

Incremental value of resting electrocardiography and echocardiography for coronary artery disease: a retrospective cross-dataset evaluation

Tao Guo

Central People's Hospital of Zhanjiang

Huan Wang

Affiliated Hospital of Guangdong Medical University

Qingqing Niu

Affiliated Hospital of Guangdong Medical University

Shuhuai Chen

Affiliated Hospital of Guangdong Medical University

Yuxin Tian

Affiliated Hospital of Guangdong Medical University

Leying Chen

Affiliated Hospital of Guangdong Medical University

Mengjie Yang

Affiliated Hospital of Guangdong Medical University

Dan Yang

Affiliated Hospital of Guangdong Medical University

Chenglin Yang

Affiliated Hospital of Guangdong Medical University

Jieying Zhi

Affiliated Hospital of Guangdong Medical University

Bin Zhang

zhangbin57630@163.com

Central People's Hospital of Zhanjiang

Research Article

Keywords: coronary artery disease, electrocardiography, echocardiography, machine learning, incremental value, transportability, MIMIC-IV

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11018498/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Incremental value of resting electrocardiography and echocardiography for coronary artery disease: a retrospective cross-dataset evaluation

Authors:

Tao Guo^{1#}, Huan Wang^{2#}, Qingqing Niu², Shuhuai Chen², Yuxin Tian², Leying Chen²,
Mengjie Yang², Dan Yang², Chenglin Yang^{2*}, Jieying Zhi^{2*}, Bin Zhang^{3*}

Affiliations

1. Department of Ultrasound Medicine, Central People's Hospital of Zhanjiang, Zhanjiang,
Guangdong, China.

2. Digital Medicine Research Laboratory, Affiliated Hospital of Guangdong Medical
University, Zhanjiang, Guangdong, China.

3. Department of Cardiovascular Diagnostics, Central People's Hospital of Zhanjiang,
Zhanjiang, Guangdong, China.

[#] These authors contributed equally to this work and share first authorship.

^{*} Correspondence:

Chenglin Yang

Digital Medicine Research Laboratory, Affiliated Hospital of Guangdong Medical University,
Zhanjiang, Guangdong, China.

Email: chenglinyang2019@outlook.com

Jieying Zhi

Digital Medicine Research Laboratory, Affiliated Hospital of Guangdong Medical University,
Zhanjiang, Guangdong, China.

Email: 13432467291@163.com

Bin Zhang

Department of Cardiovascular Diagnostics, Central People's Hospital of Zhanjiang, Zhanjiang, Guangdong, China.

Email:zhangbin57630@163.com

Abstract

Background: Resting electrocardiography (ECG) and transthoracic echocardiography are widely available in patients evaluated for coronary artery disease (CAD), but their incremental value beyond clinical predictors and their transportability across data sources remain uncertain.

Methods: We performed a retrospective modeling study using a curated stenosis-defined angiographic cohort (Z-Alizadeh Sani; n = 303) and a real-world US cohort assembled from MIMIC-IV (n = 8,147). In MIMIC-IV, CAD was operationalized from diagnosis and revascularization records around coronary angiography, and candidate clinical, ECG, and echocardiographic features were extracted within a primary ± 30 -day peri-angiographic window. Clinical, single-modality, and fusion models were compared using five-fold out-of-fold predictions. We evaluated discrimination, calibration, reclassification, decision-curve utility, subgroup performance, temporal-window sensitivity, and cross-dataset transportability.

Results: In the curated cohort, adding ECG and echocardiography to the clinical pretest model increased the AUC from 0.894 to 0.926 (increase, 0.031; P = 0.021) and outperformed the CAD-Consortium basic score (AUC, 0.848; difference, 0.078; P = 0.00015). A broad laboratory panel did not improve discrimination (P = 0.330). In MIMIC-IV, the fusion model achieved an AUC of 0.833 (95% CI, 0.817-0.848), compared with 0.812 (95% CI, 0.795-0.828) for the clinical model (increase, 0.021; P = 5.30×10^{-7}), with a calibration slope of 0.991. The fusion advantage persisted after excluding troponin and across ± 7 - to ± 90 -day windows. Cross-dataset transfer was asymmetric: the curated-cohort model decreased from an internal

AUC of 0.845 to 0.682 in MIMIC-IV, whereas the MIMIC-IV model decreased from 0.806 to 0.789 in the curated cohort.

Conclusions: Resting ECG and echocardiography provided modest but reproducible incremental information beyond clinical predictors in two angiography-related public datasets. However, the MIMIC-IV endpoint was an EHR-derived phenotype and the primary feature window was peri-angiographic rather than strictly pretest. These findings support further pre-angiography validation in lower-prevalence populations before clinical use.

Keywords: coronary artery disease; electrocardiography; echocardiography; machine learning; incremental value; transportability; MIMIC-IV

Background

Invasive coronary angiography remains the reference procedure for defining obstructive coronary artery disease (CAD), but it is resource intensive and carries procedural risk. Contemporary European guidance emphasizes estimating the clinical likelihood of obstructive CAD and selecting subsequent testing accordingly rather than using invasive angiography as a universal first-line test [1]. Clinical likelihood models based on age, sex, symptoms, and cardiovascular risk factors therefore remain an important benchmark for evaluating whether additional tests provide meaningful incremental information [1,2].

Resting 12-lead ECG and transthoracic echocardiography capture complementary phenotypes. ECG can reflect ischemia, prior infarction, conduction abnormalities, and ventricular remodeling, whereas echocardiography provides information on systolic function, regional wall motion, cardiac structure, and valve disease [3,4]. Recent machine-learning studies have shown that combining clinical, laboratory, and ECG information can improve discrimination for obstructive CAD [13]. However, a high overall AUC does not establish how much a new modality adds beyond routine clinical predictors. In addition, performance from small curated

datasets may not transport to heterogeneous clinical data collected under different referral patterns and measurement processes.

