This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Beyond Per-Frame Accuracy: Temporally Consistent Carotid Strain Tracks Cardiovascular Risk

Alvaro Navarro-Guzmán[1], Virginia Mass[2], Inés García-Lunar[2], Valentin Fuster[2], Javier Sánchez-González[1] and Borja Ibáñez[2]

[1] Philips healthcare, Madrid, Spain
alvaro.navarroguzman@philips.com
[2] CNIC, Madrid, Spain

Abstract. Carotid strain is a promising marker of arterial stiffening and subclinical atherosclerosis, measured from the change in wall perimeter across the cardiac cycle. Per-frame wall segmentation is essentially solved, but the hard property for strain is temporal consistency, a stable contour across frames, which Dice misses. Validation against one expert's sparse annotations is doubly problematic: experts label few cycles, and a model trained on that expert imitates it, so high agreement can mislead. We therefore anchor and test against references the labels cannot provide: the ECG, and the ESC SCORE2 risk score that strain should track. On a held-out subgroup of participants from CNIC's prospective cardiovascular cohorts (40 participants, 71 scans) with synchronised ECG, a full-field multi-frame nnU-Net reproduced the expert frame by frame (Dice 0.97, perimeter r=0.99) and recovered the heartbeat (r=0.80, MAE 2.5 bpm). Dense automatic strain and the expert's sparse reading both tracked risk (ρ=−0.42 and −0.36), and after adjusting for age and blood pressure, only dense strain stayed significant (ρ_{adj}=−0.33, p_{adj}=0.045). Per-frame Dice was flat (0.93 to 0.97) while the SCORE2 association ranged from ρ=−0.26 for the single-frame model to −0.42 for temporal-context models, so per-frame accuracy does not predict clinical validity but temporal consistency does. A single-seed optical-flow tracker recovered similar association (ρ=−0.40), confirming temporal consistency as the active ingredient. When the reference is sparse, agreement is the wrong target: a consistent automatic strain, anchored to ECG and associated with SCORE2, offers a scalable cardiovascular-risk read-out from a routine carotid scan.

Keywords: Vascular ultrasound, Strain, Subclinical atherosclerosis, Temporal coherence, Segmentation, ECG

1 Introduction

Atherosclerosis progresses silently for years, so detecting subclinical disease and associated arterial ageing is central to prevention. Arterial stiffening is a functional manifestation, associated with atherosclerosis throughout the vascular tree [1], and the carotid offers an ultrasound window onto it [2]. Pulse-wave velocity is a validated marker of arterial stiffness [3], but it captures only the global condition of the vasculature and

cannot distinguish between territories such as the carotid and femoral arteries [4]. Vascular strain, measured from the relative change of the vessel-wall perimeter across the cardiac cycle, has been studied as a local screening marker for subclinical atherosclerosis [5-8].

Estimating strain requires contouring the wall in every frame of the ultrasound loop of images, following standardized carotid wall-measurement conventions [9]. Per-frame segmentation is essentially a solved problem: deep networks, and nnU-Net [10] in particular, delineate the carotid wall accurately on individual frames [11-13]. The genuinely hard property for strain is temporal consistency, a stable contour that behaves coherently from frame to frame. This problem has been studied predominantly in echocardiography [14–22]. One way to enforce it is to train a model with explicit temporal information (for example, a transformer). However, such training typically requires every frame to be annotated, which sharply limits the number of clinical cases available. We therefore explore approaches that do not depend on dense annotation.

For a temporal biomarker, per-frame accuracy is necessary but not sufficient, and agreement with a single expert tests imitation rather than correctness. Strain depends on the extremes of the cycle, so it amplifies exactly the frame-to-frame inconsistencies that a per-frame Dice score averages away: two models with identical Dice can differ by an order of magnitude in strain quality. Moreover, a model trained on one expert's sparse labels learns that expert's annotation style, so close agreement with the expert certifies imitation, not correctness.

