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

Munib Mesinovic [1] **Peter Watkinson** [2] **Tingting Zhu** [1]

Abstract

Survival prediction from electronic health records (EHRs) is crucial for clinical decision-making but remains challenging due to data heterogeneity, irregular sampling, and the presence of competing risks. We propose a novel multi-modal graph neural network that dynamically constructs and integrates modality-specific graphs from time-series, demographics, diagnostic codes, and radiographic text. Our hierarchical attention mechanism fuses intra- and inter-modality interactions while providing interpretable, cause-specific risk predictions. Trained end-to-end with a combination of negative log-likelihood, ranking, and structural losses, our model significantly outperforms existing survival and graph-based baselines across five real-world EHR datasets. We further demonstrate improved calibration and interpretability, highlighting its potential for robust and transparent clinical risk stratification.

1. Introduction

Electronic health records (EHRs) offer rich but complex data for clinical risk prediction, characterised by irregular sampling, missingness, and heterogeneity across modalities. Time-to-event models enable prediction not only of whether an event occurs, but also when, making them critical for applications such as ICU triage or transplant planning (Lee et al., 2019; Qiu et al., 2025). However, classical models like Cox Proportional Hazards assume linear effects and fixed covariates, limiting their utility in dynamic, high-dimensional EHR settings (Martinussen, 2022).

Recent deep learning methods have improved performance in survival prediction, yet most focus on single-risk settings or unimodal data (Katzman et al., 2018; Lee et al., 2018; Mesinovic et al., 2024). Models like Dynamic-DeepHit and DySurv handle temporal dependencies but lack interpretability and multi-modal integration. Graph neural networks (GNNs) offer an expressive framework to model temporal and structural relationships in EHRs (Boll et al., 2024; Xu et al., 2021), though prior works rely on S or hand-crafted graphs and remain limited to single modalities (Liu et al., 2024; Mesinovic et al., 2025).

We introduce a novel multi-modal spatio-temporal graph neural network for survival analysis with competing risks. Our model learns dynamic graphs for each modality, time-series, demographics, ICD codes, and radiographic text, and fuses them using hierarchical graph attention. This enables interpretable modelling of intra- and inter-modality dependencies without any predefined graph structures.

Our contributions are as follows:

1. We introduce the first unified framework for dynamic cross-modal graph learning in healthcare, where modality-specific spatio-temporal graphs are constructed and F through hierarchical attention without relying on predefined structures.

2. We propose a novel hierarchical interpretability mechanism that enables fine-grained attribution across features, time steps, and modalities, under competing events.

3. We demonstrate that our approach outperforms state-of-the-art survival models and graph baselines on multiple real-world EHR datasets, achieving calibrated and interpretable competing risk prediction across ICU, emergency, and transplant care.

2. Methods

Data. We evaluate our model on two datasets adapted for competing risk prediction: PBC2 and MC-MED. The PBC2 dataset contains monthly biomarker measurements and S clinical variables from a randomised trial on primary biliary cirrhosis, with competing risks of death and liver transplantation. The MC-MED dataset comprises 118,385 emergency

[1]Department of Engineering Science, University of Oxford, Oxford, UK [2]Nuffield Department of Clinical Neuroscience, Oxford, UK. Correspondence to: Munib Mesinovic <munib.mesinovic@eng.ox.ac.uk>.

Proceedings of the ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences, Vancouver, Canada. 2025.

department visits from 70,545 patients at Stanford Health Care, incorporating multimodal data such as vital signs, lab tests, demographics, ICD9/10 histories, and free-text radiography reports, with competing outcomes of hospital admission, observation, and ICU admission. We retain the top 500 ICD codes and use six time steps (windows) per patient to enable graph construction. Missing values are imputed via forward and backward filling, and sequences are padded using the latest available measurements. The pre-processing pipeline for MIMIC-IV was based on previously published workflows, and eICU was based in part on work done by (Mesinovic et al., 2024; Rocheteau et al., 2021). We used the imputation as suggested by the pipeline.

For the time-series variables, we use forward filling as clinicians in practice would only consider the last recorded measurement. If the first set of measurements is missing for some time-varying features, instead of dropping those features or patients, we backward-fill from the closest measurement in the future. The time-series features were resampled to 1-hour intervals. For the ICU datasets, we considered only observations collected up to 24 hours before the registered outcome. For MC-MED, since it is an ED dataset, the entirety of the patient cohort is within 24 hours of stay within the emergency department, and we include all of this information before the event time itself. For PBC2, we resampled the data into a monthly timescale. Patient admissions were randomly split into train, validation and test sets (8:1:1). Details of the features included can be found in Supplementary Tables 1, 2, and 3.

