When Measurement Conventions Masquerade as Calibration Gains in Cardiac Digital Twins

Dang P. M. Cao[1,2⋆] and Hieu Pham[1,2,3]

[1] College of Engineering and Computer Science, VinUniversity, Hanoi, Vietnam
[2] VinUni-Illinois Smart Health Center, VinUniversity, Hanoi, Vietnam
[3] Center for Innovations in Health Sciences, VinUniversity, Hanoi, Vietnam
24dang.cpm@vinuni.edu.vn, hieu.ph@vinuni.edu.vn

Abstract. Cardiac digital twins convert clinical images into physiological measurements through observation operators, yet calibration studies often assume a fixed reference convention. Across four shared-backbone echocardiographic EF front-ends, phase conditioning appears to remove CAMUS baseline bias. Matched-reference analysis rejects this gain: single-plane ground-truth EF error is statistically indistinguishable across models, while single-plane ground-truth EF exceeds CAMUS biplane clinical EF by +6.30 points, explaining nearly all baseline bias. A prespecified EchoNet-Dynamic replication, with released data and our extractor aligned to the apical four-chamber plane, removes baseline overestimation and reverses the CAMUS ranking. We also quantify haemodynamic effects, conformal residual-width budgets, and EF-stratum changes, yielding a Convention-Aware EF Audit protocol that separates genuine observation operator calibration from measurement artefacts. GitHub: Ejection-Fraction-Bias-in-Cardiac-Digital-Twin.git

Keywords: Cardiac digital twin · Ejection fraction calibration · Echocardiographic segmentation · Observation operator · Cross-dataset validation · Conformal prediction

1 Introduction

Cardiac digital twins integrate patient-specific data with mechanistic and statistical models [4,11,13]; recent pipelines use imaging and electrocardiography to personalise anatomy and physiology [3,15]. We call this image-to-measurement mapping the *observation operator* $\mathcal{H}_{\mathrm{op}}$. Though often treated as preprocessing, systematic error in $\mathcal{H}_{\mathrm{op}}$ biases downstream state updates regardless of model quality.

For LV function, $\mathcal{H}_{\mathrm{op}}$ typically segments echocardiograms before extracting volumes and EF. CAMUS [9] and EchoNet-Dynamic [12] advanced LV segmentation, commonly evaluated by Dice [9,12,20]. However, Dice is geometric; high

⋆ Cao Pham Minh Dang is the corresponding author. Email: 24dang.cpm@vinuni.edu.vn

overlap can still bias EDV, ESV, or their ratio and thereby the physiological innovation supplied to a twin.

We audit echo-based EF calibration using four front-ends sharing a U-Net [16]: frame-wise B1; adjacent-frame-smoothed B2; P1, combining FiLM [14] with a phase head motivated by echocardiographic phase characterisation [5]; and P2, adding cross-patient cyclic consistency and Mean-Teacher-style EMA averaging [18]. Rather than proposing an architecture or closed-loop twin, we isolate whether front-end design genuinely improves EF calibration.

Reference convention is the key challenge. CAMUS reports biplane EF from apical four- and two-chamber views [9], but our extractor uses one four-chamber plane. EchoNet-Dynamic provides four-chamber videos, clinical EF/volume labels, and expert ED/ES LV tracings [12]. CAMUS is therefore plane-mismatched; EchoNet is plane-aligned but may still differ in ED/ES selection, contours, preprocessing, and calculation. A genuine gain should survive removal of the dominant plane mismatch.

We contribute: (1) an EF observation-operator audit; (2) evidence that CAMUS phase gains reflect a measured single-plane–biplane gap; (3) a negative plane-aligned EchoNet replication; and (4) quantified twin-state, residual-width, and EF-stratum consequences within a Convention-Aware EF Audit protocol.

