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

Anonymous Author(s)

Abstract

We present a novel multi-modal graph learning framework for competing risks survival analysis from electronic health records (EHRs). While recent work has demonstrated the power of deep learning for dynamic risk prediction, most models are constrained to unimodal inputs or static graph structures and cannot model competing clinical endpoints. 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. This design enables dynamic construction of informative substructures both within and across modalities, offering interpretable predictions for multiple competing outcomes. We further propose a composite training objective combining survival likelihood, temporal ranking, and graph regularisation losses to improve risk discrimination, calibration, and structural consistency over time. Our model outperforms state-of-the-art 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. These results establish a new foundation for trustworthy and data-efficient risk estimation in clinical settings.

CCS Concepts

• **Applied computing** → **Health informatics**.

Keywords

machine learning for healthcare, survival analysis, graph neural networks, multimodal learning, competing risks, interpretability

ACM Reference Format:

Anonymous Author(s). 2025. Multi-Modal Interpretable Graph for Competing Risk Prediction with Electronic Health Records. In *Proceedings of Temporal Graph Learning Workshop, SIGKDD International Conference on Knowledge Discovery and Data Mining 2025 (Workshop, KDD 2025).* ACM, New York, NY, USA, 19 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The rapid increase in the availability of electronic health records (EHRs), coupled with advanced machine learning models, has led to an eruption of proposed solutions to clinical risk prediction challenges [8, 31]. One of the challenges is the generation of robust and dynamic risk profiles that capture the patient state accurately. Time-to-event modelling allows us to model not just whether an event, like death in the hospital, happens or not, but also the time distribution of the risk of death [27]. This enables the comparison of patients with different risk profiles throughout time using longitudinal measurements instead of solely the last measured value to inform prioritisation, treatment, and lifestyle decisions [16]. To holistically assess dynamic risk requires learning the complete picture of the patient's state beyond just one set of measurements, especially in evolving scenarios of intensive care [10]. To that end, multi-modal deep learning empowers the incorporation of multiple modalities like demographics, vital signs, laboratory measurements, medical history, and clinical text reports to inform decision-making with machine learning.

A limiting factor of time-to-event modelling is its often restriction to single-risk scenarios and constraining assumptions on the behaviour of the covariates over time. The most popular example is the Cox Proportional Hazards model, which assumes a single static modality for covariates and fixed proportionality of change in risk over time between patients [23]. Extensions of the Cox model exist for dynamic covariates and competing risk scenarios, however, they make independence assumptions on the risks themselves or do not perform as well on longitudinal time-series scenarios as deep learning models [2, 39]. Multi-modal deep learning with a competing risk extension, in which multiple risks or complications are affecting the patient, can address these challenges and avoid discarding important information about the patient when estimating risk distributions for several events.

Recent multi-modal deep learning proposals for electronic health records use the fusion of separate and different neural network modules for each modality to make single-event predictions [11, 30, 38]. These approaches rely on different encoding embeddings for different modalities that might not be self-cohesive and have to account for irregularity between the time-series and the clinical notes over time. They also often do not encode the times of the reports, which can lead to missing temporal patterns from text sources as well as leakage if the reports contain information ahead of the risk distribution estimate. And while they have shown promise at single-label classification tasks in healthcare, there has been no attempt to model the challenging yet informative competing risk scenario.

Graph representation learning encodes data into graphs that graph neural networks (GNNs) can learn to capture the intricate relationships and dependencies between medical entities to generate rich, context-aware embeddings for further downstream tasks [4]. Spatio-temporal graph solutions have been proposed to construct learnable graph representations of irregularly sampled EHR data without any pre-defined graph structures [24]. By modelling each modality through direct construction of separate learnable spatio-temporal graphs, graph representation learning offers a promising

---

Workshop, KDD 2025, Toronto, Canada

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

solution to multi-modal integration. As explainability is an important requirement for healthcare solutions, structural graph information provides an intuitive approach to interpretability challenges [5, 41]. Using graph attention within a hierarchical framework enables the combining of multiple modalities while securing global interpretability. We propose constructing learnable parallel graph representations without any predefined structures of different EHR modalities with hierarchical graph attention to estimate competing risk distributions of multiple events simultaneously. Our approach successfully integrates dynamic graphs for time-series, static graphs for demographics, similarity graphs for medical history, and text vectorised graphs for clinical reports to capture patient physiological risk over time. The global attention weights also help focus on the most important patterns between modalities, enabling the model to learn meaningful representations while providing insight into clinical feature importance.