In this work we study how different AI methodologies behave when estimating vascular strain, with and without temporal information, and we judge them by how well their strain tracks an external clinical reference, the ESC SCORE2 10-year cardiovascular-risk score [23]. We build an automatic, full-cine carotid-strain pipeline whose purpose is to certify and translate a temporal ultrasound biomarker learned from sparse, single-reader labels. Our contributions are:

1. **A single-expert validation** showing that a full-field, multi-frame estimator reproduces the expert frame by frame while remaining dense and fully automatic.
2. **An ECG-anchored, label-free check** that the extracted signal is genuinely cardiac, with heart-rate tracking that can serve as a deployable quality signal.
3. **A translational result**, dense automatic strain tracks continuous SCORE2 risk, and remains associated after adjusting for the age and blood pressure that SCORE2 already embeds.
4. **An optical-flow control** that recovers the same association without any per-frame segmentation, isolating temporal consistency as the active ingredient.

2 Materials and Methods

2.1 Controlled model variations

We compare a small family of models that differ only in a few controlled factors, so that any difference between them reveals which ingredient matters under sparse annotation. The U-Net is a light single-frame segmenter and serves as the jittery lower

bound. The 2D nnU-Net is a single-frame baseline. The 2.5D nnU-Net, our featured model, is multi-frame: it stacks each frame with its t±10 neighbours (early fusion). Two optical-flow trackers, Flow-U-Net and Flow-2.5D nnU-Net, are seeded from the U-Net and the 2.5D nnU-Net respectively. All share the same network family and the same single expert, so differences isolate the factor of interest and test whether the SCORE2 result holds across configurations.

The two optical-flow trackers use no per-frame segmentation. A single mask seeds one contour and its vertices are propagated through the loop by dense Farnebäck optical flow [24].

2.2 Training

The baseline is a MONAI BasicUNet trained on sparse annotated pairs. The nnU-Net models use nnU-Net v2 (self-configured 2D, 250 epochs, 5-fold participant-grouped cross-validation, ensemble Dice 0.974). The 2.5D model stacks each frame with its t±10 neighbours as additional channels. Inference is performed frame by frame, followed by a largest-connected-component clean-up.

2.3 Strain definition

For each binary wall mask, perimeter P is read from a fitted ellipse (Ramanujan's second approximation), with a solidity gate (≥0.92) rejecting collapsed masks. This is the perimeter form of circumferential carotid strain [5], obtained by transverse-wall fitting [25]. Pixels are uncalibrated, which scales P but cancels in the strain since strain is a within-scan ratio. Before cycle detection, P(t) is conditioned (gap interpolation, FFT period estimate, high-pass detrend, Savitzky-Golay smoothing). Peaks and troughs are then located, and each peak–trough–peak triplet defines one cardiac cycle. Strain is the mean per-cycle pulsation:

$$\varepsilon = \frac{1}{N_c}\sum_{c=1}^{Nc}\frac{{p_{max}}^{(c)} - {p_{min}}^{(c)}}{{p_{min}}^{(c)}}\,[\%]$$

P_{max} and P_{min} are read from the raw perimeter at extrema located on the smoothed trace (median over ±2 frames, to avoid outliers). Model strain is computed in two modes that share the same segmentation and differ only in sampling: matched-to-expert (annotated frames only) and dense (all cycles). The dense mode is the deployable estimator, and its consistency advantage comes from sampling every cycle.

2.4 Study cohort and reference measurements

We evaluate on a held-out subgroup of participants from CNIC's prospective cardiovascular cohorts scanned following standardized vascular ultrasound protocols [4, 26]. All networks were trained only on the expert's annotations from a development set (153 individuals, 303 scans, 1846 frames) and evaluated on a separate test set (40 participants, 80 scans). Of these 80 scans, 9 were excluded for quality reasons, leaving 71 analysed (484 frames). A single expert delineated the wall on sparse end-systolic and

end-diastolic frames with the help of ECG. These annotations also trained all models. Every loop carries a synchronised single-lead ECG, and each participant has a computed ESC SCORE2 10-year cardiovascular-risk value [23]. Transversal scans were acquired with a Philips XL14-3 probe, 1 cm below the carotid bifurcation, on left and right carotids. Clinical and demographic characteristics of the SCORE2-evaluable participants are summarised in Table 1.