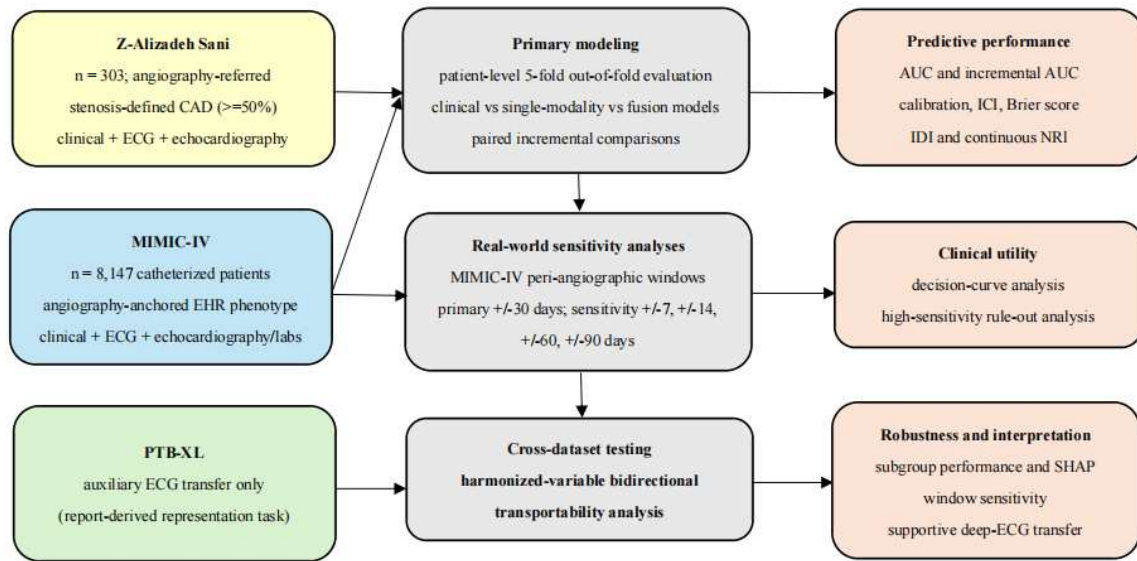
We therefore evaluated the incremental value of resting ECG and echocardiography in two public angiography-related datasets with different outcome ascertainment. The curated Z-Alizadeh Sani cohort used a stenosis-defined CAD outcome, whereas MIMIC-IV used an angiography-anchored EHR phenotype derived from diagnosis and revascularization records. We assessed discrimination, calibration, reclassification, decision-curve utility, subgroup performance, temporal-window sensitivity, and bidirectional cross-dataset transportability. We hypothesized that ECG and echocardiography would add complementary information, while transportability would depend on the population and outcome definition used for model development.

Methods

Study design and data sources

This retrospective prediction-model study used public or credentialed deidentified data. The Z-Alizadeh Sani dataset (UCI Machine Learning Repository dataset 412) contains 303 patients referred for coronary angiography at a single center, including 216 patients (71.3%) with CAD defined as at least 50% stenosis [7]. The real-world cohort was assembled from MIMIC-IV version 3.1 and linked PhysioNet resources [5], including MIMIC-IV-ECG [14] and MIMIC-IV-ECHO [15]. After applying the study criteria, 8,147 catheterized patients were retained, including 7,475 patients meeting the operational CAD phenotype and 672 patients not meeting that phenotype. PTB-XL version 1.0.3 was used only to assess transfer of the auxiliary deep-ECG representation task [6]. The study design is shown in Figure 1.

Study design and analysis workflow



Primary inference focuses on incremental value and cross-dataset transportability; the deep-ECG branch is supportive only.

Figure 1. Study design and analysis workflow. The diagram summarizes the two primary data sources, supportive ECG representation analysis, modality branches, multimodal fusion, and planned evaluation domains.

Outcome definition and temporal anchoring

In the curated cohort, CAD was defined by the source study as at least 50% luminal stenosis. In MIMIC-IV, the coronary angiography date served as the index date. CAD-positive status was operationalized using ischemic heart disease diagnosis codes or a record of percutaneous coronary intervention or coronary artery bypass grafting; CAD-negative status required angiography without a qualifying CAD diagnosis or revascularization record. This MIMIC-IV outcome should therefore be interpreted as an angiography-anchored EHR phenotype rather than a core-laboratory stenosis endpoint. Patients with pacemakers, major valvular disease, congenital heart disease, or cardiomyopathy were excluded. Candidate features were restricted to a ± 30 -day peri-angiographic window in the primary analysis, with sensitivity analyses using ± 7 -, ± 14 -, ± 60 -, and ± 90 -day windows. Because these windows can include post-index measurements, the MIMIC-IV analysis estimates peri-angiographic classification performance

and should not be interpreted as a strictly pretest prediction analysis. Cohort construction and analytic sample sizes are shown in Figure 2.

