Heuristic-Boosted Ejection Fraction Estimation from 2D U-Net Segmentation

Ernesto David Serize Portela[1][0009−0009−4527−9977], Amalia Rodríguez Sánchez[1][0009−0007−7979−7006], José Carlos Serize Portela[2][0009−0009−4527−9977], MsC. Alejandro Cespón Ferriol[1][0000−0002−8584−6958], and Dr. Jose Ignacio Ramírez Gómez[3][0000−0003−3630−5722]

[1] Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba
http://www.uclv.edu.cu
[2] Constructor University, Bremen, Germany
https://www.constructor.university
[3] Hospital Cardiocentro "Ernesto Che Guevara", Santa Clara, Cuba

Abstract. Ejection fraction estimation from apical 4-chamber echocardiogram videos is achieved through a pipeline combining deep learning segmentation and detailed analysis of ventricular geometry. A ResNet50-encoded 2D U-Net performs frame-by-frame left ventricle segmentation, with ventricular volumes subsequently calculated via the area-length method. To correct systematic biases arising from segmentation errors and heuristic volume estimation, the pipeline incorporates a regression model that predicts the signed error between estimated and ground truth ejection fractions using geometric descriptors. The most informative predictors include length ratio and volume ratio. This approach achieves a mean absolute error of 4.75% on the EchoNet-Pediatric dataset for apical four-chamber views, offering an interpretable and refined estimation of cardiac function.

Keywords: Ejection Fraction · Left-Ventricle Segmentation · 2D U-Net · ResNet50 · Heuristic · Regression

1 Introduction

Accurate estimation of ejection fraction (EF) plays a key role in assessing cardiac function. Traditionally, EF assessment relies on manual interpretation of echocardiogram videos, a process that is both time-consuming and prone to variability due to its subjective nature. This has driven a growing interest in automated methods that can provide fast, consistent, and objective measurements.

In recent years, this field has seen significant advances fueled by deep learning. Segmentation of the left ventricle has been addressed in several studies, many of which focus on 2D segmentation. Numerous efforts have improved U-Net architectures by incorporating attention mechanisms and residual connections [1][2], while GAN-based data augmentation has helped address limited annotations [3]. Lightweight U-Net and ResUnet variants have been effective for MRI segmentation [4][5], and adversarial learning combined with multi-stage pose estimation

networks has enhanced the anatomical consistency of segmentation outputs[6] .Recently, hybrid models that combine convolutional neural networks with Transformer architectures have demonstrated promising results by capturing both fine-grained local details and broader global context in echocardiographic images [7]. The publication of EchoNet-Dynamic[8] introduced a video-based deep learning algorithm that achieves state-of-the-art performance and released a large public dataset of 10,030 annotated echocardiogram videos.

Although 3D CNN have demonstrated high performance [9], their computational demands and limited interpretability represent significant challenges for practical implementations, particularly in low-resource environments.

To address these challenges, our approach begins with frame-wise segmentation in apical four-chamber view using a 2D U-Net with ResNet50 encoder backbone. Ventricular volumes and EF are then estimated using the well-established area-length method, which is grounded in cardiac geometry[10].

By recognizing the inherent inaccuracies associated with this method, we implement a second stage: a regression model that determines the difference between the predicted and true ejection fraction values, using two key geometric features—ventricular length ratio and volume ratio—as inputs. Inclusion of this module not only significantly improves accuracy, reducing the mean absolute error from 6.40% to 4.75%, but also provides an alternative to black-box deep learning models by leveraging features that are well aligned with clinical principles.

2 Materials and Methods

2.1 Dataset

This study uses the apical 4-chamber (A4C) subset from EchoNet-Pediatric data[11], developed and published by researchers at Stanford University.

It comprises 3,176 A4C video clips collected from pediatric patients aged 0 to 18 years with a balanced gender distribution. Each 112×112 pixels video consists of at least one cardiac cycle. The videos are annotated with expert manual tracings of the left ventricle at end-systole and end-diastole. The dataset covers a wide range of cardiac function, with EF values spanning from 7% to 72% , reflecting diverse clinical conditions.

Stanford's research demonstrated accurate ejection fraction estimation using A4C and parasternal short-axis views, achieving a mean absolute error of 3.66%, a reference point for evaluating new models.

To train the segmentation model, we built an image-mask dataset by extracting and pairing two frames from each video —end-systole and end-diastole—with their associated binary segmentation masks, following the annotations and tracings provided by Stanford. Then these samples were split into training, validation, and test sets, keeping the same video labels, ensuring methodological integrity.

