European Heart Journal - Digital Health (2026) 7 (Suppl 1)

e-Cardiology/Digital Health – Artificial Intelligence, Deep Learning

Development and validation of a parsimonious AI-based risk score for mortality in heart failure: a UK cohort study

N. Ahmed[1], S. Rao[1], K. Rahimi[1]

[1]University of Oxford, Oxford, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: Type of funding sources: Public grant(s) – EU funding. Main funding source(s): HORIZON EU

Background: Accurate risk stratification in heart failure (HF) populations is crucial. Existing simple scores, such as MAGGIC, show limited discrimination (C-index < 0.80) and require specialised tests like echocardiography. Conversely, while advanced Artificial Intelligence (AI) models offer greater accuracy, their operational complexity and data access requirements often hinder clinical adoption.
Purpose: To develop and validate a parsimonious, high-performing AI-based risk model for all-cause mortality in HF patients by leveraging insights from a complex AI model and feature engineering.
Methods: Using a cohort of 373,389 patients with HF (≥18 years; 1,153 English practices for derivation, 289 for validation) from the Clinical Practice Research Datalink (CPRD) Aurum dataset, we developed and validated an AI-based risk prediction model for all-cause mortality. A MAGGIC-based Cox proportional hazards model adapted for electronic health records (EHR) was established as a performance benchmark. An initial Multi-layer Perceptron (MLP) model with a survival framework was trained with MAGGIC variables. This MLP model was then enhanced by incorporating highly predictive comorbidity features—identified by an explainable Transformer model trained on longitudinal electronic health record (EHR) data—and feature distillation to derive a final, optimised 11-variable model (named MLP-Distilled). The model employed readily ascertainable variables including age, BMI, year of birth (as a proxy for birth cohort effects), and key comorbidities such as cancers.
Results: MLP-Distilled demonstrated significantly improved discriminatory performance (C-index: 0.801, 95% CI [0.796, 0.806]) compared to the benchmark MAGGIC-EHR model (0.735, [0.729, 0.742]), while maintaining excellent calibration. This parsimonious model uses fewer and more accessible variables than MAGGIC (11 vs. 13) and requires no specialised tests. The model was validated for secondary outcomes of cardiovascular events and rehospitalisation.
Conclusion: Utilising knowledge from a complex EHR-trained AI model combined with feature distillation yielded a parsimonious yet powerful risk score for HF mortality. This novel approach offers a pathway to clinically implementable tools that balance predictive accuracy with pragmatic utility, potentially improving routine HF risk stratification.