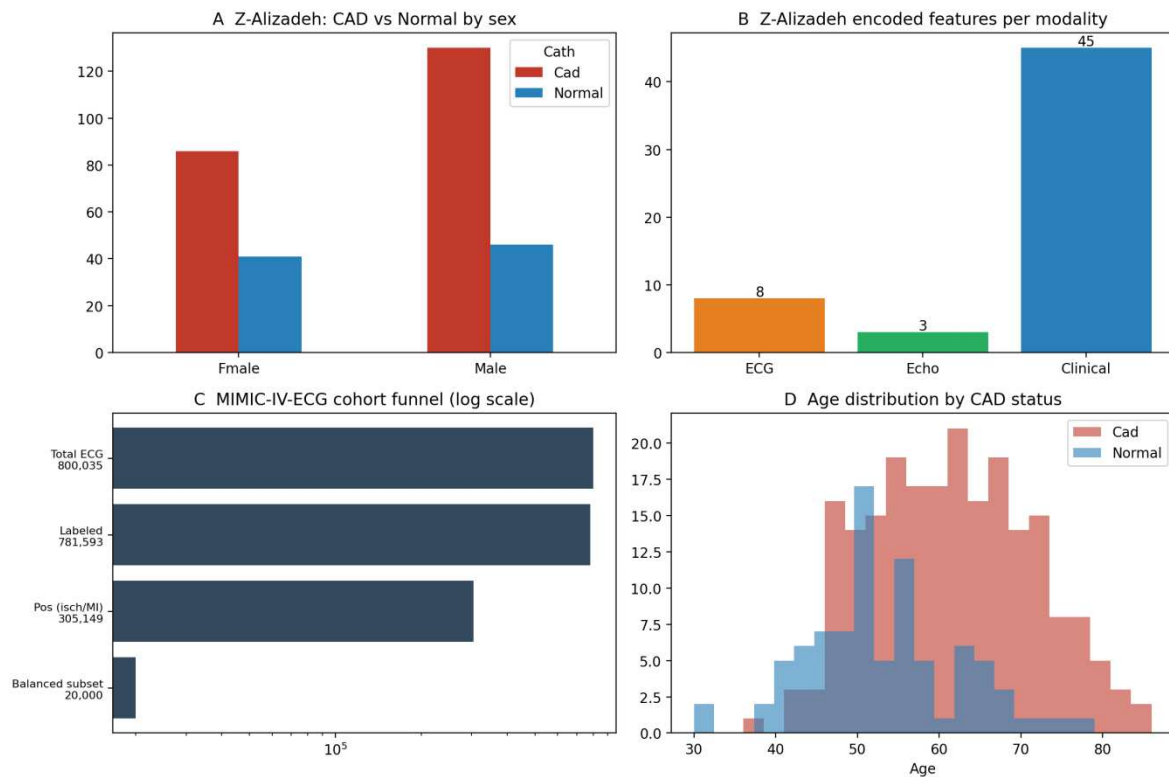


Figure 2. Cohort construction and analytic sample characteristics. Panels summarize sex distribution and modality feature counts in the curated cohort, the MIMIC-IV-ECG cohort funnel, and age distribution by CAD status.

Predictors and model development

The clinical domain included demographics, symptoms or comorbidities where available, vital signs, and laboratory measures. ECG variables included report-derived ischemic and conduction features, intervals, and axes. Echocardiographic variables included left ventricular ejection fraction, regional wall-motion abnormalities, valve severity, and related structured measurements. Cross-dataset analyses used harmonized variables with unit reconciliation. Logistic regression, random forest, and gradient-boosted tree models were evaluated. The principal MIMIC-IV comparison used XGBoost because it provided the strongest out-of-fold discrimination among the candidate algorithms. All folds were separated at the patient level.

An auxiliary deep-ECG branch used a shared one-dimensional convolutional lead encoder and anatomically constrained spatiotemporal graph models. Because the pretraining label was derived from machine-reported ECG findings rather than angiographic stenosis, this analysis was prespecified as supportive and was not used as the primary evidence for CAD discrimination.

Statistical analysis

Discrimination was summarized by receiver operating characteristic area under the curve (AUC) with 2,000 bootstrap replicates for 95% confidence intervals. Incremental AUC was tested with DeLong's method using paired out-of-fold predictions. Calibration was described by slope, intercept, integrated calibration index (ICI), and Brier score. Integrated discrimination improvement and continuous net reclassification improvement were calculated as complementary measures [8,9]. Decision-curve analysis estimated net benefit across decision thresholds. Subgroup analyses considered sex, age, and race where available; no subgroup model was proposed for clinical deployment. Analyses were performed using the reproducible scripts and fixed configuration included in the project package. Reporting was aligned with TRIPOD+AI recommendations [16], and potential sources of bias and limited applicability were interpreted with reference to PROBAST+AI [17].

Results

Curated angiographic cohort

In the Z-Alizadeh Sani cohort, the clinical pretest model achieved an AUC of 0.894. ECG and echocardiography each improved discrimination when added separately. Their combination achieved an AUC of 0.926 (95% CI, 0.894-0.955), an increase of 0.031 over the pretest model ($P = 0.021$). The full model containing a broad laboratory panel did not improve discrimination ($P = 0.330$) and had a lower calibration slope (Table 1).

149 **Table 1. Incremental performance in the Z-Alizadeh Sani cohort**

Model	AUC (95% CI)	Delta AUC vs pretest	P value	Calibration slope	ICI
Pretest	0.894 (0.848-0.935)	Reference	-	0.812	0.072
Pretest + ECG	0.917 (0.880-0.949)	+0.023	0.007	0.792	0.075
Pretest + echocardiography	0.913 (0.875-0.947)	+0.019	0.036	0.776	0.076
Pretest + ECG + echocardiography	0.926 (0.894-0.955)	+0.031	0.021	0.746	0.078
Full model (+ laboratory panel)	0.909 (0.874-0.941)	+0.015	0.330	0.502	0.056

150 Abbreviations: AUC, area under the receiver operating characteristic curve; ECG, electrocardiography; ICI, integrated

151 calibration index. P values compare paired out-of-fold predictions with the pretest model using DeLong's method.