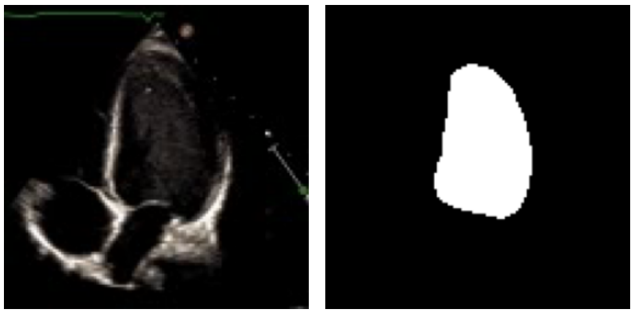

Fig. 1. Frame-Mask sample

2.2 2D U-Net

Among all deep learning approaches for medical image segmentation, encoder-decoder architectures have a notable performance. The CAMUS project demonstrated that such models can accurately segment cardiac chambers and estimate clinical indices and significantly outperformed traditional methods in echocardiographic analysis [12].

Based on these findings, to perform frame-by-frame segmentation, we implemented using PyTorch [13][14] two 2D network architectures a standard U-Net and, to enhance this baseline, a U-Net with a ResNet50 encoder already pretrained on ImageNet[15], with the goal of exploring whether transfer learning could improve segmentation accuracy.

Both models were set up with similar conditions and trained, validated and tested in the same environment, using the previously described frame-mask dataset with the same data augmentation strategies—random rotations and horizontal flips—. The models operate on a normalized grayscale input image, and outputs a probability map.

As is common in segmentation tasks, we used a combination of binary cross-entropy and Dice loss to balance pixel-wise accuracy with region-level overlap. Under this setup, both models showed stable convergence over the course of 20 training epochs. Although, as it was expected, based on the loss curves (see Fig.1 and Fig.2) the ResNet50-encoded U-Net outperformed the standard U-Net on the test set (see Table 1), achieving higher mean and lower variability for Dice similarity and intersection-over-union (IoU) scores.

Table 1. Comparison between Standard U-Net and ResNet50-Encoded U-Net performance on the test set.

Metric	**Standard U-Net**	**ResNet50-Encoded U-Net**
Dice	0.903(±0.027)	0.917(±0.020)
IoU	0.824(±0.044)	0.847(±0.034)

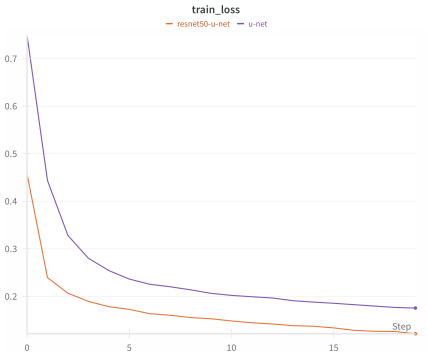

Fig. 2. Train loss over epochs

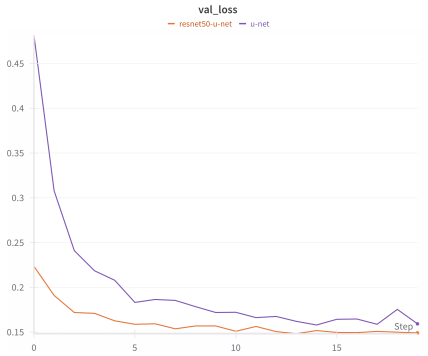

Fig. 3. Validation loss over epochs

Due to computational limitations, no hyperparameter optimization process was performed. Number of epochs, batch size, learning rate, and optimizer were selected based on commonly used defaults[9], prior experiments, and computational power. However, ResNet50 encoded U-Net still achieved a good performance compared to results reported in related studies, and was selected as the final segmentation model.

2.3 Ejection Fraction Estimation via Heuristic Area–Length Method

For an echocardiogram video, each frame at time t is passed as input to the segmentation model and the probability map outputted is thresholded by 0.5, to yield a binary mask $I_t \in \{0, 1\}$, defining the region of interest as,

$$\Omega_t \;=\; \{(x, y) \mid I_t(x, y) \;=\; 1\} \tag{1}$$

From this mask, we compute two key anatomical descriptors: the area $|\Omega_t|$, defined as the total number of pixels within the segmented ventricle, and the length L_t,

$$L_t = \max_{(x_i,y_i),\,(x_j,y_j)\in\Omega_t} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2}$$

which is estimated as the maximum Euclidean distance between any two pixels inside the predicted ventricular region. This length approximation is used because apical four-chamber views are not always vertically aligned, and the apex of the ventricle is not consistently well-defined. Therefore, this heuristic captures the ventricular long axis regardless of orientation.

