Echo-E^{3}Net: Efficient Endocardial Spatio-Temporal Network for Ejection Fraction Estimation

Moein Heidari[1*], Afshin Bozorgpour[2*], AmirHossein Zarif-Fakharnia[3], Wenjin Chen[4], Dorit Merhof[2], D. J. Foran[4], Jasmine Grewal[1], and Ilker Hacihaliloglu[1]

[1] University of British Columbia, Vancouver, BC, Canada
moein.heidari@ubc.ca
[2] University of Regensburg, Regensburg, Germany
[3] Iran University of Science and Technology, Tehran, Iran
[4] Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA
[*]Equal contribution.

Abstract. Left ventricular ejection fraction (LVEF) is a primary marker of cardiac function. However, routine estimation from endocardial measurements requires manual delineation at end-diastole (ED) and end-systole (ES), a process that is time-consuming and subject to inter-observer variability. Reliable automation is especially valuable for point-of-care ultrasound (POCUS), where computational resources are limited and the acquisition quality varies. We propose Echo-E^{3}Net, an anatomy-guided spatio-temporal network that explicitly embeds cardiac anatomy into LVEF prediction. A dual-phase Endocardial Border Detector (E^{2}CBD) uses phase-specific cross-attention to localize ED/ES endocardial landmarks and produce phase-aware landmark embeddings, while an Endocardial Feature Aggregator (E^{2}FA) fuses these embeddings with global statistical descriptors of deep feature maps to refine EF regression. Training is guided by a lightweight geometric loss that uses ED and ES endocardial landmarks to regularize EF prediction. On EchoNet-Dynamic and a PSAX subset of EchoNet-Pediatric, Echo-E^{3}Net attains competitive performance using only 1.55M parameters and 8.05 GFLOPs, an order-of-magnitude compute reduction versus recent baselines, supporting real-time deployment. Our code is publicly available at GitHub.

Keywords: Echocardiography · Ejection fraction · Point-of-care ultrasound · Efficient deep learning

1 Introduction

Left ventricular ejection fraction (LVEF) is a key indicator of cardiac function, routinely used to assess heart failure and guide clinical decision making [8,10,9,17,5]. In standard adult echocardiography, LVEF is commonly estimated from end-diastole (ED) and end-systole (ES) endocardial measurements, often using the method of disks when apical views are available [10]. This process is time-consuming and subject to substantial inter-observer variability, particularly when image quality is reduced, or views are suboptimal [11,6]. These limitations are amplified in point-of-care ultrasound (POCUS), which is increasingly used for bedside cardiac assessment but is highly operator dependent and runs on hardware with constrained compute [2,11]. Reliable LVEF estimation in this setting, therefore, demands methods that are not only accurate but also computationally efficient and robust enough for real-world deployment [3,19].

Deep learning has substantially advanced automated LVEF estimation [9,8,5]. Existing approaches include segmentation-based pipelines [12,1], direct video-level regression models [12,4,9], graph and keypoint-guided strategies that encode cardiac structure [8,18], and attention-based designs for spatio-temporal representation learning [7]. Graph-based methods such as EchoGNN [8] and EchoGraphs [18] improve interpretability through keypoint representations and shape constraints, but can underemphasize the ED and ES frames that are central to clinical EF measurement. EchoMEN [5] addresses EF-range imbalance with a multi-expert design at the cost of added computation, while CoReEcho [6] and reconstruction-based priors [20] further improve accuracy. Across these methods, however, strong accuracy is consistently coupled with heavy parameter counts and FLOPs, which limits adoption in resource-constrained POCUS settings [19].

