Multi-Modal Interpretable Graph for Competing Risk Prediction with Electronic Health Records

Munib Mesinovic (munib.mesinovic@jesus.ox.ac.uk), Peter Watkinson, Tingting Zhu

Department of Engineering Science, University of Oxford

Abstract

We present a novel multi-modal graph learning framework for competing risks survival analysis from electronic health records (EHRs). We introduce a unified, end-to-end model that learns modality-specific spatio-temporal graph representations for time-series, demographics, diagnostic histories, and radiographic text, and fuses them via hierarchical attention into a global patient graph. We further propose a composite training objective that combines survival likelihood, temporal ranking, and graph regularisation losses to improve risk discrimination, calibration, and structural consistency over time. Our model outperforms the latest baselines across five real-world EHR datasets, achieving up to 8% gains in cause-specific concordance, while offering fine-grained interpretability across temporal and modality dimensions.

Methods

Overview. We propose a dynamic, end-to-end graph learning model that integrates four heterogeneous EHR modalities—time-series vitals/labs, static demographics, diagnostic codes (ICD), and radiographic report embeddings—into a unified graph structure. Each patient is represented as a sequence of time-evolving graphs per modality, fused via a hierarchical attention mechanism into a global graph. The model captures intra- and inter-modality dependencies over time and supports missing-modality inference.

1. Dynamic Graph Construction. For each time window t, a graph $A_t^{(\mathrm{T})} \in \mathbb{R}^{f \times f}$ is computed from learnable embeddings:

$$A_t^{(\mathrm{T})} = \Theta_t^\top \cdot \Psi_t$$

We retain only the top-k strongest edges per node (excluding self-loops), creating a sparse adjacency matrix. Vertices are linked across adjacent time windows to preserve temporal continuity. This yields a dynamic sequence:

$$\mathcal{A}^{(\mathrm{T})} = \{A_1^{(\mathrm{T})}, \ldots, A_s^{(\mathrm{T})}\} \in \mathbb{R}^{f \times f \times s}$$

2. Multi-Modal Graph Encoding. Each modality is encoded as its own graph:

- ICD graphs $A^{(\mathrm{C})}$ are built using cosine similarity between code embeddings.
- Radiograph reports are embedded with Clinical-Longformer and connected with Gaussian similarity.
- High-dimensional graphs (ICD, Radiography) are reduced to 50×50 via GNN-based pooling:

$$\tilde{X}^{(\mathrm{C})}, \tilde{A}^{(\mathrm{C})} = f_\theta^{(\mathrm{C})}(X^{(\mathrm{C})}, A^{(\mathrm{C})})$$

Each modality has its own interpretability matrix $I^{(m)} \in \mathbb{R}^{d_m \times d_m}$, used for attribution.

3. Hierarchical Graph Fusion. We introduce learnable cross-modality attention matrices $W^{(m \to n)} \in \mathbb{R}^{d_m \times d_n}$, capturing directional influence across modality pairs. The final fused graph is a block matrix:

$$A^{\text{Fused}} = \begin{bmatrix} A^{(\mathrm{T})} & W^{(\mathrm{T}\to\mathrm{S})} & \cdots \\ \vdots & A^{(\mathrm{S})} & \\ & & \ddots \end{bmatrix}, \quad G^{\text{Final}} := A^{\text{Fused}} \circ \text{softmax}(I^{\text{Fused}})$$

Fused attention allows inter-modality reasoning and interpretability. Missing modalities are handled by dropping the corresponding blocks at test time.

4. Objective Function. The total training loss combines contrastive, structural, and likelihood terms:

$$\begin{aligned} \mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{focal}} + \epsilon\mathcal{L}_{\text{contrast}} + \lambda\mathcal{L}_{\text{reg}} + \mu\mathcal{L}_{\text{structure}} + \beta\mathcal{L}_{\text{VGAE}} \\ \mathcal{L}_{\text{focal}} &= -(1-\hat{y})^\gamma \log(\hat{y}) \\ \mathcal{L}_{\text{contrast}} &= -\log \frac{\exp(\text{Sim}_\theta(\mathbf{A}, \mathbf{A}^+))}{\exp(\text{Sim}_\theta(\mathbf{A}, \mathbf{A}^+)) + \sum_{\mathbf{A}^-} \exp(\text{Sim}_\theta(\mathbf{A}, \mathbf{A}^-))} \\ \mathcal{L}_{\text{reg}} &= \sum_{(i,j)\in\mathcal{E}} \|\mathbf{h}_i - \mathbf{h}_j\|^2 \\ \mathcal{L}_{\text{structure}} &= 1 - \frac{\sum A_{ij} \cdot A'_{ij}}{\|A\|_F \cdot \|A'\|_F} \end{aligned}$$

