Prediction of Fatal Cardiovascular Events After First Acute ST-Segment Elevation Myocardial Infarction Using Multidomain Data

Hamid Reza Marateb[1], Marjan Mansourian[2], Kouhyar Tavakolian[3], Hamidreza Roohafza[4], Masoumeh Sadeghi[4], Miquel Angel Mañanas[1]

[1]Institute for Research and Innovation in Health, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

[2]Epidemiology and Biostatistics Department, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran

[3]University of North Dakota, Grand Forks, United States

[4]Cardiovascular Research Institute, Isfahan University of Medical Sciences, Isfahan, Iran

Abstract

Fatal cardiovascular events after first acute ST-segment elevation myocardial infarction remain difficult to predict when psychosocial and lifestyle information is omitted. Following the published CRAS-MI methodology and ethics approval IR.NIMAD.REC.1397.295, we analyzed 1,730 modeling records from a multicenter cohort with annual follow-up over 3 years. The endpoint was fatal major adverse cardiovascular events, defined as cardiovascular death from coronary artery or cerebrovascular disease. The dataset contained 79 fatal outcomes (4.6%); mean age was 56.2 +/- 9.9 years, mean body mass index was 26.6 +/- 4.1 kg/m2, and 82.3% were male. A TabNet classifier used clinical, laboratory, lifestyle and psychosocial predictors, derived interaction terms, tuned class weighting, repeated stratified five-fold cross-validation and isotonic calibration. Out-of-fold evaluation yielded AUROC 0.836 (95% CI 0.799-0.877), Brier score 0.037 (0.030-0.044), sensitivity 0.709 (0.612-0.814), and specificity 0.795 (0.776-0.814). Feature importance showed that clinical measures, sleep efficiency and psychosocial variables carried complementary prognostic signal.

1. Introduction

Patients who survive first acute ST-segment elevation myocardial infarction remain vulnerable to fatal cardiovascular events. Conventional secondary prevention scores mostly emphasize traditional clinical factors and may underuse behavioral, sleep-related and psychosocial information. The CRAS-MI cohort was designed to address this gap by collecting clinical, paraclinical, lifestyle and psychological predictors in a multicenter first-infarction population [1].

This study developed and internally validated a calibrated artificial intelligence model for the fatal component of major adverse cardiovascular events. Reporting followed the TRIPOD+AI domains for data source, participants, outcome definition, predictors, model specification, validation, calibration, uncertainty and interpretation [2].

2. Materials and methods

2.1. Cohort and endpoint

The modeling dataset contained 1,730 analysis-ready records from the CRAS-MI multicenter cohort. Eligible participants were recruited after first acute ST-segment elevation myocardial infarction and were followed annually for 3 years. The modeled endpoint was fatal major adverse cardiovascular events, namely cardiovascular death due to coronary artery disease or cerebrovascular disease, consistent with the submitted abstract and cohort protocol [1].

2.2. Predictors and learning pipeline

Predictors covered demographic, socioeconomic, admission, treatment, laboratory, lifestyle and psychosocial domains. A TabNet classifier was selected because sequential attention can model tabular data and provides feature importance for interpretation [3]. Derived interaction terms and tuned class weighting addressed nonlinearities and the rare fatal endpoint; the final run used a positive-class weight multiplier of 4.0.

Model assessment used out-of-fold predictions from 16 repeated stratified five-fold cross-validation runs, providing 80 validation folds. Probabilities were calibrated after model fitting; isotonic regression was retained because it minimized the Brier score in the supplied results [4,5]. The operating threshold was 0.05 under a minimum specificity constraint of 0.60. Uncertainty was estimated by 1,000 bootstrap resamples.

Table 1. Cohort, model, and calibrated out-of-fold performance.

Records/events	1,730 records; 79 fatal events (4.6%)
Validation	TabNet; 16 x 5-fold stratified out-of-fold validation
Calibration/threshold	Isotonic calibration; threshold 0.05; positive-class weight multiplier 4.0
Confusion matrix	TP/FP/FN/TN = 56/339/23/1312
Discrimination/calibration	AUROC 0.836 (0.799-0.877); AUPRC 0.245 (0.167-0.346); Brier 0.037 (0.030-0.044)
Threshold metrics	Sensitivity 0.709 (0.612-0.814); specificity 0.795 (0.776-0.814); PPV 0.142 (0.111-0.177); NPV 0.983 (0.976-0.990)
Overall metrics	Accuracy 0.791 (0.772-0.810); balanced accuracy 0.752 (0.703-0.804); F1 0.236 (0.188-0.287); MCC 0.250 (0.198-0.307)

3. Results

The final calibrated analysis remained aligned with the submitted abstract in design, endpoint, sample size and event count, while updating the performance estimates. Discrimination was good for an imbalanced fatal endpoint, with AUROC 0.836 and AUPRC 0.245. The Brier score was 0.037, reflecting low average squared error for calibrated probabilities. At the selected threshold, the model detected 56 of 79 fatal events and preserved a high negative predictive value.