In this work, we present a new approach to multi-modal learning from EHR data for time-to-event modelling and advance the state-of-the-art by:

- Introducing a novel spatio-temporal graph approach that cohesively learns from and integrates multiple modalities in electronic health records, including diagnostic coding of medical history, static and time-series features, and clinical text reports.
- Proposing a hierarchical graph attention framework for multi-modal learning that provides global explainability.
- Combining ranking, longitudinal risk prediction, and structural losses to estimate competing risk distributions of multiple events simultaneously and reliably.

We evaluated the proposed model in four real-world healthcare datasets from the intensive care and liver transplant settings, demonstrating superior performance over state-of-the-art methods in deep learning for competing risk prediction while providing robust interpretability.

2 RELATED WORK

2.1 Time-to-event Modelling

Models for time-to-event prediction under competing risks span classical statistical methods and modern deep learning approaches. Traditional techniques such as the Cox proportional hazards model, the Fine–Gray subdistribution model, and Random Survival Forests offer interpretable estimates but rely on strong assumptions like proportional hazards and are ill-suited for high-dimensional, longitudinal EHR data [18, 36]. Discrete-time approaches, including logistic hazard models, provide greater flexibility and are naturally compatible with neural network architectures, though they are often constrained to static inputs and single-risk settings [40]. Deep learning methods such as DeepSurv, DeepHit, Dynamic-DeepHit, and DySurv relax many of these assumptions, enabling direct modelling of cumulative incidence functions and the integration of complex temporal patterns [12, 16, 17, 25]. In ICU contexts, dynamic models like Dynamic-DeepHit and DySurv have demonstrated strong predictive performance by capturing evolving physiological trends and non-linear feature interactions, albeit with reduced interpretability and transparency [28]. Except for Dynamic-DeepHit, most models in EHR provide only independent, cause-specific risk estimates, failing to model competing events jointly. Furthermore, learning from longitudinal and multi-modal EHR data remains challenging due to high dimensionality, irregular sampling, and pervasive missingness, contributing to poor generalizability. To date, no model has explicitly addressed the integration of multiple data modalities and the modelling of competing risks in a unified framework, two critical challenges in machine learning for healthcare. Without the capacity to capture inter-modal and temporal dependencies, existing approaches may fall short in representing patient state accurately for complex prognostic tasks.

2.2 Graph Neural Networks

Graph Neural Networks (GNNs) are well-suited to modeling temporal and structural dependencies in longitudinal health data, particularly in the presence of missingness and modality heterogeneity [3, 34]. Spatio-temporal graphs can capture evolving interactions across features and modalities [35]. Early approaches such as GCNs, GATs, and GraphSAGE rely on pre-defined graphs for downstream tasks [9, 13, 33], constructed via heuristics like Dynamic Time Warping (SimTSC) or convolutional embeddings (MedGNN) [6, 37]. However, such fixed graphs may not reflect task-relevant dynamics. Recent models like TodyNet and DynaGraph remove this constraint by learning graph structures directly from data in an end-to-end manner [22, 24]. While TodyNet struggles with scalability and interpretability, DynaGraph introduces limited temporal interpretability but remains unimodal.