To model competing risks, we estimate the joint distribution of time and event type using a survival likelihood:

$$\mathcal{L}_{\text{NLL}} = -\sum_i \log\left[\text{event or censoring probability}\right]$$

plus a time-aware risk ranking loss:

$$\mathcal{L}_{\text{rank}} = \sum_{\epsilon=1}^{E} \mu \sum_{i \neq j} M_{ij}^{\epsilon} \cdot \eta(R_\epsilon^{(i)}, R_\epsilon^{(j)})$$

with smoothed pairwise loss $\eta(a,b) = \exp\left(-\frac{a-b}{\sigma}\right)$. The final CIF for patient X at time δ is:

$$\hat{F}_\epsilon(\delta \mid X) = \frac{\sum_{t_J < \delta \leq \Delta} \widehat{a}_{\epsilon,\delta}}{1 - \sum_{\epsilon \neq \varnothing} \sum_{n \leq t_J} \widehat{a}_{\epsilon,n}}$$

Performance Summary Across Tasks and Datasets

Our model achieves top performance across five real-world clinical datasets, covering both single-risk and competing-risk survival tasks. We report time-dependent concordance (C-index ↑), Integrated Brier Score (IBS ↓), and Integrated Binomial Log-Likelihood (IBLL ↓), reflecting ranking, calibration, and probabilistic accuracy.

In single-risk settings, our model outperforms classical (Cox PH), neural (DeepSurv, DeepHit), and dynamic (DySurv, Dynamic-DeepHit) methods, with largest gains in hospital admission (MC-MED, C-index: 0.880).

In competing-risk tasks, we maintain high cause-specific C-index across outcomes (e.g., ICU admission, death), demonstrating strong multi-outcome generalisation.

Outcome / Task	C-index	IBS	IBLL	Dataset
ICU Mortality	**0.861**	**0.139**	**-0.398**	MIMIC-IV
ICU Mortality	**0.809**	**0.183**	**-0.442**	eICU
Hospital Mortality	**0.768**	**0.189**	**-0.417**	PBC2
Survival (General)	**0.797**	**0.171**	**-0.426**	SUPPORT
Hospital Admission	**0.880**	**0.128**	**-0.459**	MC-MED
ICU Admission	**0.827**	–	–	MC-MED (CR)
ED Observation	**0.797**	–	–	MC-MED (CR)
Death	**0.790**	–	–	PBC2 (CR)
Liver Transplant	**0.766**	–	–	PBC2 (CR)