152 The ECG-plus-echocardiography model yielded an integrated discrimination improvement of
153 0.085 ($P = 3.98 \times 10^{-5}$) and a continuous net reclassification improvement of 1.035. At a
154 threshold selected to preserve at least 90% sensitivity, the combined model classified 30.0% of
155 patients as rule-out candidates with a negative predictive value of 0.769, compared with 29.0%
156 and 0.761 for the pretest model. The CAD-Consortium basic score achieved an AUC of 0.848
157 and underestimated absolute risk in this high-prevalence referral cohort; the combined model
158 exceeded it by 0.078 ($P = 0.00015$). Incremental performance is shown in Figure 3.

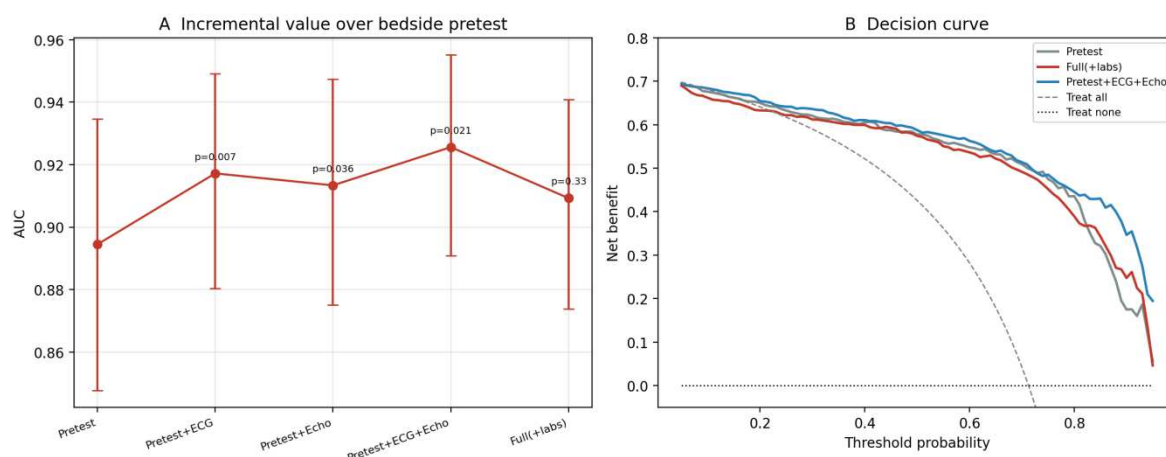


Figure 3. Incremental value of ECG and echocardiography. A, change in AUC relative to the clinical pretest model. B, decision-curve analysis across threshold probabilities.

Real-world evaluation in MIMIC-IV

The MIMIC-IV cohort had an operational CAD prevalence of 91.8%. The fusion model achieved an AUC of 0.833 (95% CI, 0.817-0.848), compared with 0.812 (95% CI, 0.795-0.828) for clinical predictors alone (increase, 0.021; $P = 5.30 \times 10^{-7}$; Table 2). The fusion model had a calibration slope of 0.991, an intercept of -0.060, an ICI of 0.079, and a Brier score of 0.062. Modality-specific and fusion-model performance is shown in Figure 4.

Table 2. Out-of-fold discrimination in MIMIC-IV

Model (XGBoost)	AUC (95% CI)
Clinical	0.812 (0.795-0.828)
ECG	0.632 (0.610-0.654)
Echocardiography	0.622 (0.601-0.644)
Fusion	0.833 (0.817-0.848)

Abbreviations: AUC, area under the receiver operating characteristic curve; ECG, electrocardiography; XGBoost, extreme gradient boosting. Values are based on patient-level out-of-fold predictions.

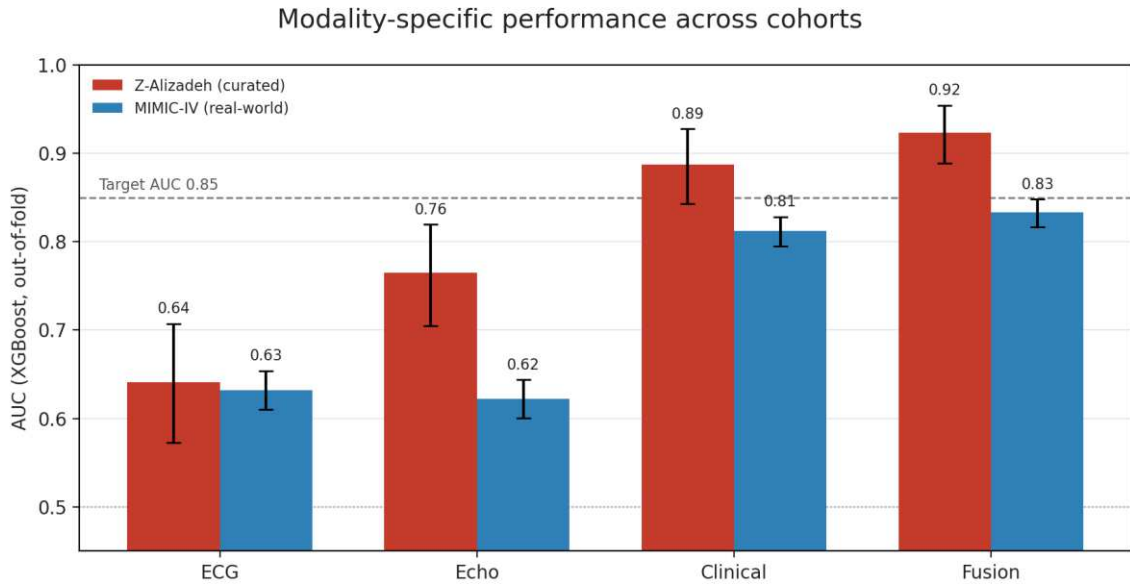


Figure 4. Modality-specific and fusion-model performance across cohorts. Bars show out-of-fold AUC estimates with uncertainty for ECG, echocardiography, clinical predictors, and multimodal fusion.

