

Prediction of heart failure based on multimodal data from MIMIC-IV [Abstract]

Anna Lutz, Fabio Hellmann, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Lutz, Anna, Fabio Hellmann, and Elisabeth André. 2026. "Prediction of heart failure based on multimodal data from MIMIC-IV [Abstract]." *Biomedical Engineering / Biomedizinische Technik* 71 (s1). <https://doi.org/10.1515/bmt-2026-1001>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0



Prediction of heart failure based on multimodal data from MIMIC-IV

Anna Lutz, Chair of Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany, annasofielutz@icloud.com

Fabio Hellmann, Chair of Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany, fabio.hellmann@informatik.uni-augsburg.de

Elisabeth André, Chair of Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany, andre@informatik.uni-augsburg.de

Introduction

Heart failure (HF) affects over 64 million people worldwide and remains a leading cause of cardiovascular mortality. Early identification of patients at risk is essential for timely treatment and to support hospital and primary care physicians. This study compares XGBoost and the Transformer-based model bidirectional Cross-Attention on multimodal clinical data to assess whether deep learning (DL) approaches outperform classical machine learning (ML) in early HF prediction.

Methods

We identified HF and non-HF patients from the MIMIC-IV database using ICD-9/10 codes, supplemented by clinical evidence from laboratory results, medications, radiology-, and discharge notes. Furthermore, we defined a 48-hour prediction window prior to the first clinical evidence of HF. Structured features were engineered from rolling-window statistics and clinical thresholds. Both XGBoost and Transformer models were trained on multimodal data and compared through an ablation study. Finally, we developed a clinical decision-support dashboard using a small set of demographic, laboratory, and medication features to deliver a 48-hour HF risk estimate, aiding clinician diagnosis.

Results

Multimodal models outperformed single-modality models across both architectures. The multimodal XGBoost model achieved the highest performance (F1 0.8773, PR-AUC 0.9437), while the multimodal Transformer achieved slightly lower performance (F1 0.8635, PR-AUC 0.9209). Structured data contributed most to XGBoost (PR-AUC of 0.9369), whereas clinical notes were better captured by the Transformer (PR-AUC of 0.8220).

Conclusion

Explainable dashboards further enhance transparency and usability by delivering quantitative 48-hour risk estimates from minimal features. This demonstrates, in this setting, that traditional ML can outperform DL models such as Transformers on tabular-dominated, multimodal clinical prediction tasks while preserving interpretability, underscoring decision-support systems' potential to aid timely diagnosis.