MM-STGNN represents a recent effort to integrate multi-modal data such as demographics, time-series, and imaging for hospital readmission prediction using a shared spatio-temporal graph [32]. However, it fuses modalities via a simple MLP, which underutilises complex temporal and inter-modality dependencies and constructs the graph prior to training using Gaussian kernel similarity on static features, thereby assuming a shared similarity metric across heterogeneous data types. This limits its ability to learn modality-specific dynamics, adapt graph structure during training, or capture fine-grained inter-feature relationships. Moreover, its interpretability and flexibility in modelling complex dependencies are constrained by its reliance on fixed graph topology and global aggregation via GraphSAGE.

In this work, we propose a novel multi-modal spatio-temporal GNN that dynamically constructs and fuses modality-specific graphs with hierarchical attention. Our model learns structural dependencies jointly with survival prediction, incorporates time-series, diagnostic codes, demographics, and radiographic text in an end-to-end framework, and provides interpretable, cause-specific risk estimates in a competing risks setting, addressing key limitations of prior GNN-based approaches.

3 Data

We evaluate our model on five real-world EHR datasets spanning ICU, emergency department (ED), and transplant settings. For all datasets, laboratory features missing in over 75% of encounters are removed, and remaining missing values are imputed via forward and backward filling. We require at least six time-steps per patient, padding shorter sequences with the most recent observation. Patients are split 8:1:1 into training, validation, and test sets.

MC-MED is a multimodal ED dataset comprising 118,385 visits from 70,545 adults at Stanford Health Care, with time-series vitals, demographics, ICD9/10 histories, labs, and radiography reports. Outcomes include ICU admission (2.3%), hospitalisation (24.9%), and ED observation (13.4%). PBC2 contains monthly longitudinal biomarkers from a primary biliary cirrhosis trial, with death and liver transplantation as competing risks. MIMIC-IV and eICU datasets contain high-resolution ICU data and use in-ICU mortality as the target event; preprocessing follows Mesinovic et al. [25]. The SUPPORT dataset includes static features for 8873 seriously ill inpatients followed for 6-month mortality. Full details are in Supplementary Section A.

4 METHODS

4.1 Notation

The input consists of multi-modal clinical data for P patients. For each patient i, the data is represented as a collection of modality-specific feature sets:

$$X_i = \left\{X_i^{(\text{TS})}, X_i^{(\text{Static})}, X_i^{(\text{Xray})}, X_i^{(\text{ICD})}\right\}$$

where:

- $X_i^{(\text{TS})} \in \mathbb{R}^{d \times l}$ denotes multivariate time-series measurements with d features over l time steps,
- $X_i^{(\text{Static})} \in \mathbb{R}^{d_{\text{Static}}}$ denotes static demographic features,
- $X_i^{(\text{Xray})} \in \mathbb{R}^{d_{\text{Xray}}}$ denotes textual embeddings of radiographic reports,
- $X_i^{(\text{ICD})} \in \mathbb{R}^{d_{\text{ICD}}}$ denotes encoded diagnostic code history.

The entire cohort of m patients is denoted as:

$$\boldsymbol{X} = \left\{X_1, X_2, \ldots, X_p\right\}$$

Corresponding labels are given as $\boldsymbol{Y} = \left\{Y_1, Y_2, \ldots, Y_p\right\}$, where each Y_i is a tuple:

$$Y_i = (\epsilon_i, t_i)$$

with $\epsilon_i \in \mathcal{E}$ representing the observed event type and $t_i \in \mathbb{R}_{\geq 0}$ representing the observed time-to-event. The set of possible events is defined as:

$$\mathcal{E} = \{\varnothing, 1, 2, \ldots, E\}$$

where $\varnothing$ denotes right-censoring (i.e., the absence of an event during the observation window), and E is the number of distinct event types. We assume that censoring is non-informative. The observed time t_i is defined as:

$$t_i = \min(T_i, C_i)$$

where T_i is the true event time and C_i is the censoring time for patient i.