These per frame descriptors are then used to estimate volumes using the area–length (bullet) method,

$$V_t = \frac{5}{6} \cdot |\Omega_t| \cdot L_t \cdot s^3, \quad s \in \mathbb{R}^+ \tag{3}$$

where s is the pixel spacing in millimeters. After computing the volumes across the video, the maximum and minimum values are identified as end-diastolic and end-systolic volumes, respectively. The heuristic ejection fraction is then computed as,

$$\widehat{\mathbf{EF}} = \left(\frac{\mathbf{V}_{\max} - \mathbf{V}_{\min}}{\mathbf{V}_{\max}} \right) \times 100 \tag{4}$$

So far, the mean absolute error for the ejection fraction estimates on the test set videos was approximately 6.40%, with a corresponding Pearson correlation coefficient of $\rho = 0.729$.

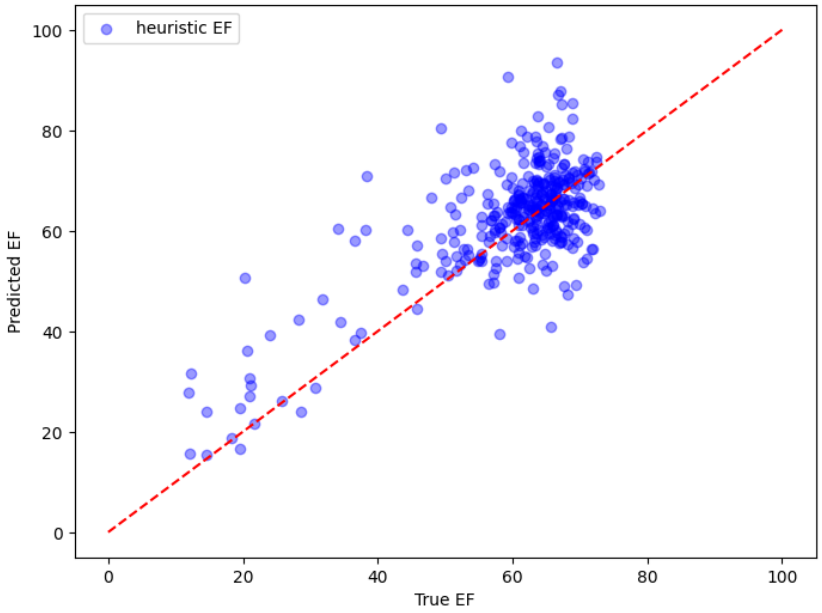

Fig. 4. True vs predicted ejection fraction

2.4 Learned Correction of Signed Error

Although this heuristic method is a clinically established approach for estimating the ejection fraction, its accuracy inherently depends on the fidelity of the segmentation and geometric assumptions can introduce systematic bias, particularly in irregular heart morphologies.

To quantify and mitigate these deviations, some studies include a correction mechanism to identify segmentation outliers, and reduce their influence in ejection fraction estimate[16].

We introduce a learned correction strategy that predicts the signed error. Let the estimated ejection fraction be denoted by $\widehat{EF}$, and the ground truth by EF_{true}. The error we aim to predict is defined as,