For eICU, MIMIC-IV, and MC-MED, the data contains de-identified patient electronic health records data, which can only be obtained after the ethical review of the proposed analysis on the PhysioNet page. Some certification of training modules is also required for access. We have cited the sources for the datasets in the text accordingly under Data. Consent for data use has been obtained by the providers, de-identification and licensing are in line with HIPAA requirements and compatible with the research conducted, which has passed ethical review and certification for data access.

We address the task of survival prediction under competing risks, where patients may experience one of several mutually exclusive event types (e.g., ICU admission, death), or be censored. Each patient i is associated with an outcome $Y_i = (\epsilon_i, t_i)$, where $\epsilon_i \in \mathcal{E} = \{1, 2, \ldots, E\} \cup \{\varnothing\}$ denotes the event type (or censoring), and $t_i \in \mathbb{R}_{\geq 0}$ is the observed time to event or censoring. The aim is to estimate cause-specific cumulative incidence functions (CIFs),

$$\hat{F}_\epsilon(t \mid X) = \mathbb{P}(T \leq t, \epsilon = \epsilon \mid X),$$

which quantify the risk over time of experiencing event ϵ, given patient covariates X, while accounting for the presence of competing risks and right-censoring. This is particularly relevant in clinical settings where understanding differential risk across multiple outcomes is critical for prognosis and intervention.

Each patient i is represented by a multi-modal input:

$$X_i = \{X_i^{(\text{TS})}, X_i^{(\text{S})}, X_i^{(\text{XR})}, X_i^{(\text{ICD})}\}$$

- $X_i^{(\text{TS})} \in \mathbb{R}^{d \times l}$: time-series of clinical measurements (e.g., vitals, labs),
- $X_i^{(\text{S})} \in \mathbb{R}^{d_s}$: S demographic variables (e.g., age, sex),
- $X_i^{(\text{XR})} \in \mathbb{R}^{d_r}$: text embeddings of radiographic reports,
- $X_i^{(\text{ICD})} \in \mathbb{R}^{d_c}$: diagnostic code histories (ICD9/10).

Together, these modalities provide complementary and longitudinal insight into the patient's state.

To capture evolving dependencies in $X_i^{(\text{TS})}$, we divide the sequence into s windows and construct one graph per window using learnable node embeddings:

$$A_t = \Theta_t^\top \Psi_t \in \mathbb{R}^{d \times d}, \quad t = 1, \ldots, s.$$

We apply top-k sparsification and add directed temporal edges from nodes in window $t-1$ to corresponding nodes in window t. To control memory, redundant nodes are aggregated. This yields a temporal graph set $\{A_t\}_{t=1}^s \in \mathbb{R}^{d \times d \times s}$.

Each modality is modelled as a separate graph:

1. S features: fully connected graph $A^{(\text{S})} \in \mathbb{R}^{d_s \times d_s}$.
2. ICD codes: Top 500 codes are embedded via co-occurrence. Cosine similarity forms $A^{(\text{ICD})} \in \mathbb{R}^{500 \times 500}$, followed by learnable pooling:

$$\tilde{A}^{(\text{ICD})} = f_\theta^{\text{ICD}}(A^{(\text{ICD})}) \in \mathbb{R}^{d' \times d'}.$$

3. Radiographic reports: We encode reports using Clinical-Longformer (Li et al., 2022):

$$x_t = \Phi(r_t), \quad A_{tt'} = \exp\left(-\frac{\|x_t - x_{t'}\|^2}{\tau}\right)$$

for all reports r_t in the patient's history. This yields patient-specific temporal graphs.

To integrate across modalities, we learn cross-modality attention matrices:

$$W^{(m \to n)} \in \mathbb{R}^{d_m \times d_n}, \quad \forall m, n \in \{\text{TS, S, XR, ICD}\}$$

Each is trained jointly with intra-modality graphs to form the fused graph:

$$A^{\text{F}} = \begin{bmatrix} A^{(\text{TS})} & W^{(\text{TS} \to \text{S})} & \cdots \\ W^{(\text{S} \to \text{TS})} & A^{(\text{S})} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Each graph and cross-attention map is paired with a trainable interpretability matrix $I^{(m)}$ or $I^{(m\to n)}$, updated via gradient attribution.

