2022;10:981019. [\[CrossRef\]](#)
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36. [\[CrossRef\]](#)
- El-Taraboulsi J, Cabrera CP, Roney C, Aung N. Deep neural network architectures for cardiac image segmentation. *Artif Intell Life Sci*. 2023;4:100083. [\[CrossRef\]](#)
- Tokodi M, Magyar B, Soós A, et al. Deep learning-based prediction of right ventricular ejection fraction using 2D echocardiograms. *JACC Cardiovasc Imaging*. 2023;16:1005-1018. [\[CrossRef\]](#)
- Li SW, Dong TX, Bai RC, et al.; study investigators. Comparison of model-based and speckle tracking-based automated three-dimensional echocardiographic quantification of left ventricular volumes and function: a multicenter study. *Int J Cardiol*. 2025;437:133426. [\[CrossRef\]](#)
- Lang RM, Bierig M, Devereux RB, et al.; Chamber Quantification Writing Group; American Society of Echocardiography's Guidelines and Standards Committee; European Association of Echocardiography. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. *J Am Soc Echocardiogr*. 2005;18:1440-63. [\[CrossRef\]](#)
- Knackstedt C, Bekkers SC, Schummers G, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. *J Am Coll Cardiol*. 2015;66:1456-1466. [\[CrossRef\]](#)
- Zhao Z, Chen Y, Wu J, et al. MMGL: Multi-scale multi-view global-local contrastive learning for semi-supervised cardiac image segmentation. 2022. *arXiv*. [\[CrossRef\]](#)
- Kusunose K, Haga A, Inoue M, Fukuda D, Yamada H, Sata M. Clinically feasible and accurate view classification of echocardiographic images using deep learning. *Biomolecules*. 2020;10:665. [\[CrossRef\]](#)
- Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580:252-256. [\[CrossRef\]](#)
- Slivnick JA, Gessert NT, Cotella JL, et al. Echocardiographic detection of regional wall motion abnormalities using artificial intelligence compared to human readers. *J Am Soc Echocardiogr*. 2024;37:655-663. [\[CrossRef\]](#)
- Cannesson M, Tanabe M, Suffoletto MS, et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. *J Am Coll Cardiol*. 2007;49:217-226. [\[CrossRef\]](#)
- Seng CH, Demirli R, Amin MG, Seachrist JL, Bouzerdoum A. Automatic left ventricle detection in echocardiographic images for deformable contour initialization. *Annu Int Conf IEEE Eng Med Biol Soc*. 2011;2011:7215-8. [\[CrossRef\]](#)
- Rahmouni HW, Ky B, Plappert T, et al. Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection. *Am Heart J*. 2008;155:562-570. [\[CrossRef\]](#)
- Shibayama K, Watanabe H, Iguchi N, et al. Evaluation of automated measurement of left ventricular volume by novel real-time 3-dimensional echocardiographic system: validation with cardiac magnetic resonance imaging and 2-dimensional echocardiography. *J Cardiol*. 2013;61:281-288. [\[CrossRef\]](#)
- Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138:1623-1635. [\[CrossRef\]](#)
- Huang MS, Wang CS, Chiang JH, Liu PY, Tsai WC. Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography. *Circulation*. 2020;142:1510-1520. [\[CrossRef\]](#)
- Kusunose K, Abe T, Haga A, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging*. 2020;13:374-381. [\[CrossRef\]](#)