2 Methodology

2.1 EF as an observation operator

The operator maps an apical four-chamber clip to EF. Each phase volume uses a single-plane method-of-disks approximation adapted from standard chamber quantification [8]:

$$V = \sum_{i=1}^{N} \frac{\pi}{4} d_i^2 \frac{L}{N}, \qquad \widehat{\mathrm{EF}} = \frac{\widehat{\mathrm{EDV}} - \widehat{\mathrm{ESV}}}{\widehat{\mathrm{EDV}}} \times 100, \tag{1}$$

where L is LV long-axis length and d_i the i-th disk diameter; both datasets use this extractor.

In a Kalman formulation [7], EF is measurement y and $\hat{y} = \mathcal{H}_{\mathrm{op}}(\mathbf{x}^-)$ its prediction from prior state $\mathbf{x}^-$. Bias $\tilde{y} = y + b$ shifts the innovation by b and therefore the posterior; a zero-mean observation model cannot represent structured convention mismatch.

CAMUS provides biplane clinical EF [9], whereas EchoNet provides four-chamber videos, clinical EF/volume labels, and expert ED/ES tracings [12]. Our extractor is plane-mismatched on CAMUS and plane-aligned, but not process-matched, on EchoNet. Pixel spacing cancels in the EF ratio; its absence in EchoNet affects neither EF nor this audit, but prevents patient-specific absolute-volume propagation.

2.2 Front-ends and metrics

All front-ends share the U-Net backbone, optimiser, augmentation, and early stopping. **B1** is frame-wise; **B2** adds symmetric-KL adjacent-frame smoothing; **P1** combines FiLM [14] with a normalised phase head [5]; and **P2** adds phase-matched cross-patient cyclic consistency and Mean-Teacher-style EMA [18]. AdamW [10] used lr $= 10^{-4}$, batch size 4, and 600/400 CAMUS/EchoNet epochs. P1/P2 require two non-streaming passes; B2 streams.

We report **EF MAE$_{\text{gt}}$** against ground-truth-mask EF from the same extractor, isolating segmentation error; **EF MAE$_{\text{clin}}$** against clinical EF, including convention mismatch; and **EF MAE$_{\text{auto}}$** using ED/ES frames selected by maximum/minimum predicted LV volume.

Before EchoNet aggregation, we specified four checks: a genuine CAMUS effect should preserve baseline positive bias, phase-related bias reduction, comparable segmentation, and extractor-matched EF advantage.

3 Experiments and Results

3.1 Evaluation protocol

For CAMUS [9], we evaluate on our held-out validation subset of $n = 50$ patients drawn from the 450 released training cases; clinical EF uses the apical four- and two-chamber views. For EchoNet-Dynamic [12], we use the official test split of 1,288 studies and exclude one study with internally inconsistent clinical volume labels (ESV > EDV), yielding $n =$ 1,287. This exclusion was specified before model evaluation. All clips are resampled to 32 frames at 256×256 and z-score normalised. We use paired Wilcoxon tests, bootstrap 95% confidence intervals, Cohen's d_z, and Holm correction.

3.2 CAMUS: apparent calibration gain with two warning signs

CAMUS initially suggests that phase conditioning improves EF calibration: despite similar LV Dice, B1/B2 overestimate clinical EF, whereas P1/P2 move bias toward zero (Table 1). Evaluated only against clinical EF, this looks like a model-level gain. Two checks reject that interpretation.

Matched-reference EF error does not improve. When EF is computed from the same single-plane ground-truth masks used by our extractor, EF MAE$_{\text{gt}}$ is statistically indistinguishable across front-ends. The ranking also changes: P1, competitive against clinical EF, has the largest extractor-matched error and most negative bias.

The apparent gain is seed-unstable and Dice-insensitive. Baselines are seed-consistent, whereas phase-conditioned models vary more. P1 shows that similar Dice need not imply calibrated EF: small phase-dependent contour shifts at ED/ES can coherently alter volumes and be amplified by the EF ratio. This is a plausible mechanism, not component attribution; FiLM, cyclic consistency, and EMA remain entangled.