The fused graph is passed through stacked GIN layers with temporal pooling and convolutional clustering. The resulting graph is flattened and passed through E MLP branches, one per event type. Each outputs $\widehat{a}_{\epsilon,t}$, a probability estimate of event ϵ at time t. CIFs are estimated via:

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

We train the model using a combined loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{reg}} + \alpha\mathcal{L}_{\text{NLL}} + \beta\mathcal{L}_{\text{rank}} + \gamma\mathcal{L}_{\text{struct}}.$$

where

- Negative Log-Likelihood (NLL) models the joint likelihood of event type and time for uncensored, and censoring distribution for censored patients.
- Ranking loss enforces consistent risk ordering across patients with comparable clinical histories.
- Structural loss encourages smooth graph evolution across time:

$$\mathcal{L}_{\text{struct}} = \mu\left(1 - \frac{\sum_{i,j} A_{ij}\cdot A'_{ij}}{\|A\|_F \cdot \|A'\|_F}\right)$$

- Regularisation smoothens the penalty across node embeddings:

$$\mathcal{L}_{\text{reg}} = \lambda \sum_{(i,j)\in E} \|h_i - h_j\|^2.$$

This combination enables accurate, robust, and interpretable cause-specific risk prediction over time from complex multimodal EHR data.

3. Results

We evaluate our model against two groups of baselines for both single-risk and competing-risk survival analysis tasks. A complete set of comparisons is in Table 1, which shows the results in the same settings with results averaged over 5 seeds. Our model consistently outperforms all baselines across different datasets and care environments. The largest improvements are observed in concordance metrics, reflecting our model's strong temporal discrimination capabilities. The combination of ranking loss and structural regularisation enables our model to predict both short- and long-term survival outcomes robustly across diverse clinical settings. Moreover, the superior IBS and IBLL scores indicate better generalisation and calibrated event probability estimates, avoiding the inflated concordance behaviour seen in prior ranking-based models.

3.1. Competing Risk Survival Performance

We evaluate our model on five real-world EHR datasets across both single-risk and competing-risk scenarios. Table 2 reports the cause-specific time-dependent concordance index ($C^{\text{td}}_{\text{ind},\epsilon}$) on two challenging competing risk datasets: PBC2 and MC-MED. Our model consistently outperforms established baselines, including Dynamic-DeepHit (Lee et al., 2019), MedGNN (Fan et al., 2025), and MM-STGNN (Tang et al., 2023). On the PBC2 dataset, we observe improvements of up to 2.5 percentage points over MM-STGNN for predicting death, and a similar gain for liver transplant prediction. On the MC-MED dataset, our model achieves especially strong gains in hospital admission prediction (0.880 vs. 0.798 for MM-STGNN), where cross-modal interactions between longitudinal vitals, diagnostic codes, and radiography reports are critical. These results underscore the value of learning dynamic graphs and hierarchical attention jointly over diverse modalities.

3.2. Interpretability and Temporal Dynamics

To interpret the model's temporal reasoning, we visualise the learned attention weights from the time-series interpretability matrix I^{TS}. As shown in Figure 1, the top 10 predictive features for each competing outcome display evolving attention patterns over six time steps. These heatmaps illustrate that the model learns not only S importance but also dynamically shifts its focus as patient trajectories evolve. This provides clinicians with temporal insight into which physiological variables are most salient for a given risk, enabling time-sensitive and interpretable risk stratification.

3.3. Cross-Modality Contributions

In addition to feature-level attention, we examine how the model attends across data modalities. Figure 2 shows 4×4 modality-level interpretability matrices extracted from the F graph for each competing outcome in MC-MED. Time-series and radiographic report embeddings contribute heavily across all tasks, reflecting their rich temporal and semantic signal. Notably, for ICU admission, we observe strong cross-modality attention from ICD code history to time-series and radiography, suggesting that diagnostic context informs escalation decisions. In contrast, ED observation relies on more diffuse inter-modality influence, consistent with its heterogeneous clinical profiles.

