

COMPARATIVE PERFORMANCE OF AI/ML VS CONVENTIONAL PROGNOSTIC MODELS IN MULTIPLE MYELOMA: A SYSTEMATIC REVIEW AND META-ANALYSIS

Basil Joy¹, Abhinav M¹, Sangeeth Pradeep², Esha Umair³

- 1. Government Medical College, Kozhikode, India
- 2. Jubilee Mission Medical College, Thrissur, India
- 3. Nishtar Medical University, Multan, Pakistan



INTRODUCTION

Background. Multiple myeloma (MM) is a heterogeneous hematological malignancy for which accurate prognostication remains challenging. Conventional staging systems, including the ISS and its revisions (R-ISS, R2-ISS), demonstrate limited discriminatory capacity.

Artificial intelligence (AI) and machine learning (ML) approaches have emerged as promising tools for survival and treatment response prediction, yet their comparative efficacy vs conventional methods has not been systematically quantified.

AIM

Objective. To systematically review and quantify, via meta-analysis, the comparative performance of AI/ML-based prognostic models versus conventional staging systems (ISS, R-ISS, R2-ISS) for predicting survival and treatment response in multiple myeloma.

METHOD

PRISMA systematic review and meta-analysis

- PubMed/MEDLINE, Embase, and Web of Science searched (2018–2025) for AI/ML prediction models in MM.
- Eligible studies compared AI models against ≥ 1 conventional prognostic method, with extractable outcomes.
- Primary outcomes: C-index for OS and PFS; secondary outcome: treatment-response AUC.
- Pooled via DerSimonian–Laird random-effects model; meta-regression identified heterogeneity sources.

CONCLUSIONS

AI/ML-based prediction models significantly outperform conventional prognostic methods in multiple myeloma across survival and treatment-response endpoints.

Genomic-data integration and deep-learning architectures yield the greatest gains. Prospective randomized validation is urgently required before clinical implementation.

RESULTS

18 studies, 19,847 patients (publication range 2021–2025) met inclusion criteria.

AI models significantly outperformed conventional staging for OS prediction: pooled C-index 0.789 (95% CI, 0.754–0.822) vs 0.697 for ISS-based models ($\Delta=+0.092$).

For PFS, pooled AI C-index was 0.741 (95% CI, 0.698–0.784) vs 0.653 for R-ISS ($\Delta=+0.088$). Treatment-response AUC was 0.831 (95% CI, 0.782–0.880).

Deep learning / transformer models performed best (C-index 0.812); genomic-data integration was the strongest predictor of accuracy ($\beta=+0.042$, $P=0.009$). Heterogeneity was substantial ($I^2=61\text{--}74\%$); GRADE certainty was moderate for survival and low-to-moderate for treatment response.

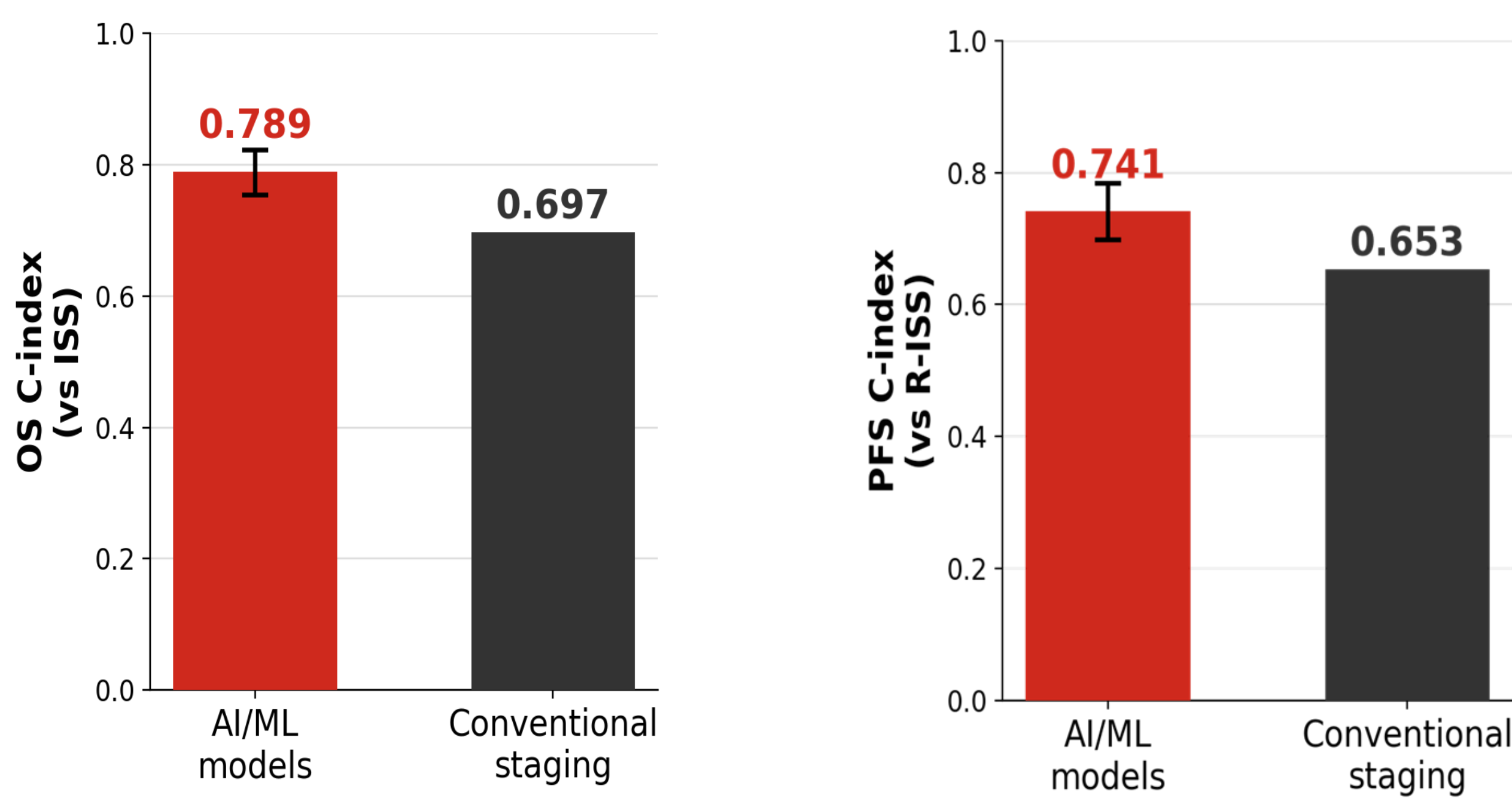


Figure 1. Pooled C-index for AI/ML vs conventional prognostic models (OS and PFS).

ACKNOWLEDGEMENTS

The authors declare no conflicts of interest and report no external funding for this study.

CONTACT INFORMATION

Basil Joy, Government Medical College, Kozhikode, India
Email: Docbasiljoy@gmail.com