3.3 The CAMUS effect is explained by a convention offset

We measured the convention gap by applying our single-plane EF extractor to CAMUS ground-truth masks and comparing the resulting EF with biplane clinical EF.[1] Table 2 confirms the confound: the measured single-plane–biplane gap closely matches the baseline clinical bias. After subtracting this convention component, B1's residual segmentation bias has a confidence interval including zero. Thus the baseline is not substantially miscalibrated once both measurements use the same convention.

Table 1. CAMUS ($n = 50$). EF metrics are three-seed mean±std; Dice is the seed mean. Phase gain appears only against biplane clinical EF; matched single-plane GT errors are flat.

Model	Dice	$\mathrm{MAE_{gt}}$	$\mathrm{Bias_{gt}}$	$\mathrm{MAE_{clin}}$	$\mathrm{MAE_{auto}}$	$\mathrm{Bias_{clin}}$
B1	0.936	8.47±0.81	−1.37±0.58	11.39±0.28	13.15±0.62	+4.83±0.53
B2	0.934	**8.08±0.93**	−1.91±1.09	10.56±0.91	11.05±1.25	+4.61±1.61
P1	0.925	10.62±0.48	−7.56±1.82	9.97±0.35	9.68±0.86	−0.56±2.26
P2	0.932	8.94±1.26	−6.27±0.81	**9.24±1.41**	**9.16±0.23**	**+0.37±1.31**

The gap is not constant. Figure 1(d) shows a slope below one between single-plane and biplane EF, indicating a proportional discrepancy. A constant correction removes mean bias but leaves structured residual error, especially in high-EF patients.

Table 2. CAMUS convention-offset analysis ($n = 49$). The measured single-plane–biplane gap explains nearly all B1 clinical bias. †Post-hoc OLS; not a validated correction.

Quantity	Estimate	Uncertainty
Single-plane GT − biplane clinical	+6.30 EF pts	95% CI $[+3.69, +8.91]$
Observed B1 clinical EF bias	$+4.83 \pm 0.53$ EF pts	seed std
Convention component	+6.30 EF pts	95% CI $[+3.69, +8.91]$
Residual segmentation component	−1.47 EF pts	95% CI $[-4.08, +1.14]$
Post-hoc OLS slope†	0.613	95% CI $[0.47, 0.76]$
Post-hoc OLS intercept†	+13.34 EF pts	–

3.4 Plane-aligned replication on EchoNet-Dynamic

EchoNet-Dynamic tests the convention-confound hypothesis because labels and extractor are single-plane aligned, although their full measurement processes

[1] One CAMUS patient was excluded because the ground-truth mask produced zero end-systolic volume; this exclusion was specified before estimating the convention offset.

need not be identical. If the CAMUS result were a genuine model effect, baselines should remain positively biased and phase-conditioned models should reduce that bias. Table 3 shows the opposite: baseline bias changes sign, the CAMUS ranking does not replicate, and B2 is numerically strongest. This reversal is expected when the dominant plane mismatch is removed.

Deployment risk of phase-conditioned inference. Segmentation remains comparable, but P1/P2 produce invalid EF when $\widehat{\mathrm{ESV}} \geq \widehat{\mathrm{EDV}}$, consistent with ED/ES sensitivity in two-pass conditioning. This implementation-level result does not show that phase conditioning is generally harmful. A safety gate can flag failures; B2 remains the strongest audited deployable option, with streaming inference, no invalid EF, and the best plane-aligned EchoNet EF $\mathrm{MAE}_{\mathrm{clin}}$.

3.5 Downstream implications for cardiac digital twins

Haemodynamic propagation. We estimate downstream EF-bias effects using a simplified single-ventricle lumped calculation. With fixed reference EDV, $\Delta\mathrm{SV} = (\Delta\mathrm{EF}/100)\,\mathrm{EDV}_{\mathrm{ref}}$ and $\Delta\mathrm{CO} = \Delta\mathrm{SV}\,\mathrm{HR}_{\mathrm{ref}}/1000$, with $\mathrm{HR}_{\mathrm{ref}} = 72$ bpm. End-systolic elastance error uses a simplified pressure–volume relationship motivated by Suga and Sagawa [17]. This is order-of-magnitude propagation, not a patient-specific closed-loop simulation; Figure 2 summarises the errors.