3.4. Ablation Analysis

To quantify the impact of individual model components and input modalities, we perform a comprehensive ablation study on the MC-MED dataset (Table 3). In the architectural ablations, removing either the ranking loss or structural regularisation term leads to consistent declines in concor-

Table 1. Single-risk evaluation results across five datasets using three common time-to-event metrics. Results are averaged over 5 seeds. For Concordance Index (↑), higher is better; for IBS and IBLL (↓), lower is better. Best values per dataset are bolded.

Concordance Index (↑)

Model	MIMIC-IV	eICU	PBC2	MC-MED	SUPPORT
Cox PH	0.711 ± 0.013	0.642 ± 0.016	0.676 ± 0.012	0.589 ± 0.018	0.664 ± 0.014
DeepSurv	0.752 ± 0.011	0.684 ± 0.014	0.702 ± 0.015	0.654 ± 0.017	0.698 ± 0.013
DeepHit	0.778 ± 0.010	0.723 ± 0.012	0.706 ± 0.011	0.739 ± 0.014	0.719 ± 0.012
Dynamic-DeepHit	0.807 ± 0.009	0.758 ± 0.010	0.716 ± 0.011	0.786 ± 0.010	0.749 ± 0.011
DySurv (Mesinovic et al., 2024)	0.832 ± 0.008	0.782 ± 0.009	0.736 ± 0.009	0.809 ± 0.008	0.779 ± 0.009
GCN	0.649 ± 0.014	0.598 ± 0.015	0.631 ± 0.012	0.579 ± 0.015	0.622 ± 0.014
GAT	0.674 ± 0.012	0.621 ± 0.013	0.659 ± 0.011	0.602 ± 0.014	0.647 ± 0.012
GraphSAGE (Liu et al., 2023)	0.697 ± 0.013	0.643 ± 0.012	0.681 ± 0.010	0.634 ± 0.014	0.669 ± 0.013
TodyNet (Liu et al., 2024)	0.738 ± 0.011	0.685 ± 0.010	0.706 ± 0.011	0.719 ± 0.010	0.708 ± 0.011
DynaGraph (Mesinovic et al., 2025)	0.803 ± 0.008	0.744 ± 0.009	0.726 ± 0.010	**0.832 ± 0.008**	0.761 ± 0.010
MedGNN	0.759 ± 0.009	0.708 ± 0.011	0.698 ± 0.010	0.776 ± 0.009	0.732 ± 0.010
MM-STGNN	0.767 ± 0.010	0.725 ± 0.010	0.693 ± 0.012	0.762 ± 0.011	0.731 ± 0.011
Ours	**0.861 ± 0.007**	**0.809 ± 0.008**	**0.768 ± 0.008**	**0.832 ± 0.007**	**0.797 ± 0.008**

Integrated Brier Score (IBS) (↓)

Model	MIMIC-IV	eICU	PBC2	MC-MED	SUPPORT
Cox PH	0.251 ± 0.012	0.304 ± 0.015	0.281 ± 0.014	0.332 ± 0.016	0.270 ± 0.012
DeepSurv	0.222 ± 0.011	0.261 ± 0.013	0.246 ± 0.013	0.299 ± 0.015	0.238 ± 0.012
DeepHit	0.209 ± 0.010	0.248 ± 0.012	0.257 ± 0.012	0.278 ± 0.013	0.247 ± 0.011
Dynamic-DeepHit	0.186 ± 0.009	0.219 ± 0.011	0.237 ± 0.011	0.199 ± 0.012	0.227 ± 0.010
DySurv	0.171 ± 0.008	0.209 ± 0.010	0.218 ± 0.009	0.177 ± 0.010	0.210 ± 0.009
GCN	0.321 ± 0.014	0.357 ± 0.016	0.336 ± 0.014	0.374 ± 0.015	0.351 ± 0.014
GAT	0.301 ± 0.012	0.341 ± 0.014	0.318 ± 0.013	0.351 ± 0.014	0.327 ± 0.013
GraphSAGE	0.282 ± 0.013	0.324 ± 0.012	0.309 ± 0.012	0.336 ± 0.014	0.298 ± 0.012
TodyNet	0.239 ± 0.011	0.279 ± 0.010	0.274 ± 0.011	0.257 ± 0.012	0.285 ± 0.011
DynaGraph	0.198 ± 0.009	0.234 ± 0.010	0.246 ± 0.011	0.216 ± 0.011	0.239 ± 0.010
MedGNN	0.186 ± 0.009	0.215 ± 0.010	0.231 ± 0.010	0.209 ± 0.010	0.227 ± 0.009
MM-STGNN	0.202 ± 0.010	0.242 ± 0.011	0.259 ± 0.011	0.212 ± 0.011	0.252 ± 0.010
Ours	**0.139 ± 0.006**	**0.183 ± 0.007**	**0.189 ± 0.007**	**0.128 ± 0.007**	**0.171 ± 0.006**