We address this gap with Echo-E^{3}Net, an efficient endocardial spatio-temporal network that explicitly embeds cardiac anatomy into LVEF prediction. Specifically, a dual-phase Endocardial Border Detector (E^{2}CBD) lets ED and ES queries attend to compressed multi-scale spatio-temporal tokens, yielding explicit landmark coordinates and phase-aware landmark embeddings that compactly encode endocardial geometry and motion. An Endocardial Feature Aggregator (E^{2}FA) fuses these embeddings with global statistical descriptors of deep feature maps to refine EF regression. Training is guided by a lightweight geometric objective that regularizes ED and ES landmark geometry. Our contributions are threefold: ❶ an anatomy-guided architecture in which dual-phase cross-attention explicitly localizes ED/ES endocardial landmarks, keeping EF prediction grounded in cardiac structure; ❷ a single-view geometric regularizer that supervises ED and ES landmark geometry during training; ❸ competitive accuracy on EchoNet-Dynamic and on a PSAX subset of EchoNet-Pediatric, at an order-of-magnitude lower compute (1.55M parameters, 8.05 GFLOPs), with real-time CPU inference and no pre-training, augmentation, or ensembling.

2 Method

Given an echocardiographic video $X \in \mathbb{R}^{B\times C\times F\times H\times W}$ with reference LVEF $y \in (0, 100)$ and dual-phase (ED/ES) endocardial landmark annotations, our goal is to predict EF while remaining consistent with the underlying cardiac geometry. As shown in Figure 1, Echo-E^{3}Net comprises three components: (i) a backbone encoder that extracts multi-scale 3D features; (ii) a dual-phase Endo-

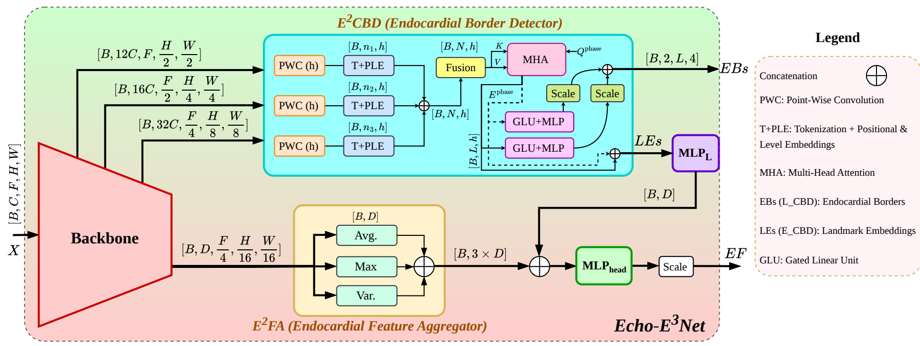

Fig. 1: Architecture of Echo-E^{3}Net. An LHUNet encoder extracts multi-scale spatio-temporal features from the input video X. E^{2}CBD tokenizes the skip features and applies phase-specific cross-attention from ED/ES landmark queries to predict dual-phase endocardial border chords (EBs) and landmark embeddings (LEs). E^{2}FA fuses global feature statistics (average, maximum, variance) with the landmark descriptor to regress EF. Here, EBs denote the predicted endocardial border chords and LEs the landmark embeddings.

cardial Border Detector (E^{2}CBD) that predicts ED/ES landmark chords and phase-aware landmark embeddings; and (iii) an Endocardial Feature Aggregator (E^{2}FA) that fuses these geometric descriptors with global feature statistics to regress EF. E^{2}CBD is trained with anatomically motivated constraints inspired by disk-based ventricular measurement, so that both local shape and global contractility cues drive the final prediction.

2.1 Backbone Network

We adopt the encoder of LHUNet [15] as a lightweight volumetric feature extractor. Unlike its original segmentation use, we keep only the encoder: initial convolutions adapt the input channels, subsequent blocks extract local features while downsampling, and hybrid blocks with large-kernel convolutions and spatial/channel attention capture longer-range context. From X, the encoder produces three multi-scale skip features $\{s_1, s_2, s_3\}$ of decreasing spatial and temporal resolution, together with a deepest feature map x. The skip features feed E^{2}CBD; the deepest feature feeds E^{2}FA.

2.2 Dual-Phase Endocardial Border Detector (E^{2}CBD)

Clinical EF estimation relies on endocardial boundaries delineated at ED and ES. E^{2}CBD mirrors this by predicting, for each phase, a set of L landmark chords ordered along the ventricular boundary, where each chord is a pair of opposing endocardial points.