Conformal observation-error budgets. We express EF residuals as split-conformal prediction intervals [19,1]. We interpret their empirical half-widths as observation-error budgets for comparing the evaluated measurement pipelines, not as direct estimates of a Kalman observation covariance. Under convention-confounded evaluation, the calibration residuals contain segmentation error and structured reference mismatch; under plane-aligned evaluation, they contain segmentation error, frame-selection error, and remaining measurement-process differences. The intervals aggregate these sources rather than identifying them separately (Table 5).

Table 3. EchoNet-Dynamic (n = 1,287 held-out after one prespecified exclusion). Under plane-aligned evaluation, the CAMUS phase-conditioning advantage disappears. Continuous metrics are mean±std over three seeds. Invalid EF counts $\widehat{\mathrm{ESV}} \geq \widehat{\mathrm{EDV}}$ as a separate deployment failure.

Model	Seeds	Invalid EF	Dice	$\mathrm{MAE}_{\mathrm{gt}}$	$\mathrm{MAE}_{\mathrm{clin}}$	$\mathrm{Bias}_{\mathrm{clin}}$
B1	3	0	$\mathbf{0.925 \pm 0.000}$	6.37 ± 0.11	6.02 ± 0.14	-1.72 ± 0.47
B2	3	0	0.924 ± 0.000	$\mathbf{6.29 \pm 0.09}$	$\mathbf{5.92 \pm 0.04}$	$\mathbf{-1.41 \pm 0.12}$
P1	3	10–20	0.918 ± 0.004	8.54 ± 0.50	8.55 ± 0.70	-5.97 ± 1.00
P2	3	8–11	0.922 ± 0.001	6.61 ± 0.21	6.33 ± 0.28	-2.72 ± 0.86

EF-stratum assignment. Using the LVEF cut points employed in the 2022 AHA/ACC/HFSA classification—$\leq$ 40%, 41–49%, and $\geq$ 50% [6]—we measure disagreement between EF strata. These strata are used only for sensitivity analysis and must not be interpreted as heart-failure diagnoses. Uncorrected CAMUS

Table 4. EchoNet replication decision specified before aggregation. Plane-aligned replication fails the checks expected under a genuine phase-conditioning calibration effect.

Criterion	Expected if CAMUS gain is genuine	Observed on EchoNet	Verdict
Baseline positive bias persists	B1/B2 remain positively biased	B1/B2 biases are negative	Fail
Phase models reduce baseline bias	P1/P2 move bias closer to zero	P1/P2 do not improve bias over B2	Fail
Segmentation remains comparable	Dice does not collapse	Dice remains comparable across front-ends	Pass
Extractor-matched EF advantage persists	Phase models improve $\mathrm{MAE_{gt}}$	B2 has lowest $\mathrm{MAE_{gt}}$	Fail

convention mismatch causes frequent EF-stratum disagreement (Table 6). P2 has the best apparent CAMUS calibration, but its disagreement rate overlaps B1 within uncertainty, so the apparent gain is not deployment-grade evidence.

4 Convention-Aware EF Audit Protocol

The results motivate a five-step Convention-Aware EF Audit (CAEA) protocol for EF observation operators. The purpose is simple: prevent a model from receiving credit for calibration gains caused by the measurement-reference pair.