Table 1. Clinical and demographic characteristics of the test cohort

Characteristic	Test cohort, n=40
Age, years	53.0 [51.0–56.0]
Male sex	27 (68%)
BMI, kg/m^2	26.9 [24.4–29.4]
Systolic BP, mmHg	124.5 [112.0–131.5]
Diastolic BP, mmHg	79.5 [74.8–88.0]
Hypertension	6 (15%)
LDL cholesterol, mg/dL	128.0 [113.8–150.2]
HDL cholesterol, mg/dL	56.5 [51.0–68.0]
Triglycerides, mg/dL	91.0 [61.5–129.0]
Ever smoker	6 (15%)
SCORE2, %	3.3 [2.5–4.7]

2.5 Statistical analysis and Metrics

We report per-frame Dice against the expert, Pearson r and Passing-Bablok regression [27] of model versus expert perimeter, Bland–Altman bias and limits [28] for strain, ECG concordance (heart-rate r and MAE, R-peak/trough locking), and the Spearman correlation of per-participant strain with continuous SCORE2 (95% CI via Fisher-z). Because SCORE2 embeds age and systolic blood pressure, we also report a partial Spearman correlation controlling for both. Per-participant strain is the mean over a participant's right and left territories. To isolate the mechanism, we compare matched-to-expert and dense sampling and check robustness across architectures.

3 Results

The automatic estimator reproduces the expert frame-by-frame. Dice is essentially flat across every configuration (0.93 to 0.97, Table 2). Figure 1 makes the evaluation chain concrete on the featured model over the 40-participant, 71-scan cohort. Per-frame Dice matches the single reader (median 0.97, Fig. 1A), and the per-frame wall perimeter tracks the expert almost perfectly (Pearson r = 0.99 over 484 frames, Passing-Bablok slope ≈ 1.0, Fig. 1B), so the geometry is correct frame by frame. Yet per-cycle strain still disagrees with the expert, at the level the sparse reference allows (Bland–Altman bias −0.9 pp, limits [−3.4, +1.6] pp, Fig. 1C). In other words, near-ceiling segmentation leaves a residual strain disagreement that is dominated by the sparseness of the human

reference, not by the segmenter. A Dice above 0.93 therefore makes per-frame overlap a poor selection metric for a temporal quantity such as strain, and motivates the label-free and clinical references that follow.

Table 2. Held-out evaluation chain (40 participants, 71 scans); the single-seed optical-flow estimator appears in the last row. Per-frame Dice does not predict the SCORE2 association: the selection metric is blind to what a translational temporal biomarker needs.

Configuration	Dice	Perim r	HR r	HR MAE (bpm)	BA bias (%)	BA LoA (%)
U-Net	0.946	0.98	0.78	2.59	-2.08	[-4.77, +0.61]
2D nnU-Net	0.973	0.99	0.72	2.83	-0.85	[-3.24, +1.53]
2.5D nnU-Net	**0.971**	**0.99**	**0.80**	**2.48**	**-0.88**	**[-3.41, +1.64]**
Flow-U-Net	0.927	0.95	0.81	2.16	-3.19	[-6.27, -0.11]
Flow-2.5D nnU-Net	0.939	0.96	0.85	1.72	-3.19	[-6.05, -0.33]

The automatic signal is genuinely cardiac. The signal built from per-frame segmentation reflects real wall pulsation. Comparing the captured ECG against the smoothed perimeter signal of the featured model, heart-rate concordance is reliable across the cohort ($r = 0.80$, MAE $\approx$ 2.5 bpm, Fig. 1D). Because the rhythm is recovered correctly on most scans, the residual disagreement with the expert reflects the sparseness of the manual reference rather than a failure to track the heartbeat. For a representative scan (292 frames, 30.2 fps, 8 cycles), the perimeter trace oscillates in phase with the ECG, with minima aligning to R-peaks beat by beat (Fig. 1E). Crucially, a correlation between a purely per-frame signal and the temporal ECG shows that the extracted signal is cardiac even without any explicit temporal modelling.