Incremental discrimination was supported by an integrated discrimination improvement of 0.022 ($P = 1.42 \times 10^{-7}$) and a continuous net reclassification improvement of 0.276. Excluding troponin reduced absolute performance but increased the observed fusion gain: AUC 0.807 versus 0.773 for the clinical model (increase, 0.034; $P = 6.23 \times 10^{-11}$). The fusion benefit remained positive across peri-angiographic windows from ± 7 to ± 90 days. Adding ICU vital-sign trajectories did not improve performance (change in AUC, -0.001; $P = 0.44$) and was limited by 39% coverage. Temporal-window sensitivity is shown in Figure 5; decision-curve and rule-out analyses are shown in Figure 6.

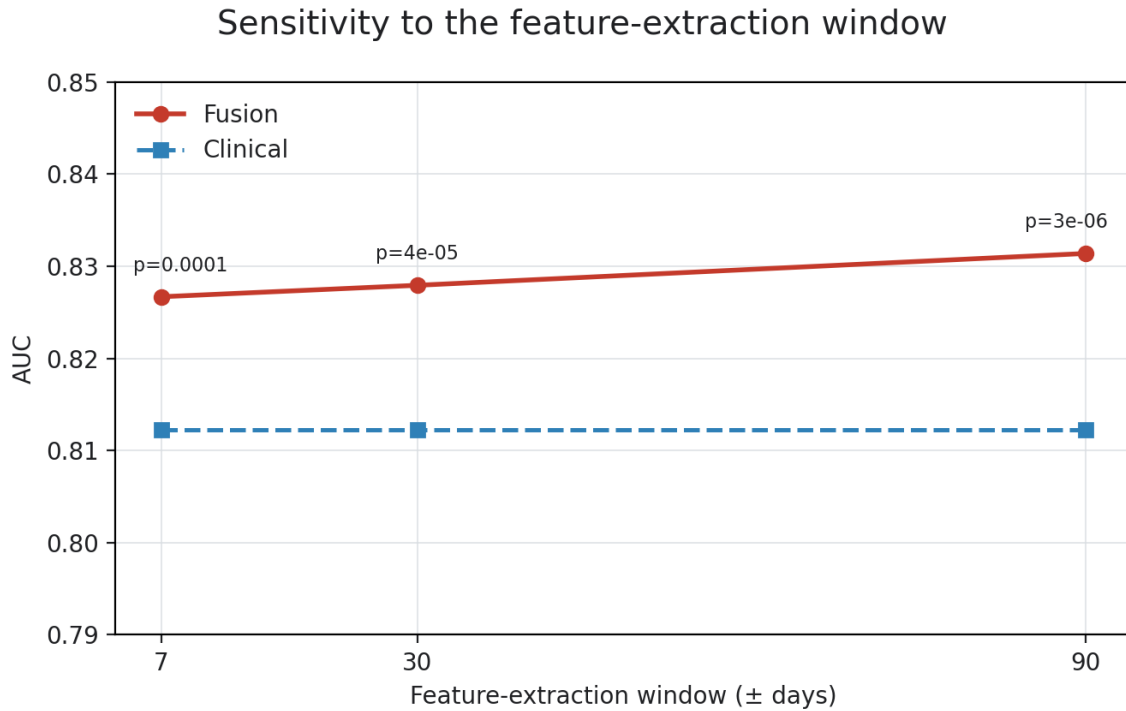


Figure 5. Sensitivity to the feature-extraction window. Clinical and fusion-model AUCs are shown across selected peri-angiographic windows; annotations indicate paired comparisons.

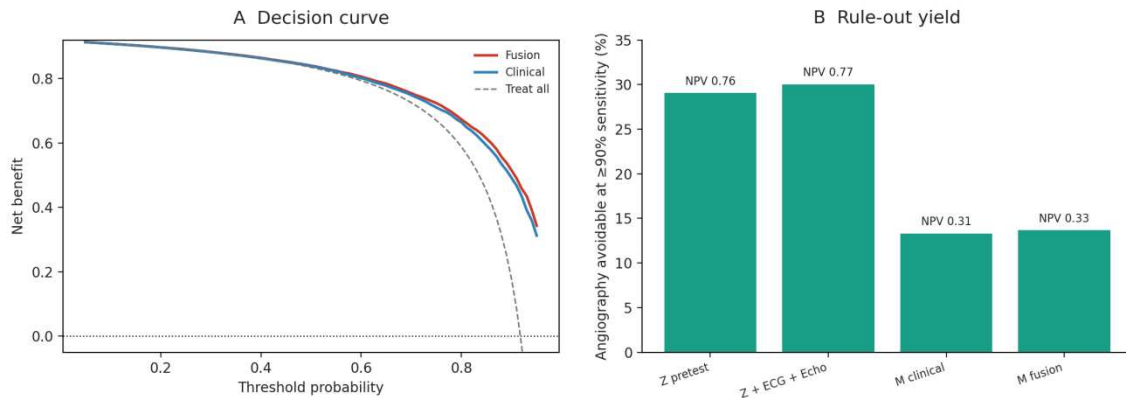


Figure 6. Decision-curve and rule-out analyses in MIMIC-IV. A, net benefit across threshold probabilities. B, proportion eligible for rule-out at the prespecified high-sensitivity operating point, with negative predictive value annotations.

Cross-dataset transportability

Transportability was strongly asymmetric (Table 3). The model developed in the curated cohort lost 0.163 AUC when evaluated in MIMIC-IV and was markedly miscalibrated. In the

reverse direction, the MIMIC-IV-trained model lost 0.018 AUC when evaluated in the curated cohort. Because the datasets differed in population, measurement processes, and outcome ascertainment, these analyses are best interpreted as cross-dataset transportability tests rather than validation of an identical target phenotype. The bidirectional comparison is shown in Figure 7.