Multi-scale tokenization. Each skip feature s_i is projected to a shared hidden width h by a point-wise convolution (PWC), $\tilde{s}_i = \mathrm{PWC}(s_i) \in \mathbb{R}^{B\times h\times F_i\times H_i\times W_i}$, then flattened over spatial/temporal axes into $n_i = F_i H_i W_i$ tokens. We add a positional and a level embedding (T+PLE in Figure 1): P_{pos} is a learned linear map of a normalized spatio-temporal coordinate grid $(t, y, x) \in [-1, 1]^3$, so that temporal and spatial positions are encoded jointly, and the learned vector $p_{\text{lvl},i} \in \mathbb{R}^h$ identifies the feature scale:

$$f_i = \text{flatten}(\tilde{s}_i) + P_{\text{pos}}(t, y, x) + p_{\text{lvl},i} \in \mathbb{R}^{B \times n_i \times h}. \tag{1}$$

The three levels are concatenated into a multi-scale token set of $N = \sum_i n_i$ tokens and refined with a lightweight token-wise MLP (Fusion):

$$z = \text{Fusion}\big([\, f_1; f_2; f_3 \,]\big) \in \mathbb{R}^{B \times N \times h}. \tag{2}$$

Phase-specific cross-attention and decoding. Two learnable query banks $Q^{\text{ED}}, Q^{\text{ES}} \in \mathbb{R}^{L \times h}$ (one query per landmark) attend to the fused tokens via multi-head cross-attention, yielding phase-aware embeddings

$$E^{\phi} = \text{MHA}(Q^{\phi}, z, z) \in \mathbb{R}^{B \times L \times h}, \qquad \phi \in \{\text{ED}, \text{ES}\}, \tag{3}$$

with $h = 32$ and 4 heads. Each embedding is decoded into a chord by a gated MLP and constrained to the image grid by a scaled tanh:

$$c^{\phi} = \text{GLU}\big(\text{MLP}(E^{\phi})\big) \in \mathbb{R}^{B \times L \times 4}, \qquad \ell^{\phi} = \tfrac{\rho}{2}\big(\tanh(c^{\phi}) + 1\big), \tag{4}$$

where each chord is parameterized by two endpoints (x_1, y_1, x_2, y_2) and $\rho = 112$ is the image size. Stacking both phases gives the predicted borders EBs $= \{\ell^{\text{ED}}, \ell^{\text{ES}}\} \in \mathbb{R}^{B \times 2 \times L \times 4}$ and landmark embeddings LEs $\in \mathbb{R}^{B \times 2 \times L \times h}$. EBs provide explicit border supervision and visualization, while LEs carry a phase-aware signal to the EF head.

2.3 Endocardial Feature Aggregator (E^{2}FA)

EF estimation also benefits from global descriptors of LV contractility over the cardiac cycle. E^{2}FA summarizes the deepest feature map x with three global statistics across the spatio-temporal dimensions, average, maximum, and variance concatenated into a global descriptor $x_{\text{glob}} \in \mathbb{R}^{B \times 3D}$. In parallel, the landmark embeddings are flattened and projected to the backbone width by MLP_{L}, giving a per-sample landmark descriptor $x_{\text{lnd}} = \text{MLP}_{\text{L}}(\text{flatten}(\text{LEs})) \in \mathbb{R}^{B \times D}$. Global and landmark descriptors are fused and regressed into a bounded EF:

$$F_{\text{final}} = [\, x_{\text{glob}}, x_{\text{lnd}} \,], \qquad \widehat{\text{EF}} = \tfrac{1}{2}\big(\tanh(\text{MLP}_{\text{head}}(F_{\text{final}})) + 1\big) \times 100. \tag{5}$$