Dynamic Graph Construction for Time-Series. For longitudinal time-series data, we partition each patient's sequence into s equally-sized time windows, constructing a dynamic graph representation for each window. Each adjacency matrix is initialized through learnable node embeddings:

$$A = \Theta^{\top} \cdot \Psi \in \mathbb{R}^{d \times d},$$

where $\Theta = \left[\theta_{t,1}, \theta_{t,2}, \ldots, \theta_{t,d}\right]$ and $\Psi = \left[\psi_{t,1}, \psi_{t,2}, \ldots, \psi_{t,d}\right]$ are randomly initialized learnable parameters representing source and target node embeddings for time window t.

To promote computational efficiency, the adjacency matrix is sparsified using top-k edge selection:

$$\begin{gathered}\text{idx}, \text{idy} = \operatorname{argtopk}(A[:, :]) \quad \text{with} \quad \text{idx} \neq \text{idy}, \\ A[-\text{idx}, -\text{idy}] = 0.\end{gathered}$$

Temporal evolution is preserved across windows by dynamically connecting features over time. For each time slot, new vertices are introduced to represent features from both the current and previous time windows, resulting in a vertex set:

$$\left\{v_{(t,1)}, v_{(t,2)}, \ldots, v_{(t,D)}, v_{(t-1,1)}, v_{(t-1,2)}, \ldots, v_{(t-1,D)}\right\},$$

where directed edges link $v_{(t-1,d)}$ to $v_{(t,d)}$ for each feature $d = 1, \ldots, D$. To prevent exponential growth in the number of nodes, node embeddings from previous windows are aggregated, and redundant vertices are pruned.

The final time-series graph representation consists of a set of adjacency matrices:

$$A = \{A_1, A_2, \ldots, A_s\} \in \mathbb{R}^{d \times d \times s},$$

capturing spatio-temporal patterns across features and time.

Multi-Modal Graph Construction. Our model simultaneously processes multiple heterogeneous data modalities, constructing an independent graph for each. Specifically, we generate graphs for multivariate time-series measurements, static demographic variables, radiographic report embeddings (X-ray text), and diagnostic code histories (ICD9/10). Each modality is represented by its own adjacency matrix:

$$A^{(m)} \in \mathbb{R}^{d_m \times d_m}$$

where $m \in \{\text{TS}, \text{Static}, \text{Xray}, \text{ICD}\}$ and d_m is the number of features in modality m.

Let $C = \{c_1, \ldots, c_{d_{\text{ICD}}}\}$ be the top $d_{\text{ICD}} = 500$ most frequent diagnostic codes in the training set. Each code c_i is embedded into a feature vector $e_i \in \mathbb{R}^d$ using a co-occurrence-based embedding model trained on ICD sequences from the training data only.

We construct a fully connected intra-modality graph $A^{(\text{ICD})} \in \mathbb{R}^{d_{\text{ICD}} \times d_{\text{ICD}}}$ using pairwise cosine similarity:

$$A_{ij}^{(\text{ICD})} = \frac{e_i^{\top} e_j}{\|e_i\| \cdot \|e_j\|}, \quad \forall i, j$$

To reduce computational cost, we apply a spatial pooling operation using a learnable function f_{θ}^{ICD} to project this graph to a lower-dimensional representation:

$$\tilde{A}^{(\text{ICD})} = f_{\theta}^{\text{ICD}}(A^{(\text{ICD})})$$

where $\tilde{A}^{(\text{ICD})} \in \mathbb{R}^{d' \times d'}$ are the pooled adjacency and node features, respectively, with $d' \ll d_{\text{ICD}}$.

For radiographic reports, textual feature embeddings are generated using Clinical-Longformer, a transformer model pre-trained on MIMIC-III chest radiograph reports [19, 20]. Let $\mathcal{R}^{(i)} = \{r_1^{(i)}, \ldots, r_T^{(i)}\}$ denote the set of radiographic reports for patient i, ordered by time. Each report is encoded via the frozen Clinical-Longformer model Φ:

$$x_t^{(i)} = \Phi(r_t^{(i)}), \quad x_t^{(i)} \in \mathbb{R}^d$$

We construct a patient-specific temporal graph $A^{(i,\text{Xray})} \in \mathbb{R}^{T \times T}$ using a Gaussian similarity kernel:

$$A_{tt'}^{(i,\text{Xray})} = \exp\left(-\frac{\|x_t^{(i)} - x_{t'}^{(i)}\|_2^2}{\tau}\right), \quad \forall t, t' \in \{1, \dots, T\}$$

where τ is a temperature hyperparameter. The corresponding node feature matrix is $X^{(i,\text{Xray})} = [x_1^{(i)}, \dots, x_T^{(i)}]^\top$.