Dense sampling, not per-frame accuracy, drives the risk association. Stiffer arteries are known to correlate with age and cardiovascular risk factors, captured through risk scores such as SCORE2. Results show that they do and, numerically, more so for the automatic signal than for the expert. Regressing per-participant strain on continuous SCORE2 (Fig. 2), both show the expected negative association, the dense automatic strain numerically stronger on the identical participants ($\rho = -0.42$ vs -0.36, Table 3).

Because SCORE2 is built largely from age and systolic blood pressure, a raw correlation is partly mechanistic by construction, so we control for both. Every association attenuates after adjustment (Table 3): the expert ($p_{adj} = 0.057$), the matched-frames model ($p_{adj} = 0.098$) and the optical-flow tracker ($p_{adj} = 0.138$) all fall below significance, while the dense strain remains the only significant estimator ($\rho_{adj} = -0.33$, $p_{adj} = 0.045$). This implies that the association is not explained by age and systolic blood pressure alone, suggesting that strain carries information about cardiovascular risk beyond these two dominant drivers.

A controlled experiment isolates that mechanism. Constrained to the expert's frames, the model simply reproduces the expert ($\rho = -0.35$, $p = 0.027$, essentially identical to the expert's -0.36); freed to sample every cycle, it strengthens to $\rho = -0.42$ ($p = 0.007$). The gain comes from which phases are sampled, not from per-frame accuracy: the expert's one or two phases are a high-variance estimate, whereas dense sampling averages

every cycle and lowers the standard error. The isolated per-frame analyses of the AI and the human are therefore comparable, but complete time series yield a lower-variance biomarker than a few expert-annotated frames.