Integrated Binomial Log-Likelihood (IBLL) (↓)

Model	MIMIC-IV	eICU	PBC2	MC-MED	SUPPORT
Cox PH	-0.223 ± 0.011	-0.257 ± 0.012	-0.243 ± 0.011	-0.284 ± 0.012	-0.249 ± 0.011
DeepSurv	-0.258 ± 0.010	-0.299 ± 0.011	-0.281 ± 0.011	-0.317 ± 0.011	-0.288 ± 0.010
DeepHit	-0.278 ± 0.009	-0.322 ± 0.010	-0.304 ± 0.010	-0.337 ± 0.011	-0.308 ± 0.010
Dynamic-DeepHit	-0.318 ± 0.008	-0.369 ± 0.009	-0.351 ± 0.009	-0.381 ± 0.010	-0.356 ± 0.009
DySurv	-0.339 ± 0.007	-0.391 ± 0.009	-0.367 ± 0.008	-0.413 ± 0.009	-0.379 ± 0.008
GCN	-0.184 ± 0.012	-0.212 ± 0.013	-0.194 ± 0.012	-0.228 ± 0.013	-0.202 ± 0.012
GAT	-0.201 ± 0.011	-0.238 ± 0.012	-0.219 ± 0.011	-0.251 ± 0.012	-0.226 ± 0.011
GraphSAGE	-0.215 ± 0.011	-0.248 ± 0.011	-0.234 ± 0.010	-0.271 ± 0.011	-0.238 ± 0.010
TodyNet	-0.292 ± 0.010	-0.335 ± 0.010	-0.311 ± 0.010	-0.353 ± 0.010	-0.319 ± 0.010
DynaGraph	-0.298 ± 0.009	-0.342 ± 0.010	-0.318 ± 0.010	-0.362 ± 0.010	-0.329 ± 0.010
MedGNN	-0.308 ± 0.009	-0.362 ± 0.010	-0.337 ± 0.009	-0.373 ± 0.009	-0.347 ± 0.009
MM-STGNN	-0.301 ± 0.010	-0.354 ± 0.010	-0.328 ± 0.010	-0.361 ± 0.010	-0.338 ± 0.010
Ours	**-0.398 ± 0.006**	**-0.442 ± 0.007**	**-0.417 ± 0.007**	**-0.459 ± 0.007**	**-0.426 ± 0.007**

dance, particularly for hospital admission, highlighting the effectiveness of multi-objective training. Replacing the GIN backbone with a standard GCN also results in performance degradation across all outcomes. In the modality addition analysis, we observe the largest performance gain when time-series features are added to S inputs, with further improvements from the inclusion of ICD codes and radiography reports. These results confirm the complementary value of each modality and highlight the importance of structured multi-modal fusion. The same trends hold on MIMIC-IV and eICU (see Supplement), supporting the robustness of our design across diverse clinical settings.

4. Discussion

Our results demonstrate that dynamic, cross-modal graph learning significantly enhances survival prediction in clinical settings involving heterogeneous and irregular EHR data. Unlike prior models that rely on S or early-fusion strategies, our approach constructs modality-specific spatio-temporal graphs and fuses them via hierarchical attention, enabling fine-grained, interpretable risk estimation across time and modalities.

The model consistently outperforms strong baselines on five datasets, particularly in high-stakes settings like emergency and ICU care. Its interpretability matrices reveal clinically meaningful patterns, such as rising influence of renal markers over time or ICD-derived context contributing to ICU escalation. These findings support real-world deployment in triage, monitoring, or discharge planning workflows.

Limitations include the absence of treatment variables and external validation on non-US datasets. Future work will extend our framework to incorporate intervention effects and explore subgroup-specific attention patterns to audit for potential biases.

Table 2. Cause-specific concordance index ($C^{\text{td}}_{\text{ind},\epsilon}$) on the PBC2 and MC-MED datasets for competing risks. Our model consistently outperforms all baselines. Higher is better.