$$\varepsilon \ = EF_{true} \ - \ \widehat{EF} \tag{5}$$

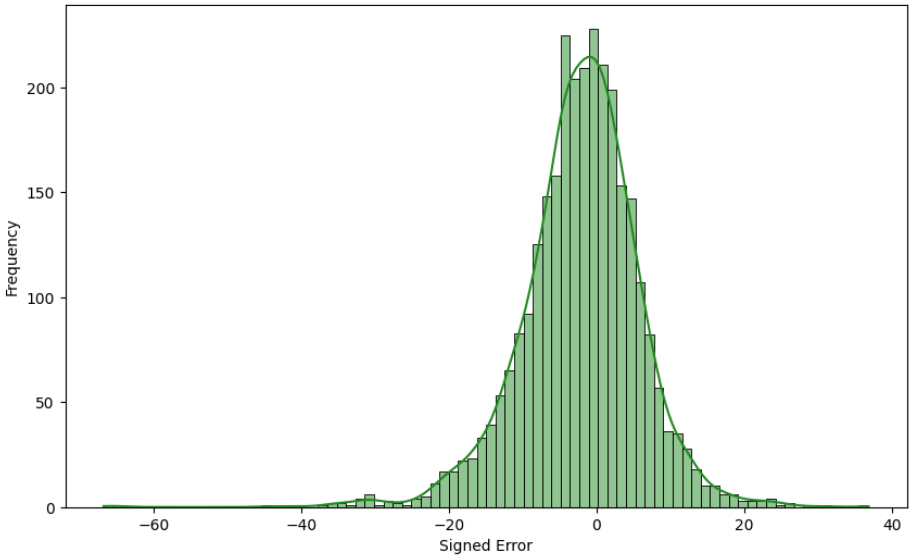

Fig. 5. Signed error distribution on test set

To estimate ε, we trained a machine learning regression model using features extracted from the frame-by-frame segmentation process. These features encapsulate volumetric consistency, morphological stability. The model outputs a predicted correction $\hat{\varepsilon}$, which we use to obtain a new ejection fraction estimate.

$$\widehat{EF}_{corrected} \ = \ \widehat{EF} \ + \ \hat{\varepsilon} \tag{6}$$

2.5 Geometric Features from Frame-by-Frame Segmentation

During the frame-by-frame segmentation process, various geometric descriptors were tracked over time to capture the dynamic behavior of the left ventricle. Based on their correlation with the signed error between the bullet method's ejection fraction and the true clinical values, two features stood as the most predictive: the length ratio and the volume ratio.

The length ratio is defined as the ratio between the minimum and maximum ventricular lengths throughout the cardiac cycle. This feature captures the extent of longitudinal contraction, which is strongly associated with systolic function.

$$Length\ Ratio\ =\ \frac{L_{min}}{L_{max}} \tag{7}$$

Similarly, the volume ratio approximates stroke efficiency and provides an estimate of the heart's pumping capacity over time.

$$Volume\ Ratio\ =\frac{V_{min}}{V_{max}} \tag{8}$$

To evaluate the relevance of these two features in predicting the signed error, we computed both Pearson's and Spearman's correlation coefficients (see Table 2)

Table 2. Correlation coefficients (Pearson and Spearman) between derived features and the signed error.

Feature	**Pearson**	**Spearman**
Length ratio	0.375	0.425
Volume ratio	0.293	0.457

Furthermore, a strong linear correlation was observed between volume ratio and length ratio, with Pearson's $\tau_p = 0.89$ and Spearman's $\tau_s = 0.90$. This expected relationship reflects the geometric coupling between ventricular area and length across the cardiac cycle. Despite this, combining both features into a joint descriptor did not improve model performance.

Although these features illustrate that the heuristic method tends to slightly underestimates in cases where the ventricle contracts as expected and overestimates on reduced motion or abnormal contraction.

2.6 Machine Learning Model Selection

We ran experiments using the scikit-learn[17] implementations of three regressor models—Random Forest, K-Nearest Neighbors, and Gradient Boosting—using the inference of the segmentation and bullet method with the predefined split annotations.

Gradient Boosting achieved the highest R^2 and the lowest mean absolute error and mean squared error on the validation set (see Table 3), demonstrating a superior capacity to capture relationships present in the feature space. This model was therefore selected for hyperparameter tuning.

After hyperparameter tuning, the model was evaluated on the held-out test set. This final performance on unseen data yielded a mean squared error of 42.4, R^2 of 0.408 and a mean absolute error of 4.75, which is the final ejection fraction error of the hole pipeline.

To visually assess the impact of the signed error predictor, we present a comparison between the original heuristic predictions and the corrected values obtained by applying the predicted error to the initial estimate (see Fig. 7).