e-Cardiology/Digital Health – Artificial Intelligence, Deep Learning

Discrimination_Calibration.jpeg

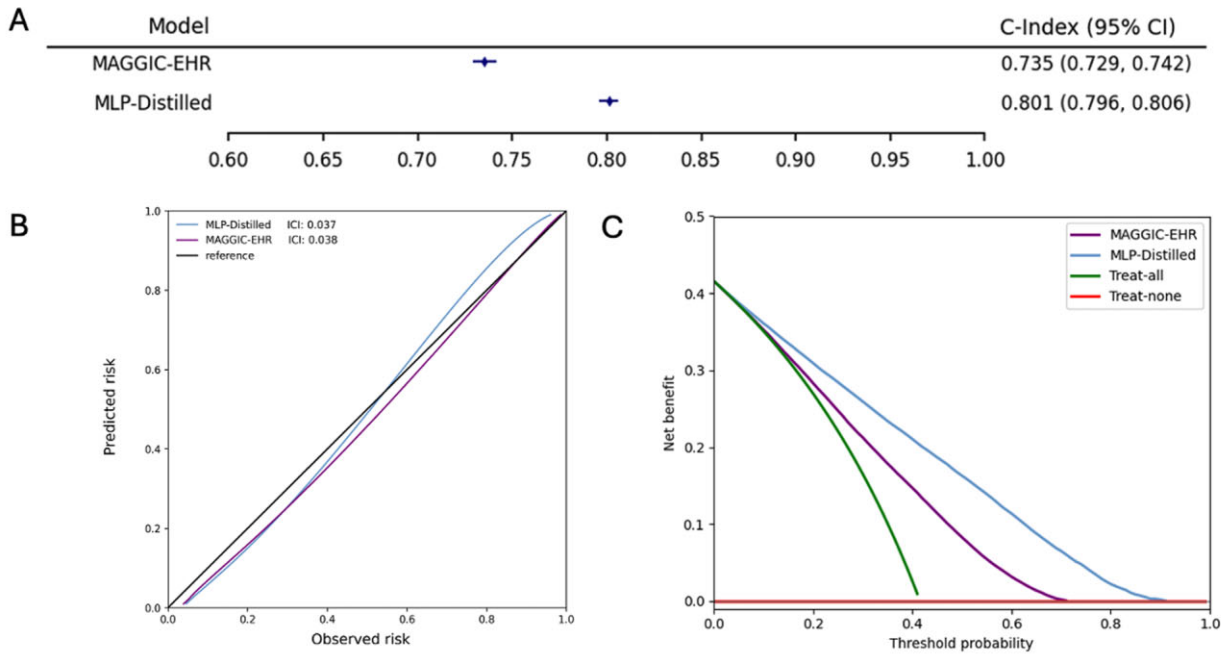

Figure 1: Discrimination, Calibration curves, and net benefit analysis for 12-month risk prediction of all-cause mortality in UK validation data

Discrimination (C-Index) with 95% CI (A), Calibration curves with integrated calibration indices (ICI) (B), and decision curve analysis (C) are presented for all-cause mortality at 12-month timepoint for MAGGIC-EHR and MLP-Distilled Models. For (B) ICI, lower is better with reference (black line) presenting optimal ICI of 0.0.

Impact_Analysis_50_Threshold.jpeg

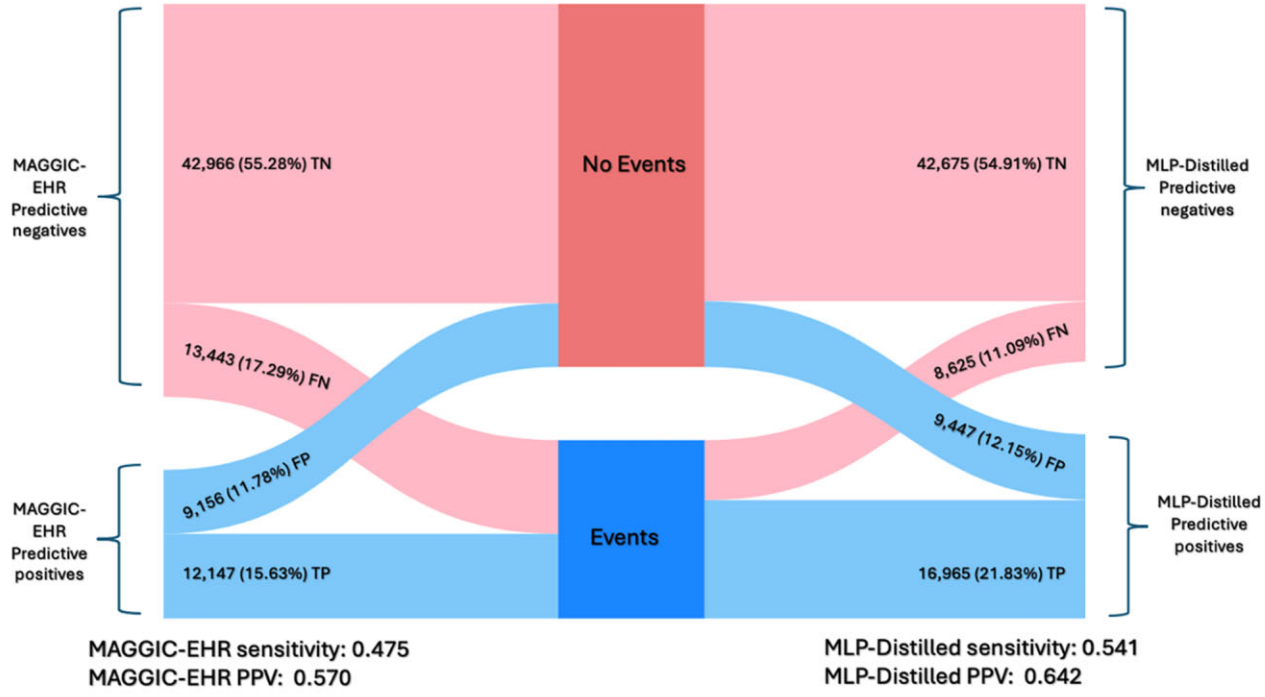

Figure 2. Impact analyses at the 50% decision threshold for 12- month all-cause mortality prediction on UK validation data.

Sankey diagrams compare predicted outcomes between the models, showing how patient classification compares to actual outcomes, "Events" (i.e., death) and "No events " categories at 50% threshold (denoted as dark blue and red respectively). TP: true positive; TN: true negative; FP: false positive; FN: false negatives.