Table 3. Bidirectional cross-dataset transportability

Training to test cohort	Internal AUC	External AUC (95% CI)	AUC decrease	External calibration slope
Z-Alizadeh Sani (Iran) to MIMIC-IV (United States)	0.845	0.682 (0.661-0.704)	0.163	0.336
MIMIC-IV (United States) to Z-Alizadeh Sani (Iran)	0.806	0.789 (0.729-0.845)	0.018	1.289

Internal AUC denotes out-of-fold performance in the development dataset; external AUC denotes performance after transfer to the other dataset. The datasets differ in population and outcome ascertainment, so this analysis is interpreted as transportability rather than validation of an identical phenotype.

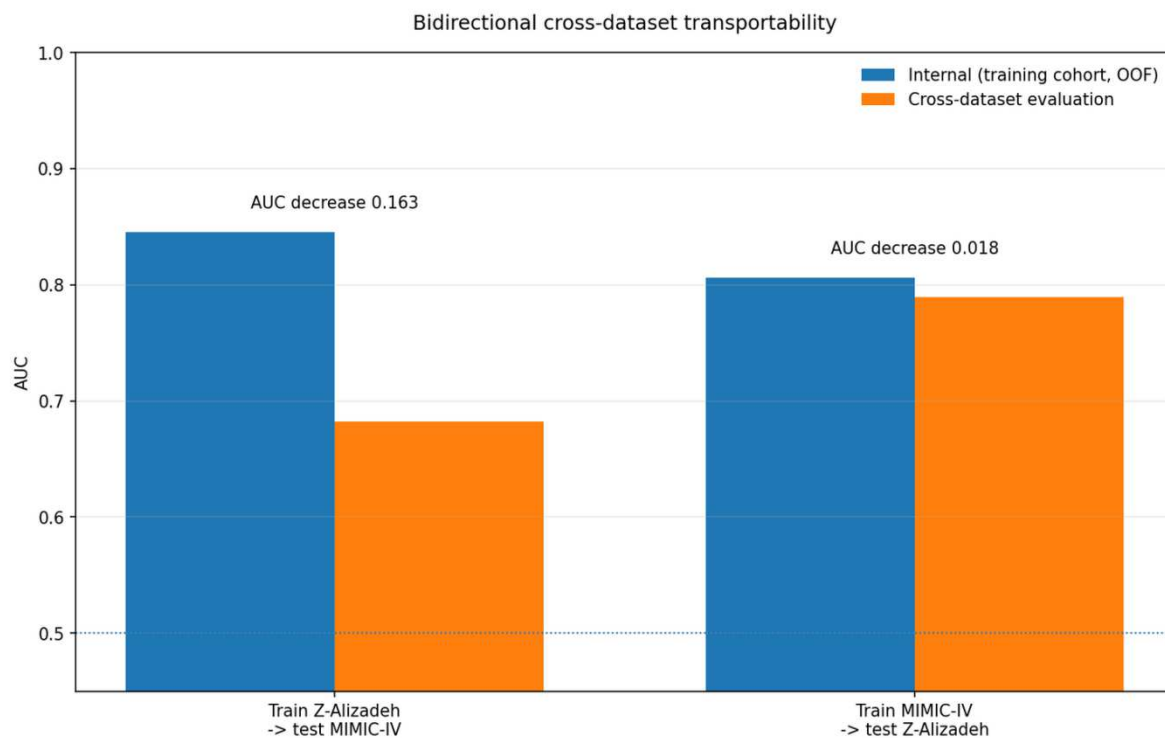


Figure 7. Bidirectional cross-dataset transportability of harmonized models. Internal out-of-fold AUC is compared with performance after transfer to the other dataset.

Subgroup performance and explainability

In MIMIC-IV, the maximum subgroup AUC gap was 0.074. Discrimination was lower in women than in men (AUC, 0.779 vs 0.843), declined with age, and ranged from 0.786 to 0.853 across race groups. These exploratory analyses identify potential limitations rather than establish fairness or subgroup-specific clinical validity. SHAP attribution indicated that clinical, echocardiographic, and ECG domains accounted for approximately 79%, 16%, and 5% of aggregate importance, respectively [11,12]. Subgroup performance is shown in Figure 8.

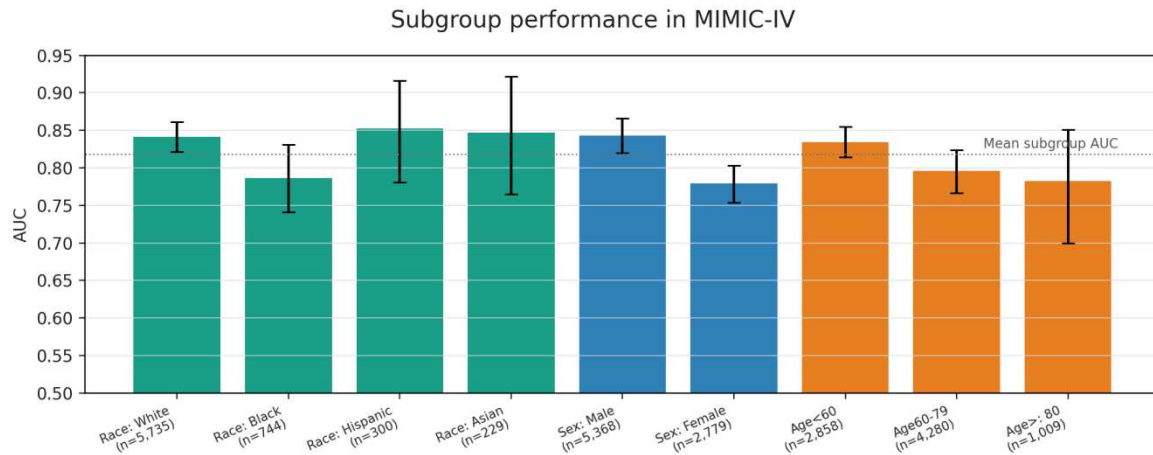


Figure 8. Subgroup performance in MIMIC-IV. AUC estimates with uncertainty are shown across race, sex, and age strata; analyses are exploratory and are not intended to establish fairness.