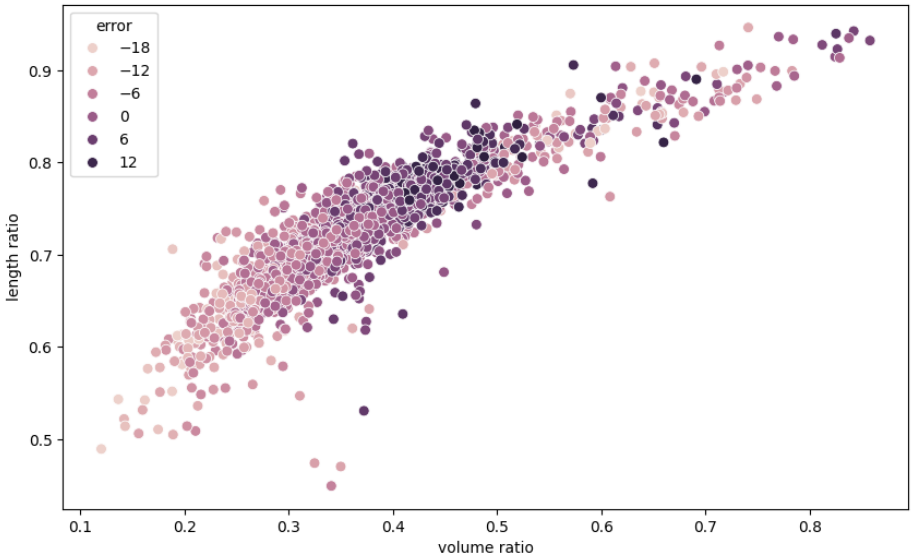

Fig. 6. Volume ratio versus length ratio with signed error

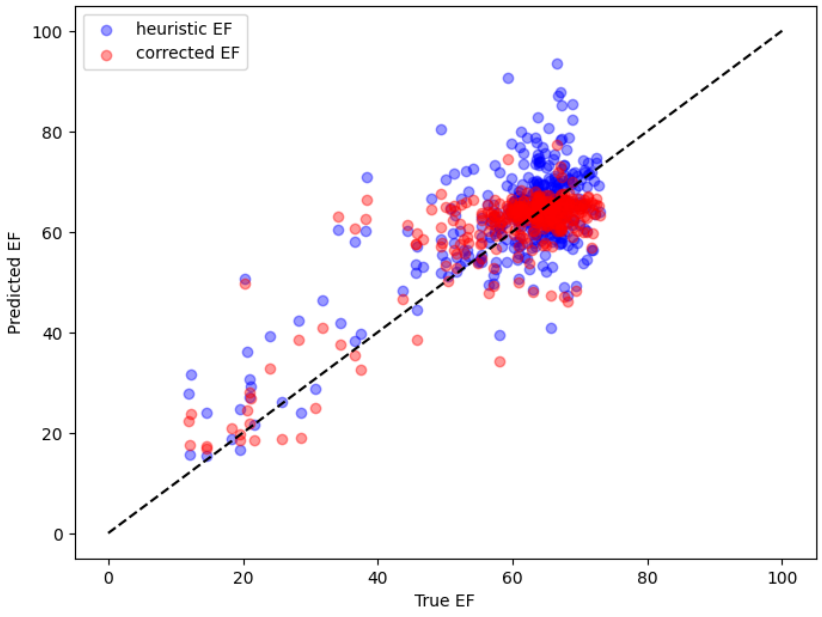

Fig. 7. Signed error predictor impact

Table 3. Performance comparison of candidate models on validation set

Model	**MAE**	**MSE**	R^2
Gradient Boosting	3.959980	26.459827	0.370116
K-Nearest Neighbors	4.032440	26.894415	0.359771
Random Forest	4.264104	29.986607	0.286160

3 Results

The ResNet50 encoded 2D U-Net accurately outlined the left ventricle, achieving a mean Dice score of 0.92 and an IoU of 0.85 in the test data. After frame-by-frame segmentation, the initial ejection fraction estimates had a mean absolute error of 6.40%, reflecting the usual challenges of heuristic volume calculations when dealing with different heart shapes. By applying a regression model focused on length and volume ratio, the heuristic method was enhanced, bringing the corrected mean absolute error down to 4.75%.

Both already trained models, Segmentation and Signed Error Predictor, were applied with the same idea on the Echonet-Dynamic test set to measure the pipeline performance on adult data. The initial mean absolute error estimation ejection fraction was 8.94% and after summing the predicted error this result was reduced to 7.34%. However, this is not a state-of-the-art score, differences in heart behavior between children and adults were expected to lead to a higher error. Still, we can see an improvement coming from the signed-error predictor, indicating that our approach remains valid.

4 Conclusions

By combining a reliable heuristic approach with data-based corrections, this work offers a way to improve accuracy while keeping the process understandable for clinicians. Building on the well-known area-length method ensures that the pipeline stays connected to physiological principles, addressing the common issue of limited transparency in many deep learning models.

Adding geometric descriptors helps to capture the heart's complex movements, reducing errors that often affect traditional methods. These results highlight the promise of hybrid approaches that strike a balance between clinical interpretability and improved precision. Such solutions not only fit naturally into existing clinical workflows but also have the potential to make automated cardiac function assessment more accessible and widely adopted across different healthcare settings.