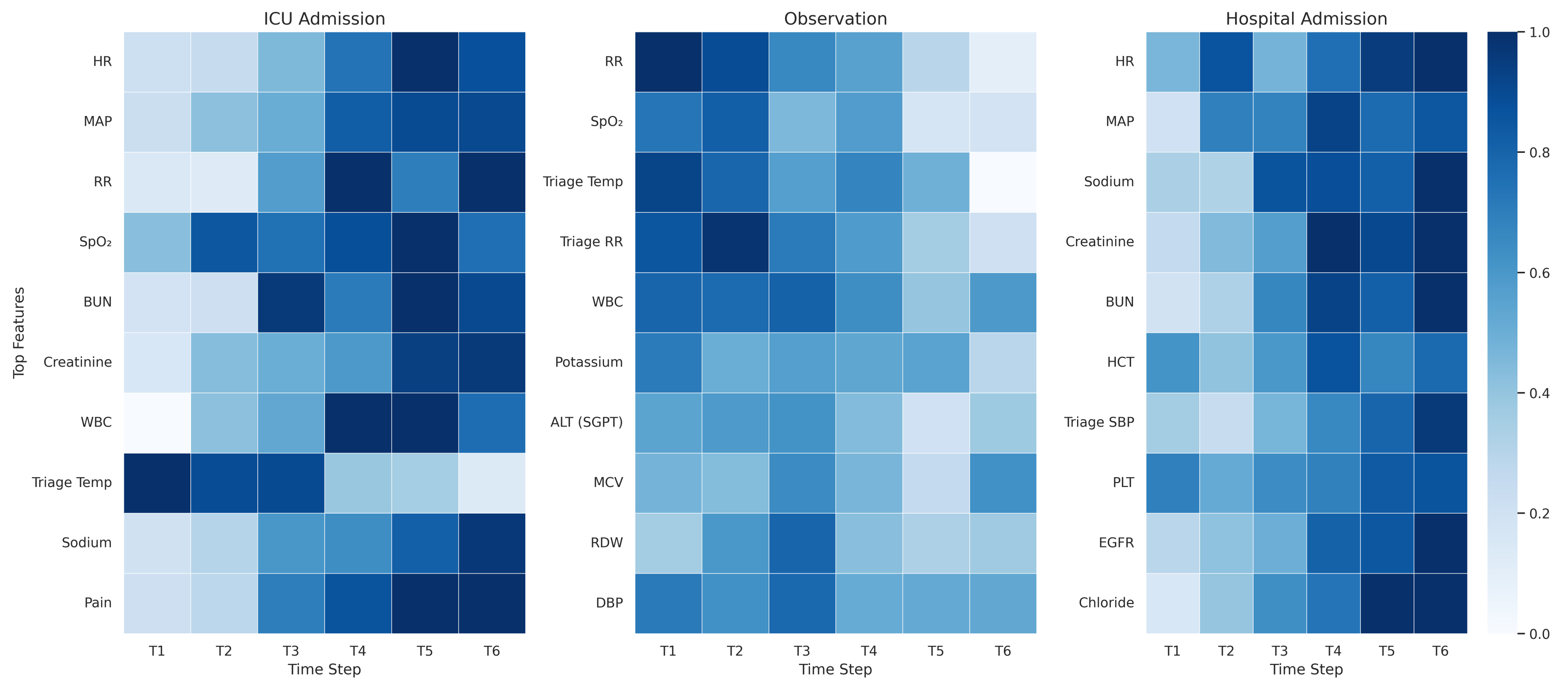
Ablation Study on MC-MED

We evaluate model variants to quantify the importance of each component. Removing ranking or structural loss reduces concordance. Excluding ICD or radiographic data harms performance on outcomes needing medical history. This confirms the value of our multi-modal attention-based design.

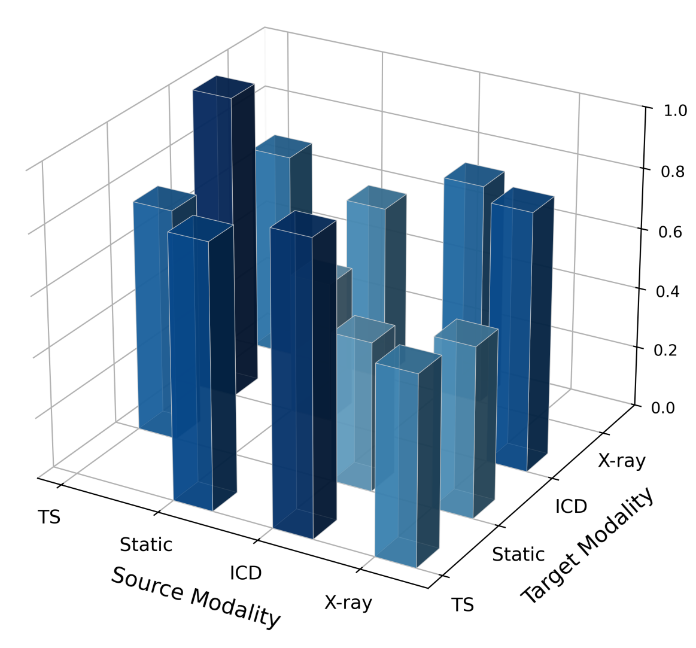
Variant	ICU Adm.	Hosp. Adm.	ED Obs.
w/o Ranking Loss	0.802	0.783	0.776
w/o Structural Loss	0.808	0.789	0.779
w/o ICD Codes	0.822	0.828	0.785
Ours (Full)	**0.827**	**0.880**	**0.797**

Interpretability

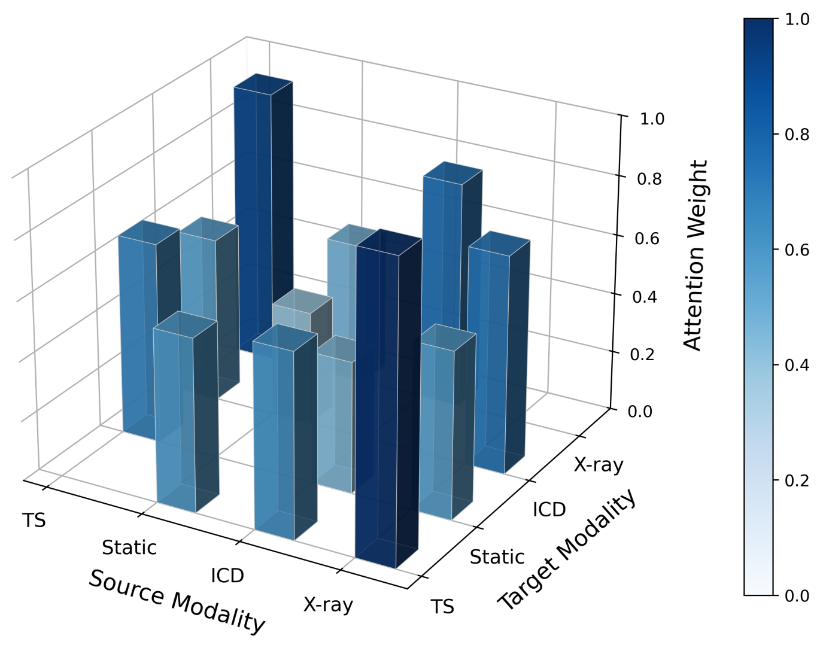
We visualise both temporal and modality-specific attention weights. Panel (a) shows how feature importance evolves in time for ICU prediction. Panels (b–c) show cross-modal contributions for ICU and hospital admission, revealing interpretable, outcome-specific attention patterns.



(a) Temporal attention on top ICU features (eICU).



(b) Cross-modality weights for ICU admission (MC-MED).



(c) Cross-modality weights for hospital admission (MC-MED).