Model	PBC2 (Death)	PBC2 (Transplant)	ICU Adm.	Hosp. Adm.	ED Obs.
MedGNN	0.758	0.740	0.812	0.785	0.773
Dynamic-DeepHit	0.765	0.743	0.816	0.791	0.778
MM-STGNN	0.770	0.749	0.821	0.798	0.781
Ours	**0.790**	**0.766**	**0.827**	**0.880**	**0.797**

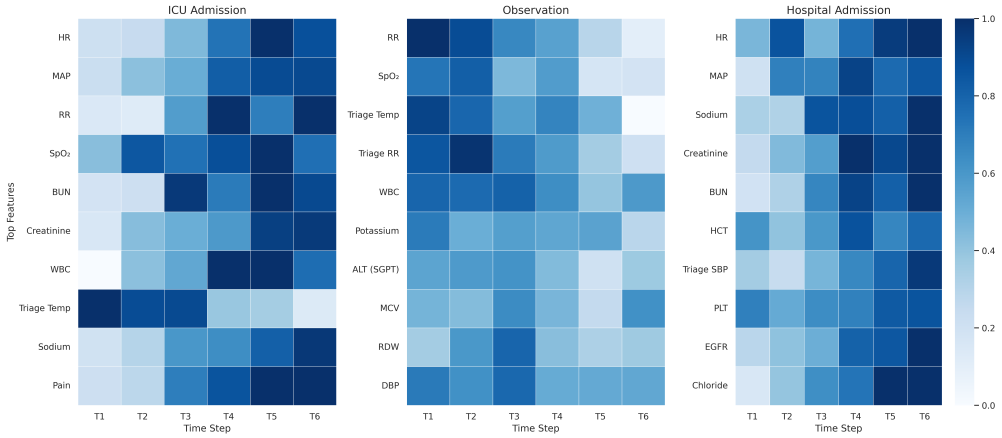

Figure 1. Temporal heatmap of the top 10 time-series features in MC-MED for ICU admission, ED observation, and hospital admission. Attention weights from the interpretability matrix I^{TS} highlight the model's temporal reasoning.

Impact Statement

The model supports clinical decision-making through transparent, time-aware predictions across multiple modalities. The ability to attribute predictions to specific features and modalities may improve trust and adoption in clinical settings. Our study uses de-identified, ethically sourced datasets.

References

Boll, H. O., Amirahmadi, A., Ghazani, M. M., de Morais, W. O., de Freitas, E. P., Soliman, A., Etminani, F., Byttner, S., and Recamonde-Mendoza, M. Graph neural networks for clinical risk prediction based on electronic health records: A survey. *J. Biomed. Informatics*, 151:104616, 2024.

Fan, W., Fei, J., Guo, D., Yi, K., Song, X., Xiang, H., Ye, H., and Li, M. Towards multi-resolution spatiotemporal graph learning for medical time series classification. In *Proceedings of the ACM on Web Conference 2025*, pp. 5054–5064, 2025.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18 (1):1–12, 2018.

Lee, C., Zame, W., Yoon, J., and Van Der Schaar, M. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Lee, C., Yoon, J., and Van Der Schaar, M. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, 67(1):122–133, 2019.

Li, Y., Wehbe, R. M., Ahmad, F. S., Wang, H., and Luo, Y. Clinical-longformer and clinical-bigbird: Transformers for long clinical sequences. *arXiv preprint arXiv:2201.11838*, 2022.

Liu, H., Yang, D., Liu, X., Chen, X., Liang, Z., Wang, H., Cui, Y., and Gu, J. Todynet: temporal dynamic graph neural network for multivariate time series classification. *Information Sciences*, 677:120914, 2024.

Liu, T., Jiang, A., Zhou, J., Li, M., and Kwan, H. K. Graphsage-based dynamic spatial–temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 24(10): 11210–11224, 2023.

Martinussen, T. Causality and the cox regression model. *Annual Review of Statistics and Its Application*, 9(1):249–259, 2022.

