

AI-Guided Risk Stratification for Aortic Stenosis using Large Language Models Enhanced with Guidelines

Authors: *Dorian Garin,¹ Stéphane Cook,¹ Charlie Ferry,¹ Wesley Bennar,¹ Mario Togni,¹ Pascal Meier,¹ Peter Wenaweser,¹ Serban Puricel,¹ Diego Arroyo¹

1. Department of Cardiology, University and Hospital Fribourg, Switzerland

*Correspondence to dorian.garin@icloud.com

Disclosure: The authors have declared no conflicts of interest.

Keywords: Aortic stenosis, European System for Cardiac Operative Risk Evaluation II (EuroSCORE II), large language model (LLM), risk stratification.

Citation: EMJ Int Cardiol. 2025;13[1]:43-44. <https://doi.org/10.33590/emjintcardiol/QRFR8070>

BACKGROUND

Traditional operative risk calculators, such as the European System for Cardiac Operative Risk Evaluation II (EuroSCORE II), may misclassify patients with severe aortic stenosis by insufficiently considering comorbidities and anatomical variables; particularly when guiding between transcatheter aortic valve implantation and surgical aortic valve replacement. The authors developed a guidelines-integrated large language model (LLM) that incorporates the 2021 European Society of Cardiology (ESC) guidelines for managing valvular heart disease, aiming to determine whether this approach could improve risk stratification compared to a purely EuroSCORE II-based strategy.¹

METHODS

The authors retrospectively analysed 231 patients with severe aortic stenosis who underwent formal Heart Team evaluation for low- versus high-operative risk between 1st January 2022–4th December 2024. For each patient, a clinical vignette was created to mimic a Heart Team presentation. A Forest-of-Thought prompting technique was then employed, simulating a multi-specialist discussion to yield either a 'low' or 'high' risk classification. The guidelines-integrated LLM (GPT-4o Version 2024-08-06; OpenAI,

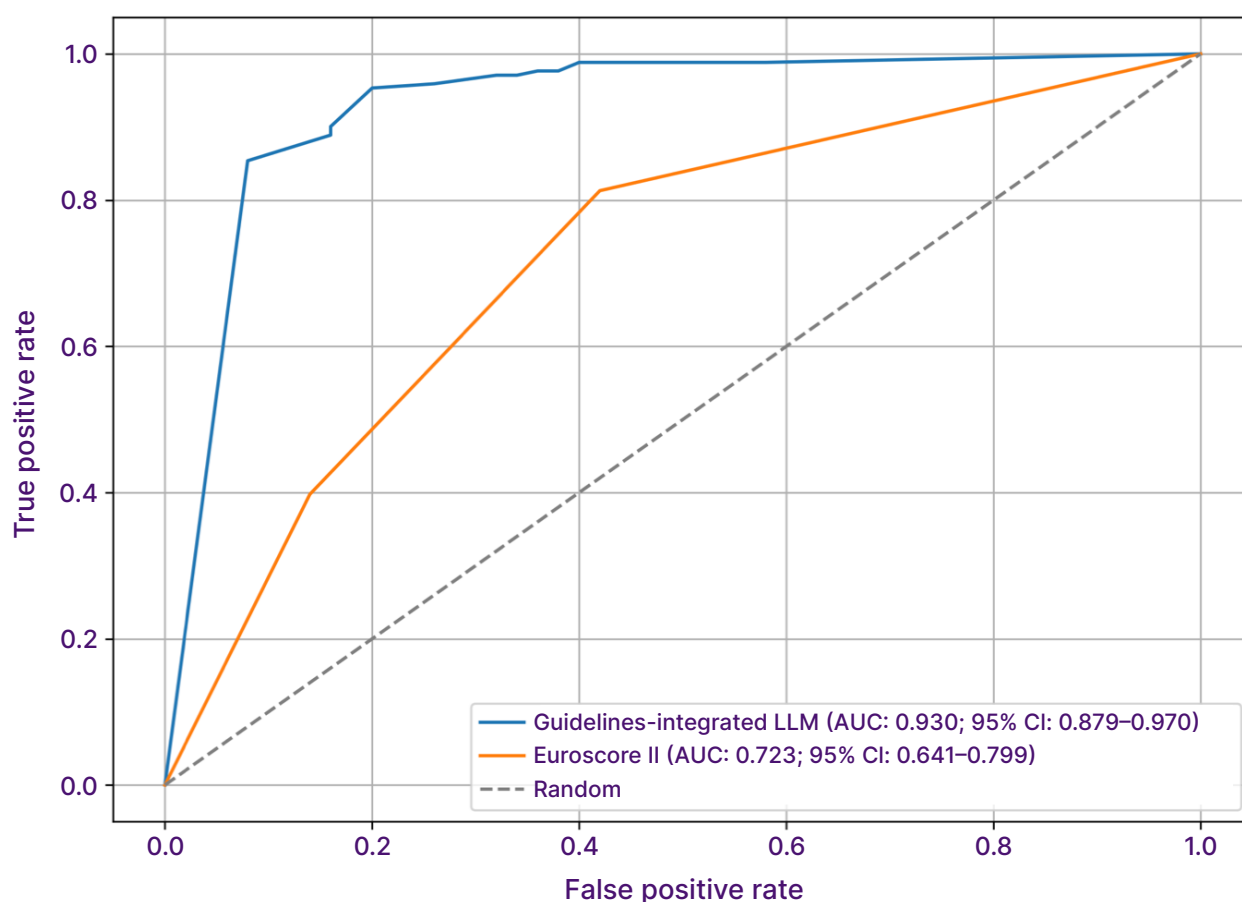
San Francisco, California, USA) received each vignette 40 times, and responses were consolidated using a self-consistency 'voting' procedure. The output from this guidelines-integrated LLM was compared to a EuroSCORE II-based approach, which defined low risk as EuroSCORE II <4% and age <75 years, and high risk as EuroSCORE II >8%. The primary endpoint was mean accuracy (proportion of correct low/high classifications versus the Heart Team's reference), while secondary endpoints included sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve. Logistic regression was used to assess the relative importance of EuroSCORE II versus other clinical variables. A subanalysis evaluated the guidelines-integrated LLM with versus without explicit EuroSCORE II input.