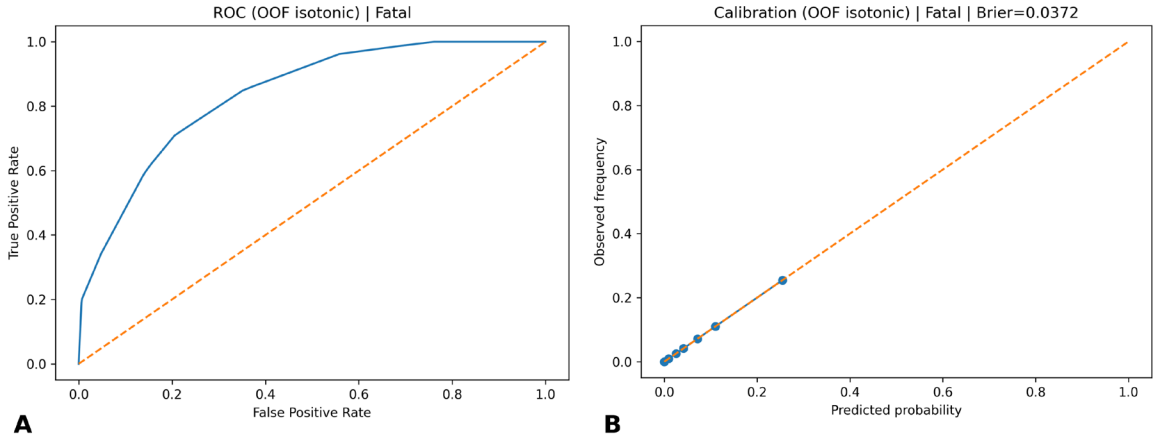

Figure 1. Out-of-fold performance of the final isotonic-calibrated model. A: receiver operating characteristic curve. B: final calibrated calibration plot with Brier score = 0.0372.

Feature-importance analysis showed that the highest-ranked variables combined traditional clinical measures and patient-reported or behavioral factors. The top 30 predictors included height, sleep efficiency, fasting blood sugar, age, weight, hemoglobin, lipids, waist circumference, social isolation, hospital length, socioeconomic status, sense of coherence, quality of life, sleep duration, type D personality, coping strategy and stress.

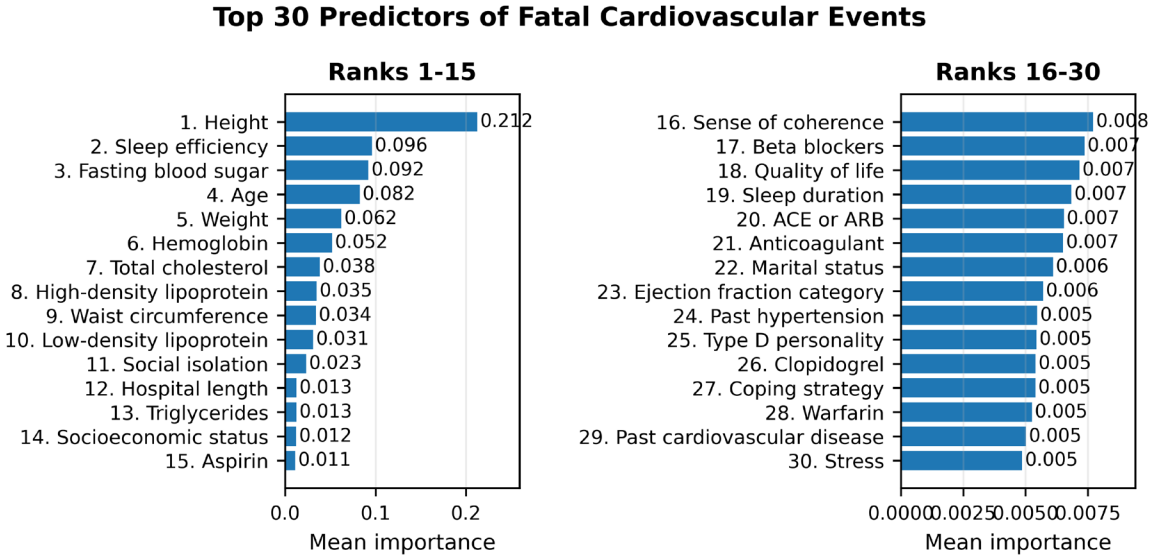

Figure 2. Top 30 predictors ranked by mean feature importance across repeated out-of-fold TabNet analyses. ACE: angiotensin-converting enzyme; ARB: angiotensin receptor blocker; OOF: out-of-fold.

4. Discussion

The model supports the premise that fatal event prediction after first acute ST-segment elevation myocardial infarction benefits from integrating clinical, metabolic, sleep and psychosocial data. Calibration was emphasized because predicted risks may influence follow-up intensity, rehabilitation prioritization and patient counseling [5,6].

Important limitations remain. Validation was internal and out-of-fold; external geographic or temporal validation is required before clinical use. The endpoint was rare, so positive predictive value was modest and sensitive to prevalence and threshold selection. Predicted high-risk status should therefore prompt clinical review rather than automatic action.

Future work should freeze preprocessing, model hyperparameters, calibration strategy and threshold-selection rule, then report deployment setting, intended use, fairness checks, missing-data handling and model availability in line with TRIPOD+AI [2].

Acknowledgments

We thank the CRAS-MI investigators, study nurses, psychologists, cardiologists and participants. This analysis used de-identified modeling outputs derived from the CRAS-MI cohort.

References

[1] H. Roohafza et al., "A cardiovascular risk assessment model according to behavioral, psychosocial and traditional factors in patients with ST-segment elevation myocardial infarction (CRAS-MI): review of literature and methodology of a multi-center cohort study," Curr. Probl. Cardiol., vol. 48, Art. no. 101158, 2023.

[2] G. S. Collins et al., "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, Art. no. e078378, 2024.

[3] S. O. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," Proc. AAAI Conf. Artif. Intell., vol. 35, pp. 6679-6687, 2021.

[4] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in Proc. 22nd Int. Conf. Mach. Learn., 2005, pp. 625-632.

[5] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg, "Calibration: the Achilles heel of predictive analytics," BMC Med., vol. 17, Art. no. 230, 2019.

[6] E. W. Steyerberg, Clinical Prediction Models, 2nd ed. Cham, Switzerland: Springer, 2019.

Address for correspondence:

Hamid Reza Marateb
Institute for Research and Innovation in Health, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain
hamid.reza.marateb@upc.edu