References

1. Azarmehr, N., Ye, X., Janan, F., Howard, J.P., Francis, D.P., Zolgharni, M.: Automated Segmentation of Left Ventricle in 2D echocardiography using deep learning. arXiv preprint arXiv:2003.07628 (2020). https://doi.org/10.48550/arXiv.2003.07628

2. Moradi, S., Ghelich Oghli, M., Alizadehasl, A., Shiri, I., Oveisi, N., Oveisi, M., Maleki, M., Dhooge, J.: MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Physica Medica **67**, 58–69 (2019). https://doi.org/10.1016/j.ejmp.2019.10.001
3. Kumar, V., Sharma, N.M., Mahapatra, P.K., Dogra, N., Maurya, L., Ahmad, F., Dahiya, N., Panda, P.: Enhancing Left Ventricular Segmentation in Echocardiograms Through GAN-Based Synthetic Data Augmentation and MultiResUNet Architecture. Diagnostics **15**(6), 663 (2025). https://doi.org/10.3390/diagnostics15060663
4. Irshad, M., Yasmin, M., Sharif, M.I., Rashid, M., Sharif, M.I., Kadry, S.: A Novel Light U-Net Model for Left Ventricle Segmentation Using MRI. Mathematics **11**(14), 3245 (2023). https://doi.org/10.3390/math11143245
5. Xu, S., Lu, H., Cheng, S., Pei, C.: Left Ventricle Segmentation in Cardiac MR Images via an Improved ResUnet. International Journal of Biomedical Imaging **2022**, 8669305 (2022). https://doi.org/10.1155/2022/8669305
6. Wu, H., Lu, X., Lei, B., Wen, Z.: Automated left ventricular segmentation from cardiac magnetic resonance images via adversarial learning with multi-stage pose estimation network and co-discriminator. Medical Image Analysis **68**, 101891 (2021). https://doi.org/10.1016/j.media.2020.101891
7. Shi, S., Alimu, P., Mahemut, P.: The Study of Echocardiography of Left Ventricle Segmentation Combining Transformer and Convolutional Neural Networks. International Heart Journal **65**(5), 889–897 (2024). https://doi.org/10.1536/ihj.23-638
8. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., Zou, J.Y.: EchoNet-Dynamic: A Large New Cardiac Motion Video Data Resource for Medical Computer Vision. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Lecture Notes in Computer Science **12264**, 66–75 (2020). https://doi.org/10.1007/978-3-030-59719-1_7
9. Ali, W., Alsabban, W., Shahbaz, M., Al-Laith, A., Almogadwy, B.: EFNet: estimation of left ventricular ejection fraction from cardiac ultrasound videos using deep learning. *PeerJ Computer Science* **11**, e2506 (2025). https://doi.org/10.7717/peerj-cs.2506
10. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., Lancellotti, P., Muraru, D., Picard, M.H., Rietzschel, E.R., Rudski, L., Spencer, K.T., Tsang, W., Voigt, J.-U.: Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *European Heart Journal - Cardiovascular Imaging* **16**(3), 233–271 (2015). https://doi.org/10.1093/ehjci/jev014
11. Reddy, C.D., Lopez, L., Ouyang, D., Zou, J.Y., He, B.: EchoNet-Pediatric: A Large Pediatric Echocardiography Video Data Resource for Medical Machine Learning. Stanford University (2023). https://doi.org/10.71718/d05h-gy43
12. Leclerc, B., Smistad, O., Caballero, J., Frangié, P.A., Bernard, O., et al.: Automatic left ventricular myocardium delineation in echocardiographic images with deep learning: the CAMUS project. IEEE Transactions on Medical Imaging **38**(9), 2198–2210 (2019). https://doi.org/10.1109/TMI.2019.2900516
13. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z.,

Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32 (NeurIPS 2019), 8024–8035 (2019). https://pytorch.org

14. Qubvel: Segmentation Models PyTorch. GitHub repository and documentation, 2019–2024. https://github.com/qubvel/segmentation_models.pytorch https://segmentation-models-pytorch.readthedocs.io
15. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
16. Zhang, Y., Liu, B., Bunting, K.V., Brind, D., Thorley, A., Karwath, A., Lu, W., Zhou, D., Wang, X., Mobley, A.R., Tica, O., Gkoutos, G.V., Kotecha, D., Duan, J.: Development of automated neural network prediction for echocardiographic left ventricular ejection fraction. *Frontiers in Medicine* **11**, 1354070 (2024). https://doi.org/10.3389/fmed.2024.1354070
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011). https://scikit-learn.org