RESULTS

In identifying high-risk patients, the guidelines-integrated LLM achieved 90.05% accuracy (95% CI: 86.07–94.02), notably surpassing the EuroSCORE II-based method at 50.23% (95% CI: 43.58–56.87), with a mean difference of -39.82% (95% CI: -47.96 – -31.68; $p < 0.0001$). For low-risk stratification, it again outperformed the EuroSCORE II-based model (90.05% versus 85.97%; mean difference -4.07%; 95% CI: -7.93 – -0.21; $p = 0.039$). Comparing LLM variants with and without EuroSCORE II information showed a 7.69% mean accuracy gain (95% CI: 2.82–12.56; $p = 0.002$) when EuroSCORE II was omitted. Sensitivity, specificity, and ROC analyses were consistent with these findings (Figure 1).

Logistic regression indicated that excluding EuroSCORE II did not significantly alter the LLM's overall weighting of EuroSCORE II variables (Mann-Whitney $p = 0.34$). However, the lower performance with EuroSCORE II appeared linked to overemphasis on a limited subset of predictors, notably pulmonary artery systolic pressure (odds ratio [OR]: 1.70; $p = 0.007$), age (OR: 1.39; $p < 0.001$), and kidney disease (OR: 7.64; $p = 0.032$).

Figure 1: Aortic stenosis procedural risk stratification.



AUC: area under the curve; Euroscore II: European System for Cardiac Operative Risk Evaluation II; LLM: large language model.

In contrast, the guidelines-integrated LLM without EuroSCORE II maintained a balanced weighting across multiple variables, except for age (OR: 1.62; $p < 0.0001$) and male gender (OR: 1.11; $p = 0.038$).

CONCLUSION

A guidelines-integrated LLM strategy leveraging ESC guidelines provided superior high- and low-procedural risk stratification of patients with severe aortic stenosis, compared to a EuroSCORE II-based

approach. By encompassing a wider range of clinically relevant factors, this approach may enhance both clinical decision-making and individualised patient management, potentially better identifying candidates for transcatheter aortic valve implantation.

Reference

1. Garin D et al. AI-guided risk stratification for aortic stenosis using large language models enhanced with guidelines. Abstract A62829DG. EuroPCR, 2025 20-23 May, 2025.