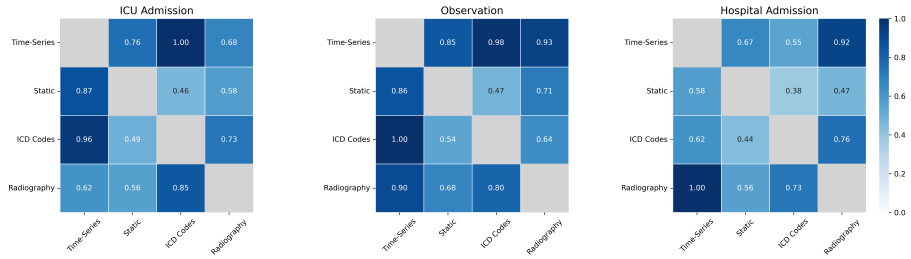

Figure 2. Modality-level attention heatmaps for ICU admission, ED observation, and hospital admission. Diagonal entries represent intra-modality contributions; off-diagonal entries show cross-modality interactions.

Table 3. Ablation study on the MC-MED dataset. We report the cause-specific time-dependent concordance index ($C^{\text{td}}_{\text{ind},\epsilon}$) for three competing outcomes. Results are averaged over 5 seeds. Higher is better.

Model Variant / Modalities Included	**ICU Admission**	**Hospital Admission**	**ED Observation**
Architectural and Loss Ablations			
w/o Ranking Loss	0.802 ± 0.012	0.783 ± 0.015	0.776 ± 0.017
w/o Structural Loss	0.808 ± 0.010	0.789 ± 0.013	0.779 ± 0.015
w/o History (ICD9/10)	0.822 ± 0.009	0.828 ± 0.011	0.785 ± 0.012
w/o Radiography Embeddings	0.820 ± 0.009	0.820 ± 0.011	0.782 ± 0.012
GCN backbone	0.811 ± 0.008	0.813 ± 0.011	0.782 ± 0.013
Ours (Full Model)	$\mathbf{0.827 \pm 0.007}$	$\mathbf{0.880 \pm 0.010}$	$\mathbf{0.797 \pm 0.012}$
Modality Addition Ablation (mean $\pm$ std)			
S only	0.792 ± 0.011	0.769 ± 0.013	0.766 ± 0.014
S + Time-Series	0.815 ± 0.010	0.790 ± 0.012	0.776 ± 0.013
S + ICD Codes	0.798 ± 0.010	0.790 ± 0.012	0.772 ± 0.013
S + Radiography	0.796 ± 0.011	0.785 ± 0.013	0.770 ± 0.014
S + Time-Series + ICD Codes	0.820 ± 0.009	0.820 ± 0.011	0.782 ± 0.012
S + Time-Series + Radiography	0.822 ± 0.009	0.828 ± 0.011	0.785 ± 0.012
S + ICD Codes + Radiography	0.809 ± 0.010	0.829 ± 0.012	0.781 ± 0.013
All Modalities	$\mathbf{0.827 \pm 0.007}$	$\mathbf{0.880 \pm 0.010}$	$\mathbf{0.797 \pm 0.012}$

Mesinovic, M., Watkinson, P., and Zhu, T. Dysurv: dynamic deep learning model for survival analysis with conditional variational inference. *Journal of the American Medical Informatics Association*, pp. ocae271, 2024.

Mesinovic, M., Molaei, S., Watkinson, P., and Zhu, T. Dynagraph: Interpretable multi-label prediction from ehrs via dynamic graph learning and contrastive augmentation. *arXiv preprint arXiv:2503.22257*, 2025.

Qiu, J., Hu, Y., Li, L., Erzurumluoglu, A. M., Braenne, I., Whitehurst, C., Schmitz, J., Arora, J., Bartholdy, B. A., Gandhi, S., et al. Deep representation learning for clustering longitudinal survival data from electronic health records. *Nature Communications*, 16(1):2534, 2025.

Rocheteau, E., Liò, P., and Hyland, S. Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit. In *Proceedings of the conference on health, inference, and learning*, pp. 58–68, 2021.

Tang, S., Tariq, A., Dunnmon, J. A., Sharma, U., Elugunti, P., Rubin, D. L., Patel, B. N., and Banerjee, I. Predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks. *IEEE Journal of Biomedical and Health Informatics*, 27(4):2071–2082, 2023.

Xu, H., Duan, Z., Wang, Y., Feng, J., Chen, R., Zhang, Q., and Xu, Z. Graph partitioning and graph neural network based hierarchical graph matching for graph similarity computation. *Neurocomputing*, 439:348–362, 2021.