Supportive deep-ECG analysis

On the auxiliary machine-report task, a shared lead encoder trained in MIMIC-IV-ECG transferred to PTB-XL with AUCs above 0.92. However, the resting-ECG branch achieved an AUC of only 0.659 for the angiography-anchored MIMIC-IV CAD phenotype. This contrast shows that strong performance on report-derived ischemia labels should not be interpreted as equivalent evidence for obstructive CAD.

Discussion

This study has three main findings. First, resting ECG and echocardiography added modest, complementary discrimination beyond a clinical pretest model in the curated stenosis-defined cohort, and their combination outperformed the CAD-Consortium basic score. Second, a positive incremental signal was also observed in the much larger MIMIC-IV cohort and persisted after excluding troponin and across alternative peri-angiographic windows. Third, cross-dataset transfer was strongly directional: the model developed in heterogeneous real-world data was more transportable to the curated cohort than the reverse. These findings

should be interpreted in the context of different outcome definitions in the two datasets rather than as replication of an identical angiographic endpoint.

The size of the incremental gain is important. In MIMIC-IV, the fusion model improved AUC by 0.021 over the clinical model, despite an overall AUC of 0.833. This distinction between overall discrimination and incremental value is clinically relevant because a multimodal model can appear highly accurate while most of its performance is driven by clinical predictors. A contemporary study of 7,907 patients similarly reported that combining clinical, laboratory, and ECG data improved prediction of obstructive CAD [13]. Our analysis extends that perspective by explicitly quantifying paired incremental performance, calibration, reclassification, decision-curve utility, and cross-dataset transportability. The limited rule-out gain and low negative predictive value in MIMIC-IV are expected in a catheterized cohort with 91.8% phenotype prevalence and should not be extrapolated to outpatient screening.

Temporal interpretation is a key limitation of the real-world analysis. The primary MIMIC-IV window extended from 30 days before to 30 days after angiography, and the sensitivity analyses also used symmetric windows. Consequently, some predictor measurements may have occurred after the index procedure. The current MIMIC-IV results therefore support peri-angiographic classification rather than a strictly pretest diagnostic use case. A future analysis restricted to measurements obtained before angiography, with prespecified rules for selecting the most recent pre-index value, would provide a stronger test of prospective clinical utility.

The directional transportability result is methodologically informative but should not be overinterpreted. Curated datasets can encode narrow referral patterns and cleaner measurement processes, whereas MIMIC-IV contains greater heterogeneity and measurement noise. At the same time, MIMIC-IV is derived from a single health system with an inpatient and critical-care emphasis, and its operational CAD phenotype is based on diagnosis and procedure records

rather than core-laboratory stenosis [5]. Differences in outcome definition, prevalence, missingness, referral, and measurement may all contribute to the asymmetric transfer. Subgroup differences likewise require caution: lower discrimination in women and older patients may reflect differences in presentation, referral, measurement, or label quality, and the present analyses were not designed to certify equitable deployment [11].

Strengths and limitations

Strengths include the use of two public angiography-related datasets, patient-level out-of-fold evaluation, explicit temporal anchoring, paired incremental testing, calibration and decision-curve assessment, window sensitivity analyses, subgroup evaluation, and bidirectional cross-dataset testing. Several limitations are important. First, both primary cohorts were selected around coronary angiography and had high CAD prevalence, limiting spectrum applicability and transport of absolute risk. Second, the MIMIC-IV endpoint was an EHR-derived phenotype based on diagnoses and revascularization rather than a stenosis-defined angiographic reference standard. Third, the symmetric peri-angiographic windows can include post-index information, preventing a strict pretest interpretation of the MIMIC-IV analysis. Fourth, structured ECG and echocardiographic measurements reflect clinical acquisition and missingness processes that may introduce selection bias. Fifth, the small curated cohort limits model complexity and subgroup inference, and selecting the best-performing algorithm from candidate models may add optimism if model-selection procedures are not fully nested. Finally, this was a retrospective study without prospective clinical-impact evaluation. The supportive deep-ECG analysis used report-derived labels and should not be interpreted as independent validation of obstructive CAD.

Conclusions

Resting ECG and echocardiography provided modest but reproducible incremental information beyond clinical predictors across two angiography-related public datasets. The combined model was more informative than either modality alone, while a broad laboratory panel did not add discrimination in the curated cohort. Cross-dataset transfer was asymmetric, favoring model development in the larger heterogeneous real-world cohort. Because the MIMIC-IV outcome was EHR-derived and its feature windows were peri-angiographic, these results should be viewed as evidence for further evaluation rather than clinical deployment. Prospective, pre-angiography validation in lower-prevalence populations, rigorous recalibration, and prespecified subgroup assessment are needed before clinical use.

List of abbreviations

AUC, area under the receiver operating characteristic curve; CABG, coronary artery bypass grafting; CAD, coronary artery disease; ECG, electrocardiography/electrocardiogram; EHR, electronic health record; ICI, integrated calibration index; PCI, percutaneous coronary intervention; SHAP, SHapley Additive exPlanations; XGBoost, extreme gradient boosting.