2.4 Landmark-Based Geometric Training Objective

Simpson's method of disks estimates LV volume by dividing the chamber into a stack of elliptical disks along the LV long axis; EF then follows from end-diastolic and end-systolic volumes as EF = (EDV−ESV)/EDV. We translate this geometry into a differentiable regularizer over the predicted chords. The primary loss term is the mean-squared EF regression loss, $\mathcal{L}_{\mathrm{EF}} = \frac{1}{N}\sum_n(\widehat{\mathrm{EF}}_n - \mathrm{EF}_n^{\mathrm{gt}})^2$. For the geometric surrogate, we treat the chord i of phase ϕ as a disk cross-section: its diameter is the chord width $B_{\phi,i} = \|\mathbf{p}_{\phi,i}^{(1)} - \mathbf{p}_{\phi,i}^{(2)}\|_2$, and its height $H_{\phi,i}$ is the perpendicular distance from the center of chord $i-1$ to the line through chord i. These yield a disk stack with per-disk volume $\pi(B_{\phi,i}/2)^2 H_{\phi,i}$, recovering EDV/ESV. Three geometric terms supervise complementary aspects of this construction:

$$\mathcal{L}_{\mathrm{geo}} = \underbrace{\mathcal{L}_{\mathrm{pts}}}_{\text{chord localization}} + \underbrace{\mathcal{L}_{B}}_{\text{disk diameters}} + \underbrace{\mathcal{L}_{H}}_{\text{long-axis spacing}}, \tag{6}$$

where $\mathcal{L}_{\mathrm{pts}}$ is the MSE between predicted and reference chord endpoints, $\mathcal{L}_B$ and $\mathcal{L}_H$ match disk diameters and heights. Each term captures a distinct geometric property (landmark position, disk diameter, and long-axis spacing). The full objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{EF}} + \lambda_{\mathrm{geo}}\,\mathcal{L}_{\mathrm{geo}}, \qquad \lambda_{\mathrm{geo}} = 0.05, \tag{7}$$

with the geometric terms equally weighted under λ_{geo}.

3 Experiments

Datasets and Implementation. We evaluate on two public benchmarks. EchoNet-Dynamic [12] comprises 10,030 apical four-chamber (A4C) echocardiography videos with expert LVEF labels and ED/ES left-ventricular tracings; we use the official training, validation, and test splits. EchoNet-Pediatric [13] provides pediatric studies acquired in a parasternal short-axis (PSAX) view, where we follow an 80%/10%/10% split, yielding 3443 training, 430 validation, and 431 held-out test videos. Videos are processed at 112×112 resolution. Following prior work [9], we sample 64 frames at a sampling period of 2; for shorter clips we pad with the running average of previous frames. Echo-E^{3}Net is trained for 45 epochs on a single NVIDIA RTX 4070 (12 GB) with batch size 8, using AdamW with learning rate 5×10^{-4} and weight decay of 10^{-4}. A key practical advantage over prior work is that our model attains strong performance without external pre-training, or heavy data augmentation; in contrast, EchoGNN [8] relies on ED/ES classification pre-training, EchoCoTr [9] on vision-domain pre-trained weights, and EchoGraphs [18] on extensive augmentation. We measure agreement with mean absolute error (MAE), root mean squared error (RMSE), and R^2, and characterize efficiency with parameter count and GFLOPs.

3.1 Comparison with State of the Art

Table 1 compares Echo-E^{3}Net against recent EF regression baselines on EchoNet-Dynamic, each evaluated in its original configuration. Echo-E^{3}Net achieves competitive accuracy (MAE 3.92, RMSE 5.20, R^2 0.82) while operating at a fraction

Table 1: Comparison of EF estimation models on EchoNet-Dynamic. Parameter count in millions. Green indicates the best result and yellow the second best.

Model	**Frames**	**FLOPs**	**Params**	**MAE**$\downarrow$	**RMSE**$\downarrow$	$\mathbf{R^2}\uparrow$
UVT [14]	128	130.00G	-	5.95	8.38	0.52
MC3 [12]	32	97.656G	-	4.45	5.68	0.76
EchoGNN [8]	64	-	1.7	4.45	-	0.77
EchoNet-Dynamic [12]	32	91.974G	32	4.22	5.56	0.79
EchoGraphs [18]	16	40.739G	27.6	4.01	5.36	0.81
EchoCoTr [9]	36	44.907G	21	3.98	5.34	0.81
CoReEcho [6]	36	58.92G	21	3.90	5.13	0.82
CardiacNet [20]	16	7949G	28	3.83	-	-
Echo-E^{3}Net (Ours)	64	8.05G	1.55	3.92	5.20	0.82