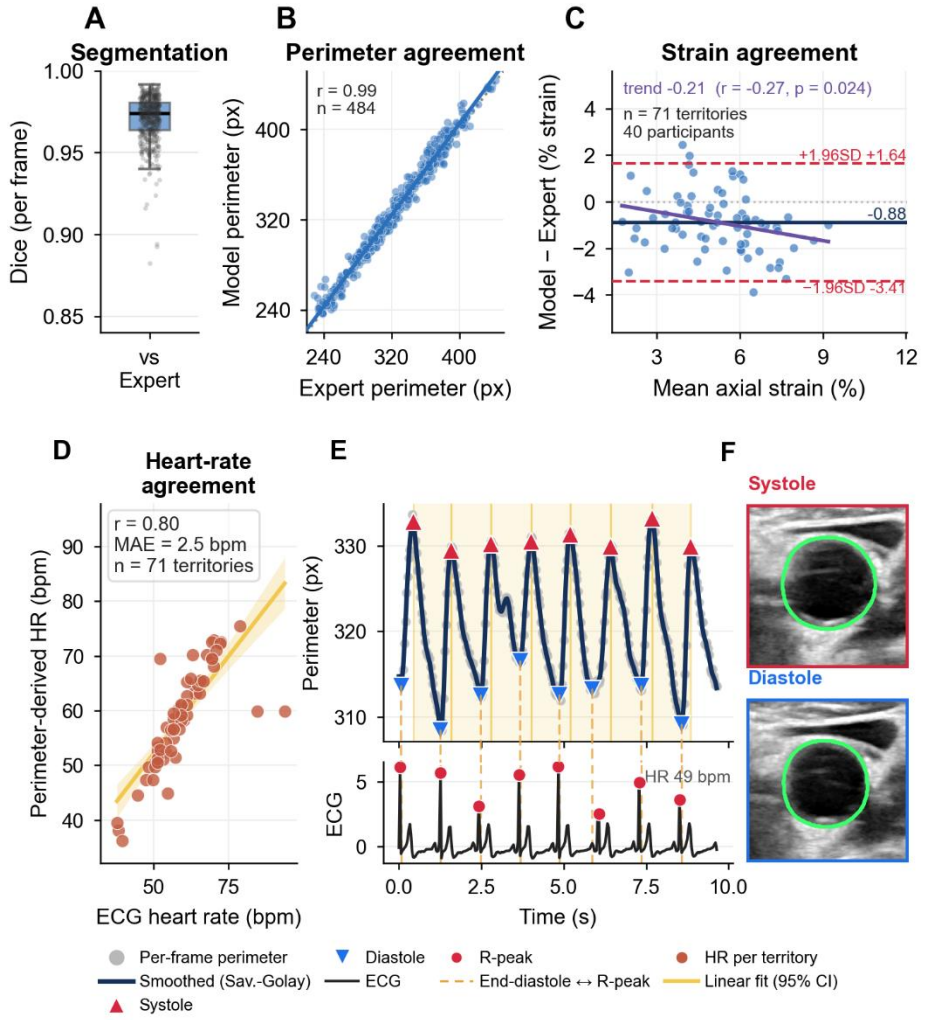

Fig. 1. Evaluation chain for the featured 2.5D nnU-Net (40 participants, 71 scans). (A) per-frame Dice vs. expert; (B) per-frame perimeter with Passing-Bablok fit; (C) per-cycle strain Bland-Altman; (D) heart-rate agreement vs. ECG; (E) one cine: perimeter trace with detected extrema, phase-locked to ECG R-peaks; (F) systole/diastole overlays. A–D cohort-wide; E–F illustrative.

Temporal consistency, not per-frame segmentation, is what carries the signal. The single-seed optical-flow estimator does no per-frame segmentation and tracks only one seed contour, yet it recovers the association ($\rho = -0.40$, $p = 0.011$) and still tracks the

heartbeat (r = 0.85, MAE 1.72 bpm). Its contour is slightly less accurate (median Dice 0.939 vs. 0.971), the expected cost of propagating a single drifting contour. This confirms that the biomarker reflects wall motion rather than a quirk of the segmentation.

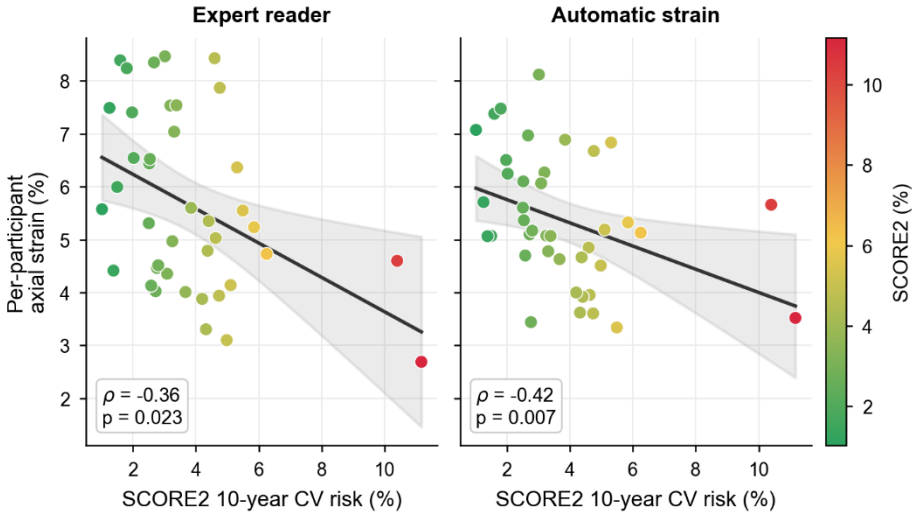

Fig. 2. Per-participant strain vs. continuous SCORE2 (n=40), expert (left) and dense automatic estimator (right); OLS fit with 95% band, Spearman ρ and p annotated. The automatic signal shows a ρ=−0.42 association despite never being trained for risk.