1. **Compute EF $\mathbf{MAE_{gt}}$ under the extractor convention.** Apply the deployed EF extractor to ground-truth masks. If $\mathrm{MAE_{gt}}$ is flat while $\mathrm{MAE_{clin}}$ varies, treat the apparent clinical-reference gain as suspicious.
2. **Characterise the convention offset.** Estimate $\widehat{\delta}$ between GT EF under the extractor convention and clinical EF. Offsets comparable to claimed model gains are practically relevant. Check residual trends because a constant correction may remove mean bias but leave structured error.
3. **Check seed consistency.** Run independent seeds when feasible. A clinical EF advantage with large cross-seed variation is unstable, especially for phase-conditioned models where phase-estimation shifts can change ED/ES volumes.
4. **Add a deployment safety gate.** Flag $\widehat{\mathrm{ESV}} \geq \widehat{\mathrm{EDV}}$ as an internally inconsistent estimate. Separately flag $\widehat{\mathrm{EF}} < 5\%$ as a study-specific extreme-value quality-control threshold rather than a physiologically impossible value. Both conditions should trigger review before the estimate enters a downstream twin.
5. **Replicate under an aligned reference.** Treat gains as genuine only if they persist after aligning the measurement plane and auditing residual differences in frame selection, contours, preprocessing, and calculation.

CAEA is architecture-independent in design; validation on direct EF regressors, video models, stronger backbones, and hybrid pipelines remains required.

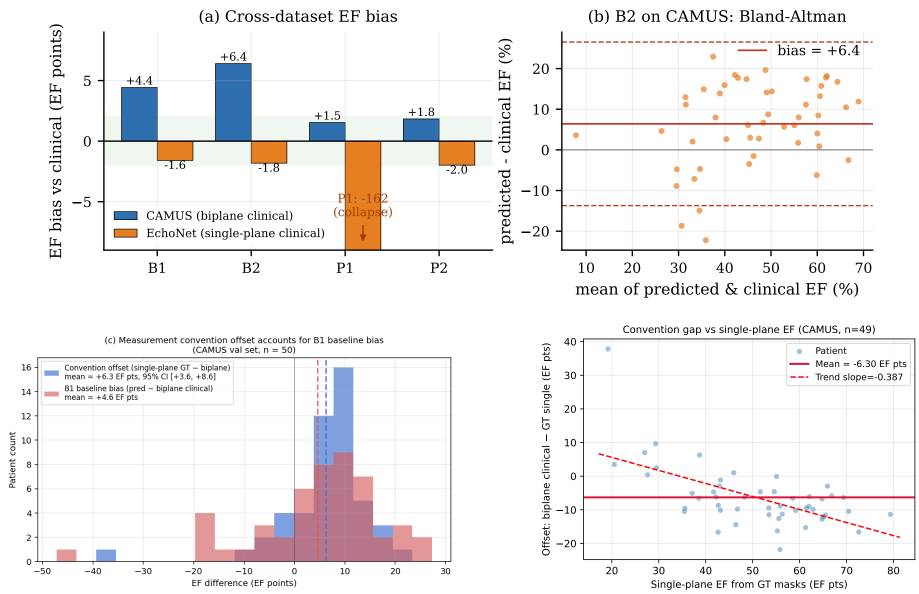

Fig. 1. Cross-dataset EF calibration audit. (a) CAMUS baselines overestimate clinical EF while phase-conditioned models appear near zero; plane-aligned EchoNet evaluation removes this pattern. Shaded band: ±2 EF pts. (b) B2 Bland–Altman on CAMUS [2]. (c) CAMUS convention offset ($n = 49$): single-plane GT EF exceeds biplane clinical EF by +6.30 points. (d) OLS mapping; slope < 1 indicates a proportional gap.

5 Discussion

Main lesson. Phase conditioning appeared to improve CAMUS EF calibration despite similar Dice, but the gain disappeared after plane alignment. Because EF and ventricular-volume estimates depend on imaging and measurement conventions [21], calibration should be assessed for the full model–extractor–reference pair, using extractor-matched EF, convention auditing, seed consistency, and external replication.

Clinical and digital-twin implications. Differences between single-plane and recommended biplane quantification are well recognised [8,21]; here, they create a benchmark artefact that favours phase-conditioned models without improving extractor-matched EF. The audit separates convention bias from model error and quantifies downstream effects. Among the evaluated front-ends, B2 is most deployable: it supports streaming, produces no invalid EF estimates, and achieves the best plane-aligned EchoNet EF $\mathrm{MAE}_{\mathrm{clin}}$.