of the computational cost. Specifically, at just 8.05 GFLOPs it reduces compute by 86.3% relative to CoReEcho [6] (58.92G) and by over 99.9% relative to CardiacNet [20] (7949G), and with only 1.55M parameters it is 92.7% and 94.5% more compact than these two methods, respectively. We further assess the real-time behavior of Echo-E^{3}Net by measuring inference latency. At the EchoNet-Dynamic acquisition rate of $\sim$50 fps, an input clip spans roughly 2.56 s of native acquisition, covering two to three cardiac cycles. Echo-E^{3}Net processes a clip in a median of 112.7 ms on a CPU-only workstation (12 threads, batch size 1), roughly 23$\times$ faster than the acquisition itself. EF is therefore available effectively as soon as the loop is recorded, without GPU hardware, supporting point-of-care use.

To test the generalization of our method, we then extend our evaluation to the EchoNet-Pediatric dataset under matched settings (Table 3). Echo-E^{3}Net achieves the best MAE (3.88) and R^2 (0.73) while remaining the most compact model, outperforming EchoCoTr [9] and CoReEcho [6] despite using an order of magnitude fewer parameters. Although the disk-based formulation is defined along the A4C long axis, the underlying supervision constrains ED and ES endocardial geometry. In PSAX videos, these constraints serve as a view-agnostic prior on LV shape and contraction, while the surrogate is not interpreted as exact disk integration. The consistent improvement on PSAX suggests that the anatomical signal remains useful beyond the apical view.

3.2 Ablation Study & Interpretability

Table 2 reports an ablation over the two modules and the three geometric loss terms. Removing either module degrades all three metrics: dropping E^{2}CBD (and, with it, the geometric supervision) raises RMSE from 5.20 to 5.40, and dropping E^{2}FA raises it to 5.31, confirming that explicit endocardial geometry and global contractility statistics are complementary. Each geometric loss term yields a small but consistent gain: removing $\mathcal{L}_{\text{pts}}$, $\mathcal{L}_B$, or $\mathcal{L}_H$ individually increases RMSE and reduces R^2 relative to the full model. This indicates the

Table 2: Ablation over Echo-E^{3}Net components on EchoNet-Dynamic. Results are averaged over 3 runs. Green best, yellow second best.

Component					**Metric**		
E^{2}CBD	E^{2}FA	$\mathcal{L}_{\text{pts}}$	$\mathcal{L}_B$	$\mathcal{L}_H$	MAE↓	RMSE↓	R^2↑
✔	✔	✘	✔	✔	3.98	5.30	0.81
✔	✔	✔	✘	✔	4.00	5.31	0.81
✔	✔	✔	✔	✘	3.95	5.29	0.81
✘	✔	–	–	–	4.05	5.40	0.80
✔	✘	✔	✔	✔	4.01	5.31	0.81
✔	✔	✔	✔	✔	3.92	5.20	0.82

Table 3: Generalization to EchoNet-Pediatric. Green best, yellow second best.

Model	Params	MAE↓	R^2↑
EchoGNN [8]	1.7 M	5.18	0.44
EchoCoTr [9]	21 M	4.13	0.70
CoReEcho [6]	21 M	4.44	0.67
Echo-E^{3}Net	1.55 M	3.88	0.73

terms are non-redundant, each constraining a distinct geometric property (landmark position, disk diameter, and long-axis spacing), and that the full Simpson-inspired objective is best.

Figure 3 shows a temporal Grad-CAM [16] sequence across the cardiac cycle for Echo-E^{3}Net and CoReEcho [6]. Our model's activations remain tightly concentrated on the left-ventricular cavity and endocardial border across frames, with limited response in surrounding tissue, whereas the baseline's are more diffuse. We further assess clinical utility for detecting heart failure with reduced ejection fraction (LVEF $\leq$ 40) on the EchoNet-Dynamic test set (n=1276, 160 positives). Using the continuous predicted LVEF as a decision score, Echo-E^{3}Net attains an area under the ROC curve of 0.981 (95% CI 0.973–0.988) and an average precision of 0.905 (95% CI 0.870–0.937), far exceeding the prevalence baseline of 0.125 (Figure 2).