The seeding experiment then pins down the mechanism. Seeded from the 2.5D nnU-Net, flow adds nothing (ρ = −0.40, close to model's dense strain), because the model is already temporally consistent. Seeded instead from U-Net, the tracker reaches ρ = −0.33 (p = 0.035), stronger than that U-Net's own dense strain (ρ = −0.26, p = 0.104), because flow imposes the continuity the single-frame network lacks. But this gain is largely age and blood-pressure-correlated: it falls to ρ_{adj} = −0.13 (p_{adj} = 0.453) after adjustment, whereas the 2.5D nnU-Net model retains the only significant adjusted association (ρ_{adj} = −0.33). Flow-tracking from a jittery seed thus recovers a raw association but not the age/BP-independent signal that genuine multi-frame consistency carries.

Temporal consistency is therefore the property that carries the risk: it is supplied directly by a multi-frame network, partly recovered by flow-tracking, and absent from a single-frame segmenter alone. A full 3-D (spatiotemporal) model was outside our scope due to lack of dense ground truth segmentation. Whether it would improve on the multi-frame nnU-Net remains open, but our results indicate that temporal consistency, supplied here by 2.5D nnU-Net, is sufficient to recover the clinical association with CV risk scores. The learned segmenter remains the operating point of choice since it offers better fidelity and ECG locking at the same association. Single-seed flow serves as a robustness control rather than a replacement.

Table 3. Per-participant strain vs. continuous ESC SCORE2 (n = 40): raw Spearman ρ (95% Fisher-z CI) and a partial ρ_{adj} controlling for age and systolic blood pressure, the dominant stiffness drivers that SCORE2 contains. The featured model appears under two sampling regimes: the expert and single-seed trackers are references.

Strain source	SCORE2 raw ρ (95% CI)	p	SCORE2 adj ρ_{adj} (95% CI)	p_{adj}
Expert (sparse)	-0.36 (-0.60, -0.05)	0.023	-0.31 (-0.57, +0.01)	0.057
2.5D nnU-Net (expert-matched frames)	-0.35 (-0.60, -0.04)	0.027	-0.27 (-0.54, +0.05)	0.098
2.5D nnU-Net (dense)	**-0.42 (-0.65, -0.13)**	**0.007**	**-0.33 (-0.59, -0.01)**	**0.045**
U-Net (dense)	-0.26 (-0.53, +0.06)	0.104	-0.19 (-0.48, +0.14)	0.262
2D nnU-Net (dense)	-0.37 (-0.61, -0.07)	0.019	-0.29 (-0.56, +0.03)	0.074
Flow-U-Net (dense)	-0.33 (-0.58, -0.02)	0.035	-0.13 (-0.43, +0.20)	0.453
Flow-2.5D nnU-Net (dense)	-0.40 (-0.65, -0.13)	0.011	-0.24 (-0.55, +0.04)	0.138

4 Discussion

This study reframes how an automatic temporal ultrasound biomarker should be evaluated and translated. The conventional pipeline (pick a segmenter by Dice, then validate strain against one sparse expert) fails. First, Dice is flat (0.93 to 0.97) while the clinical signal runs from a non-significant −0.26 up to −0.42, so the selection metric is blind to what matters. Second, models trained on this sparse, high-variance reference imitate it by construction rather than achieving external validation.

The resolution is to test against references the labels cannot provide. The ECG establishes, without any annotation, that the oscillation is the wall's cardiac pulsation (R-peak locking, r = 0.80 heart-rate concordance) and provides a deployable, label-free quality signal. SCORE2 delivers the translational result: dense automatic strain tracks guideline risk on identical participants, with a numerically larger point estimate than the expert's sparse reading (ρ = −0.42 vs −0.36), despite never being trained for risk. The gain comes not from a better objective or a heavier network but from consistency: dense, every-cycle, full-field measurement turns the same segmentation skill into a lower-variance, more clinically valid biomarker. The optical-flow control confirms temporal consistency, whether supplied by the network or by tracking, as the active ingredient. A temporal biomarker should be judged by its consistency and against label-free references, not by per-frame overlap or sparse-reader agreement. We expect this principle to transfer to other 2D+time ultrasound biomarkers.