Limitations. The audit covers two datasets, one EF extractor, and four shared-backbone U-Net front-ends; broader datasets and direct or video-based regressors remain untested. CAMUS includes only 49–50 patients, so EF-stratum and conformal analyses are exploratory. EchoNet aligns the imaging plane but not

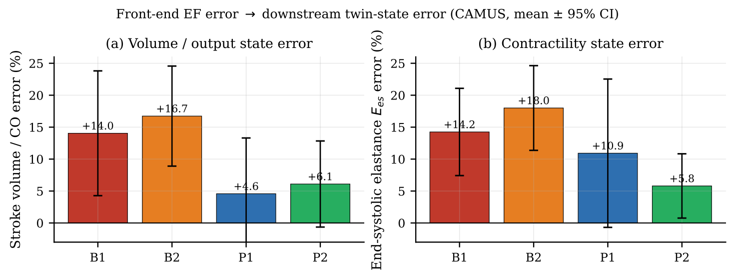

Fig. 2. Downstream twin-state errors under CAMUS biplane-reference conditions (mean±95% CI), using Section 3.5. Under plane-aligned EchoNet conditions, SV and CO errors fall to about 3–4%, so much of the penalty is recoverable after removing the convention gap.

necessarily ED/ES selection, contours, preprocessing, or calculation, and CMR validation is still needed. P2 entangles FiLM, cyclic consistency, and EMA, preventing general conclusions about phase conditioning. Haemodynamic propagation assumes fixed volumes and heart rate rather than a patient-specific closed-loop twin.

Acknowledgments. The authors thank Mr. Ha-Hieu Pham for his valuable advice on manuscript preparation and the refinement of the research ideas. The computational resources used for model training were provided by the VinUni-Illinois Smart Health Center (VISHC) Server infrastructure. Travel support for presenting this work at the workshop was provided by VinUniversity's Student Research Grant Program.

Table 5. Split-conformal EF observation-error budgets. Half-width in EF points. The squared-width index is $W_{\mathrm{rel}} = (h/h_{\mathrm{EchoNet,B2}})^2$; it is a comparative scale index, not a direct estimate of Kalman observation covariance. CAMUS uses a 50/50 calibration/evaluation split of our 50-patient validation subset, so its intervals are exploratory.

Dataset / reference	Model	90% half-width	Coverage	W_{rel}
CAMUS clinical, biplane-mismatched	B1	23.46	96.0%	2.89
CAMUS clinical, biplane-mismatched	P2	16.96	100.0%	1.51
EchoNet clinical, plane-aligned	B2	13.80	92.4%	1.00

References

1. Angelopoulos, A.N., Bates, S.: Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning **16**(4), 494–591 (2023). https://doi.org/10.1561/2200000101

Table 6. EF-stratum disagreement under CAMUS convention mismatch ($n = 50$). Strata use guideline LVEF cut points for sensitivity analysis and must not be interpreted as standalone heart-failure diagnoses.