4 Discussion and Limitations

Our proposed Echo-E^{3}Net, with explicit anatomical guidance embedded into LVEF estimation, achieves competitive accuracy at a fraction of the compute of existing methods, generalizes across views, and remains deployable on CPU hardware.

Several limitations remain. First, on EchoNet-Dynamic, our accuracy is on par

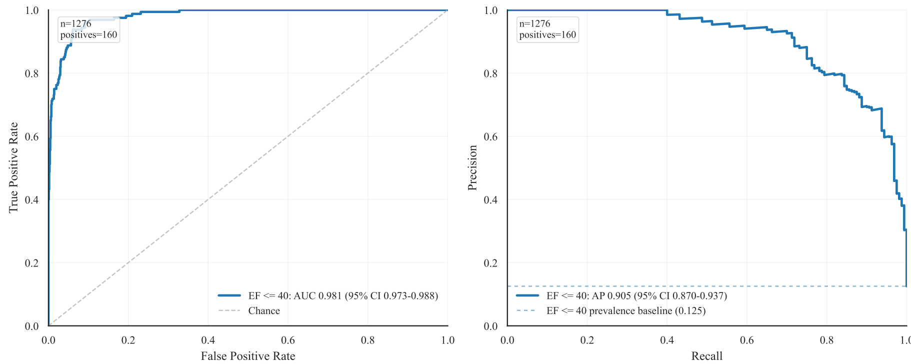

Fig. 2: Discrimination for reduced ejection fraction on the EchoNet-Dynamic test set using a threshold of LVEF $\leq$ 40. Left, receiver operating characteristic curve with area under the curve and confidence interval. Right, precision recall curve with average precision and confidence interval; the dashed line indicates the prevalence baseline.

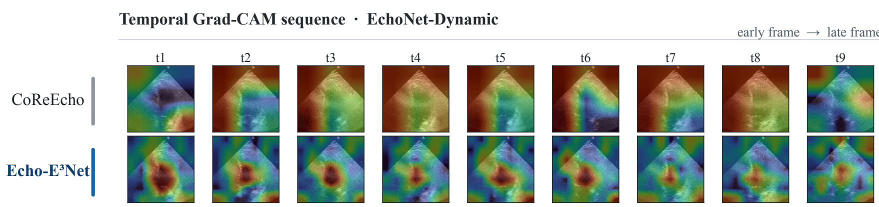

Fig. 3: Grad-CAM [16] across the cardiac cycle on an EchoNet-Dynamic video. Echo-E^{3}Net's saliency tracks the contracting left-ventricular cavity frame-to-frame, whereas CoReEcho's drifts into surrounding tissue, an effect of E^{2}CBD's explicit endocardial supervision.

with the strongest baselines; we view the present contribution as establishing a strong accuracy–efficiency operating point, and leave to future work the question of pushing accuracy higher while preserving the lightweight design. Second, although we evaluate on two views and two populations, generalization across scanners, vendors, and acquisition protocols remains to be established on external multi-centre cohorts. Third, the predicted endocardial chords serve only as a training-time geometric regularizer; producing clinically faithful, orderable contours suitable for segmentation would require explicit ordering supervision, which we leave to future work.

5 Conclusion

We introduced Echo-E^{3}Net, an efficient, anatomy-guided spatio-temporal network for LVEF estimation. A dual-phase endocardial border detector and an endocardial feature aggregator, trained with a Simpson-inspired geometric objective, deliver accuracy on par with the state of the art on EchoNet-Dynamic and the best results on EchoNet-Pediatric, using only 1.55M parameters and 8.05 GFLOPs with real-time CPU inference. These properties make Echo-E^{3}Net well suited to resource-constrained point-of-care ultrasound.