Declarations

Ethics approval and consent to participate

This secondary analysis used deidentified public or credentialed datasets and involved no new participant recruitment or intervention. The original data collections were conducted under their respective institutional approvals, and access to restricted MIMIC resources requires credentialing, completion of required human-subjects training, and acceptance of the PhysioNet data use agreement.

Consent for publication

Not applicable.

299 **Clinical trial registration**

300 Not applicable.

301 **Availability of data and materials**

302 The Z-Alizadeh Sani dataset is publicly available from the UCI Machine Learning Repository
303 (<https://doi.org/10.24432/C5Q31T>) [18]. MIMIC-IV version 3.1 is available through
304 PhysioNet to credentialed users under its access requirements
305 (<https://doi.org/10.13026/kpb9-mt58>) [19]. MIMIC-IV-ECG, MIMIC-IV-ECHO, and
306 PTB-XL are available through PhysioNet under their stated access conditions [6,14,15].
307 Dataset licenses and data use agreements prohibit redistribution of credentialed source data.
308 Analysis code, fixed configuration files, and figure-generation scripts are contained in the
309 accompanying project package.

310 **Competing interests**

311 The authors declare that they have no competing interests.

312 **Funding**

313 No funding

314 **Authors' contributions**

315 According to the CRediT taxonomy, the contributions were as follows: BZ: Conceptualization,
316 Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft,
317 Visualization. HW: Conceptualization, Methodology, Validation, Formal analysis,
318 Investigation, Data curation, Writing – original draft. QQN: Data curation, Formal analysis,
319 Investigation, Visualization. SHC: Data curation, Formal analysis, Validation. YXT: Data
320 curation, Investigation. LYC: Data curation, Investigation. MJY: Data curation,
321 Investigation. DY: Data curation, Investigation. TG: Conceptualization, Resources,
322 Supervision, Writing – review & editing. CLY: Conceptualization, Methodology, Resources,

Supervision, Project administration, Writing – review & editing. JYZ: Conceptualization, Methodology, Resources, Supervision, Project administration, Writing – review & editing. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Bin Zhang and Huan Wang contributed equally to this work and share first authorship.

Acknowledgements

Not applicable.

References

- [1] Vrints C, Andreotti F, Koskinas KC, et al. 2024 ESC Guidelines for the management of chronic coronary syndromes. *Eur Heart J*. 2024;45:3415-3537. doi:10.1093/eurheartj/ehae177.
- [2] Genders TSS, Steyerberg EW, Hunink MGM, et al. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. *BMJ*. 2012;344:e3485. doi:10.1136/bmj.e3485.
- [3] Kligfield P, Gettes LS, Bailey JJ, et al. Recommendations for the standardization and interpretation of the electrocardiogram. *J Am Coll Cardiol*. 2007;49:1109-1127. doi:10.1016/j.jacc.2006.12.012.
- [4] Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults. *J Am Soc Echocardiogr*. 2015;28:1-39.e14. doi:10.1016/j.echo.2014.10.003.
- [5] Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10:1. doi:10.1038/s41597-022-01899-x.
- [6] Wagner P, Strodthoff N, Bousseljot RD, et al. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data*. 2020;7:154. doi:10.1038/s41597-020-0495-6.

- 347 [7] Alizadehsani R, Zangooei MH, Hosseini MJ, et al. Coronary artery disease detection using
348 computational intelligence methods. *Knowl Based Syst.* 2013;48:1-10.
349 doi:10.1016/j.knosys.2013.04.001.
- 350 [8] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added
351 predictive ability of a new marker. *Stat Med.* 2008;27:157-172. doi:10.1002/sim.2929.
- 352 [9] Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: development,
353 internal validation, and assessing incremental value. *Heart.* 2012;98:683-690.
354 doi:10.1136/heartjnl-2011-301246.
- 355 [10] Steyerberg EW, Moons KGM, van der Windt DA, et al. PROGRESS 3: prognostic model
356 research. *PLoS Med.* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381.
- 357 [11] Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine
358 learning to advance health equity. *Ann Intern Med.* 2018;169:866-872.
359 doi:10.7326/M18-1990.
- 360 [12] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural*
361 *Inf Process Syst.* 2017;30:4765-4774.
- 362 [13] Lee HG, Park SD, Bae JW, et al. Machine learning approaches that use clinical, laboratory,
363 and electrocardiogram data enhance the prediction of obstructive coronary artery disease.
364 *Sci Rep.* 2023;13:12635. doi:10.1038/s41598-023-39911-y.
- 365 [14] Gow B, Pollard T, Nathanson LA, et al. MIMIC-IV-ECG: Diagnostic Electrocardiogram
366 Matched Subset (version 1.0) [dataset]. *PhysioNet.* 2023.
367 <https://doi.org/10.13026/4nqg-sb35>.
- 368 [15] Gow B, Pollard T, Greenbaum N, et al. MIMIC-IV-ECHO: Echocardiogram Matched
369 Subset (version 1.0) [dataset]. *PhysioNet.* 2026. <https://doi.org/10.13026/nrjh-5r77>.

- 370 [16]Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for
371 reporting clinical prediction models that use regression or machine learning methods. BMJ.
372 2024;385:e078378. doi:10.1136/bmj-2023-078378.
- 373 [17]Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias,
374 and applicability assessment tool for prediction models using regression or artificial
375 intelligence methods. BMJ. 2025;388:e082505. doi:10.1136/bmj-2024-082505.
- 376 [18]Alizadehsani R, Roshanzamir M, Sani Z. Z-Alizadeh Sani [dataset]. UCI Machine
377 Learning Repository. 2013. <https://doi.org/10.24432/C5Q31T>.
- 378 [19]Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S, et al. MIMIC-IV (version
379 3.1) [dataset]. PhysioNet. 2024. <https://doi.org/10.13026/kpb9-mt58>.
- 380