Clinical translation. Our cohorts focus on subclinical atherosclerosis in asymptomatic middle-aged adults, where an opportunistic risk read-out from a routine carotid scan would help most. SCORE2 is the ESC-recommended instrument for long-term risk and primary prevention. Aortic PWV tracks it at R = 0.495 in a comparable population cohort [29], a magnitude our local carotid strain matches with opposite sign (ρ = −0.42). These results position wall-motion strain to add mechanical information for

cardiovascular prediction on top of the clinical score at scale, particularly for plaque-free individuals.

Limitations. The SCORE2 association is cross-sectional, based on 40 participants and a single expert reader, so it shows strain correlates with overall cardiovascular risk but not independent prognostic value, which needs a larger cohort with event follow-up. As all models are trained and evaluated against this one reader, inter- and intra-observer variability and cross-annotator generalisation remain open. After adjusting for age and systolic blood pressure, only the dense strain remained significant, so independence from these dominant drivers is suggestive rather than established at this sample size. Perimeters are uncalibrated, which does not affect strain but means absolute dimensions are not reported. We therefore study associations with objective references such as ECG and SCORE2 rather than a second reader. Future work includes multi-annotator training, a paired test of the automatic–expert difference and a prospective association with remote atherosclerosis.

Acknowledgments. The Progression of Early Subclinical Atherosclerosis (PESA) study is co-funded by the Centro Nacional de Investigaciones Cardiovasculares (CNIC), and Santander Bank. The CNIC is supported by the Instituto de Salud Carlos III (ISCIII), the Ministry of Science and Innovation, and the Pro CNIC Foundation and is a Severo Ochoa Center of Excellence (grant CEX2020-001041-S, funded by MCIN/AEI/10.13039/501100011033). IGL is supported by the ISCIII (Project "PI24/00260", funded by ISCIII and co-funded by the European Union).

Disclosure of Interests. ANG and JSG are Philips employees. No additional conflicts of interest.

References

1. van Popele, N.M., Grobbee, D.E., Bots, M.L., et al.: Association between arterial stiffness and atherosclerosis: the Rotterdam Study. Stroke 32(2), 454–460 (2001)
2. Yang, W.I., Ha, J.W.: Non-invasive assessment of vascular alteration using ultrasound. Clin. Hypertens. 21, 25 (2015)
3. Laurent, S., Cockcroft, J., Van Bortel, L., et al.: Expert consensus document on arterial stiffness: methodological issues and clinical applications. Eur. Heart J. 27(21), 2588–2605 (2006)
4. Fernández-Friera, L., Peñalvo, J.L., Fernández-Ortiz, A., et al.: Prevalence, vascular distribution, and multiterritorial extent of subclinical atherosclerosis in a middle-aged cohort: the PESA study. Circulation 131(24), 2104–2113 (2015)
5. Park, H.E., Cho, G.Y., Kim, H.K., Kim, Y.J., Sohn, D.W.: Validation of circumferential carotid artery strain as a screening tool for subclinical atherosclerosis. J. Atheroscler. Thromb. 19(4), 349–356 (2012)
6. Bu, Z., Ma, J., Fan, Y., et al.: Ascending aortic strain analysis using 2-dimensional speckle tracking echocardiography improves the diagnostics for coronary artery stenosis in patients with suspected stable angina pectoris. J. Am. Heart Assoc. 7(14), e008802 (2018)
7. Podgórski, M., Grzelak, P., Kaczmarska, M., Polguj, M., Łukaszewski, M., Stefańczyk, L.: Feasibility of two-dimensional speckle tracking in evaluation of arterial stiffness. Vascular 26(1), 63–69 (2018)
8. Patton, et al.: Speckle tracking carotid artery circumferential strain is a marker of arterial sclerosis but not coronary atherosis. J. Clin. Ultrasound 46(9), 575–581 (2018)