Acknowledgments. This work was supported by the Canadian Foundation for Innovation-John R. Evans Leaders Fund (CFI-JELF) program grant number 42816. Mitacs Accelerate program grant number AWD024298-IT33280. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2023-03575]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [RGPIN-2023-03575].

References

1. Dai, W., Li, X., Ding, X., Cheng, K.T.: Cyclical self-supervision for semi-supervised ejection fraction prediction from echocardiogram videos. IEEE Transactions on Medical Imaging **42**(5), 1446–1461 (2022)
2. Fraleigh, C.D., Duff, E.: Point-of-care ultrasound: An emerging clinical tool to enhance physical assessment. The Nurse Practitioner (2022)
3. Heidari, M., Azad, R., Kolahi, S.G., Arimond, R., Niggemeier, L., Sulaiman, A., Bozorgpour, A., Aghdam, E.K., Kazerouni, A., Hacihaliloglu, I., et al.: Enhancing efficiency in vision transformer networks: Design techniques and insights. arXiv preprint arXiv:2403.19882 (2024)
4. Kazemi Esfeh, M.M., Luong, C., Behnami, D., Tsang, T., Abolmaesumi, P.: A deep bayesian video analysis framework: towards a more robust estimation of ejection fraction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 582–590. Springer (2020)
5. Lai, S., Zhao, M., Zhao, Z., Chang, S., Yuan, X., Liu, H., Zhang, Q., Meng, G.: Echomen: Combating data imbalance in ejection fraction regression via multi-expert network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 624–633. Springer (2024)
6. Maani, F.A., Saeed, N., Matsun, A., Yaqub, M.: Coreecho: Continuous representation learning for 2d+ time echocardiography analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 591–601. Springer (2024)
7. Mokhtari, M., Ahmadi, N., Tsang, T.S., Abolmaesumi, P., Liao, R.: Gemtrans: A general, echocardiography-based, multi-level transformer framework for cardiovascular diagnosis. In: International Workshop on Machine Learning in Medical Imaging. pp. 1–10. Springer (2023)
8. Mokhtari, M., Tsang, T., Abolmaesumi, P., Liao, R.: Echognn: explainable ejection fraction estimation with graph neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 360–369. Springer (2022)
9. Muhtaseb, R., Yaqub, M.: Echocotr: Estimation of the left ventricular ejection fraction from spatiotemporal echocardiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 370–379. Springer (2022)

10. Muldoon, M., Khan, N.: Lightweight and interpretable left ventricular ejection fraction estimation using mobile u-net. IEEE ISBI (2023)
11. Ostvik, A.e.a.: Real-time automatic ejection fraction and foreshortening detection using deep learning. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2020)
12. 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., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
13. Reddy, C.D., Lopez, L., Ouyang, D., Zou, J.Y., He, B.: Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. Journal of the American Society of Echocardiography **36**(5), 482–489 (2023)
14. Reynaud, H., Vlontzos, A., Hou, B., Beqiri, A., Leeson, P., Kainz, B.: Ultrasound video transformers for cardiac ejection fraction estimation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 495–505. Springer (2021)
15. Sadegheih, Y., Bozorgpour, A., Kumari, P., Azad, R., Merhof, D.: Lhu-net: A lean hybrid u-net for cost-efficient, high-performance volumetric segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 326–336. Springer (2025)
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
17. Thomas, S., Cao, Q., Novikova, A., Kulikova, D., Ben-Yosef, G.: Echonarrator: Generating natural text explanations for ejection fraction predictions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 634–644. Springer (2024)
18. Thomas, S., Gilbert, A., Ben-Yosef, G.: Light-weight spatio-temporal graphs for segmentation and ejection fraction prediction in cardiac ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 380–390. Springer (2022)
19. Varudo, R.e.a.: Machine learning for the real-time assessment of left ventricular ejection fraction in critically ill patients. Critical Care (2022)
20. Yang, J., Lin, Y., Pu, B., Guo, J., Xu, X., Li, X.: Cardiacnet: Learning to reconstruct abnormalities for cardiac disease assessment from echocardiogram videos. In: European Conference on Computer Vision. pp. 293–311. Springer (2025)