Model	Different stratum / n	Rate	95% CI
B1	27/50	54%	39–68%
P2	22/50	44%	30–59%

2. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet **327**(8476), 307–310 (1986). https://doi.org/10.1016/S0140-6736(86)90837-8
3. Camps, J., Wang, Z.J., Doste, R., Berg, L.A., Holmes, M., Lawson, B., Tomek, J., Burrage, K., Bueno-Orovio, A., Rodriguez, B.: Harnessing 12-lead ECG and MRI data to personalise repolarisation profiles in cardiac digital twin models for enhanced virtual drug testing. Medical Image Analysis **100**, 103361 (2025). https://doi.org/10.1016/j.media.2024.103361
4. Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J.F., Bukhari, H.A., Wajdan, A., et al.: The 'digital twin' to enable the vision of precision cardiology. European Heart Journal **41**(48), 4556–4564 (2020). https://doi.org/10.1093/eurheartj/ehaa159
5. Dezaki, F.T., Dhungel, N., Abdi, A.H., Luong, C., Tsang, T., Jue, J., Gin, K., Hawley, D., Rohling, R., Abolmaesumi, P.: Deep residual recurrent neural networks for characterisation of cardiac cycle phase from echocardiograms. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Lecture Notes in Computer Science, vol. 10553, pp. 100–108. Springer (2017). https://doi.org/10.1007/978-3-319-67558-9_12
6. Heidenreich, P.A., Bozkurt, B., Aguilar, D., Allen, L.A., Byun, J.J., Colvin, M.M., Deswal, A., Drazner, M.H., Dunlay, S.M., Evers, L.R., Fang, J.C., Fedson, S.E., Fonarow, G.C., Hayek, S.S., Hernandez, A.F., Khazanie, P., Kittleson, M.M., Lee, C.S., Link, M.S., Milano, C.A., Nnacheta, L.C., Sandhu, A.T., Stevenson, L.W., Vardeny, O., Vest, A.R., Yancy, C.W.: 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. Circulation **145**(18), e895–e1032 (2022). https://doi.org/10.1161/CIR.0000000000001063
7. Kalman, R.E.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering **82**(1), 35–45 (1960). https://doi.org/10.1115/1.3662552
8. 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., 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. Journal of the American Society of Echocardiography **28**(1), 1–39.e14 (2015). https://doi.org/10.1016/j.echo.2014.10.003
9. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., Lartizien, C., D'hooge, J., Lovstakken, L., Bernard, O.: Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Transactions on Medical Imaging **38**(9), 2198–2210 (2019). https://doi.org/10.1109/TMI.2019.2900516

10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=Bkg6RiCqY7
11. Niederer, S.A., Sacks, M.S., Girolami, M., Willcox, K.: Scaling digital twins from the artisanal to the industrial. Nature Computational Science **1**(5), 313–320 (2021). https://doi.org/10.1038/s43588-021-00072-5
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., Zou, J.Y.: Video-based AI for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020). https://doi.org/10.1038/s41586-020-2145-8
13. Peirlinck, M., Sahli Costabal, F., Yao, J., Guccione, J.M., Tripathy, S., Wang, Y., Ozturk, D., Segars, P., Morrison, T.M., Levine, S., Kuhl, E.: Precision medicine in human heart modeling: Perspectives, challenges, and opportunities. Biomechanics and Modeling in Mechanobiology **20**(3), 803–831 (2021). https://doi.org/10.1007/s10237-021-01421-z
14. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32, pp. 3942–3951 (2018). https://doi.org/10.1609/aaai.v32i1.11671
15. Qian, S., Ugurlu, D., Fairweather, E., Dal Toso, L., Deng, Y., Strocchi, M., Cicci, L., Jones, R.E., Zaidi, H., Prasad, S., Halliday, B.P., Hammersley, D., Liu, X., Plank, G., Vigmond, E., Razavi, R., Young, A., Lamata, P., Bishop, M., Niederer, S.: Developing cardiac digital twin populations powered by machine learning provides electrophysiological insights in conduction and repolarization. Nature Cardiovascular Research **4**(5), 624–636 (2025). https://doi.org/10.1038/s44161-025-00650-0
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Suga, H., Sagawa, K.: Instantaneous pressure-volume relationships and their ratio in the excised, supported canine left ventricle. Circulation Research **35**(1), 117–126 (1974). https://doi.org/10.1161/01.RES.35.1.117
18. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems. vol. 30, pp. 1195–1204 (2017)
19. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer (2005). https://doi.org/10.1007/b106715
20. Wei, H., Cao, H., Cao, Y., Zhou, Y., Xue, W., Ni, D., Li, S.: Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science, vol. 12264, pp. 623–632. Springer (2020). https://doi.org/10.1007/978-3-030-59713-9_60
21. Wood, P.W., Choy, J.B., Nanda, N.C., Becher, H.: Left ventricular ejection fraction and volumes: It depends on the imaging method. Echocardiography **31**(1), 87–100 (2014). https://doi.org/10.1111/echo.12331