9. Touboul, P.J., Hennerici, M.G., Meairs, S., et al.: Mannheim carotid intima-media thickness and plaque consensus (2004–2006–2011). Cerebrovasc. Dis. 34(4), 290–296 (2012)
10. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18(2), 203–211 (2021)
11. Zahnd, G., Kapellas, K., van Hattem, M., et al.: A fully-automatic method to segment the carotid artery layers in ultrasound imaging. Ultrasound Med. Biol. 43(1), 239–257 (2017)
12. Lainé, N., Zahnd, G., Liebgott, H., Orkisz, M.: Carotid artery wall segmentation in ultrasound image sequences using a deep convolutional neural network. arXiv:2201.12152 (2022)
13. Xie, M., Li, Y., Xue, Y., et al.: Vessel lumen segmentation in carotid artery ultrasounds with the U-Net convolutional neural network. In: IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM), pp. 2680–2684. IEEE (2020)
14. Leclerc, S., Smistad, E., Pedrosa, J., et al.: Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans. Med. Imaging 38(9), 2198–2210 (2019)
15. Ling, H.J., Garcia, D., Bernard, O.: Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net. In: IEEE Int. Ultrasonics Symp. (IUS). IEEE (2022)
16. Painchaud, N., Duchateau, N., Bernard, O., Jodoin, P.M.: Echocardiography segmentation with enforced temporal consistency. IEEE Trans. Med. Imaging 41(10), 2867–2878 (2022)
17. Hu, J., Smistad, E., Salte, I.M., Dalen, H., Lovstakken, L.: Exploiting temporal information in echocardiography for improved image segmentation. In: IEEE Int. Ultrasonics Symp. (IUS). IEEE (2022)
18. Ouyang, D., He, B., Ghorbani, A., et al.: Video-based AI for beat-to-beat assessment of cardiac function. Nature 580, 252–256 (2020)
19. DyL-UNet: a dynamic-learning framework for temporally consistent echocardiographic segmentation. arXiv (2025)
20. Lu, W., et al.: Echo-ODE: a dynamics-modelling network with neural ODE for temporally consistent segmentation of video echocardiograms. Front. Physiol. (2025)
21. EchoCardMAE: video masked auto-encoders customized for echocardiography. In: MICCAI. Springer (2025)
22. CoReEcho: continuous representation learning for 2D+time echocardiography analysis. In: MICCAI. Springer (2024)
23. SCORE2 working group and ESC Cardiovascular Risk Collaboration: SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur. Heart J. 42(25), 2439–2454 (2021)
24. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Scandinavian Conf. Image Analysis (SCIA), LNCS 2749, pp. 363–370. Springer (2003)
25. Golemati, S., Stoitsis, J., Sifakis, E.G., Balkizas, T., Nikita, K.S.: Using the Hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery. Ultrasound Med. Biol. 33(12), 1918–1932 (2007)
26. Fernández-Ortiz, A., Jiménez-Borreguero, L.J., Peñalvo, J.L., et al.: The Progression and Early detection of Subclinical Atherosclerosis (PESA) study: rationale and design. Am. Heart J. 166(6), 990–998 (2013) PubMed
27. Passing, H., Bablok, W.: A new biometrical procedure for testing the equality of measurements from two different analytical methods. J. Clin. Chem. Clin. Biochem. 21(11), 709–720 (1983)

28. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327(8476), 307–310 (1986)
29. Strömberg, S., Stomby, A., Engvall, J., Östgren, C.J.: Systematic Coronary Risk Evaluation 2 (SCORE2), arterial stiffness, and subclinical coronary atherosclerosis in a population-based study. Scand. J. Prim. Health Care 43(2), 455–462 (2025)