Each intra-modality graph $A^{(m)}$ (including TS, Static, ICD, Xray) is paired with a corresponding interpretability matrix $I^{(m)} \in \mathbb{R}^{d_m \times d_m}$ as described below. These matrices are trained jointly with the model and updated based on gradient attribution to quantify feature-level and structural importance for survival prediction.

Together, these modality-specific graphs contribute to the final fused multi-modal graph A^{Fused} and its interpretability matrix I^{Fused} used in downstream graph learning.

4.2 Hierarchical Attention

To enable interactions between modalities, we introduce learnable cross-modality attention matrices:

$$W^{(m \to n)} \in \mathbb{R}^{d_m \times d_n}$$

for each ordered modality pair (m, n), where $W^{(m \to n)}$ captures the directed influence from features in modality m to features in modality n. These matrices are randomly initialised and updated during training to optimize the survival prediction objective. To capture the relative importance of features and modalities in the final prediction, we associate each intra-modality adjacency matrix $A^{(m)}$ and each cross-modality matrix $W^{(m \to n)}$ with a corresponding interpretability matrix. These interpretability matrices are updated end-to-end during training based on the contribution of nodes and edges to the model's loss. Details on the gradient computation of the interpretability weights are in Supplementary section C, and proof of convergence for the weight matrices in section 4. This hierarchical attention mechanism allows our model to provide fine-grained interpretability across individual features, within modalities, and between modalities.

The full fused multi-modal graph $A^{\text{Fused}} \in \mathbb{R}^{d_{\text{Fused}} \times d_{\text{Fused}}}$ is constructed as:

$$\begin{bmatrix} A^{(\text{TS})} & W^{(\text{TS}\to\text{Static})} & W^{(\text{TS}\to\text{Xray})} & W^{(\text{TS}\to\text{ICD})} \\ W^{(\text{Static}\to\text{TS})} & A^{(\text{Static})} & W^{(\text{Static}\to\text{Xray})} & W^{(\text{Static}\to\text{ICD})} \\ W^{(\text{Xray}\to\text{TS})} & W^{(\text{Xray}\to\text{Static})} & A^{(\text{Xray})} & W^{(\text{Xray}\to\text{ICD})} \\ W^{(\text{ICD}\to\text{TS})} & W^{(\text{ICD}\to\text{Static})} & W^{(\text{ICD}\to\text{Xray})} & \tilde{A}^{(\text{ICD})} \end{bmatrix}$$

where diagonal blocks represent intra-modality graphs and off-diagonal blocks represent learnable cross-modality interactions.

Each modality-specific adjacency matrix $A^{(m)}$ and each cross-modality matrix $W^{(m \to n)}$ is paired with a corresponding interpretability matrix:

$$I^{(m)}, \quad I^{(m \to n)} \in \mathbb{R}^{d_m \times d_n}$$

initialised uniformly and updated during training via gradient attribution with respect to the model's loss. These interpretability matrices quantify the importance of nodes and edges within and between modalities. The final fused interpretability matrix I^{Fused} mirrors the structure of A^{Fused}, allowing for fine-grained attribution of model predictions across heterogeneous feature types.

The adjacency and interpretability matrices are then aggregated for a final graph representation G:

$$G^{\text{Final}} = (A^{\text{Fused}} \parallel I^{\text{Fused}})$$

capturing intra-modality structures, cross-modality interactions, and hierarchical feature importance in a unified framework optimised for competing risks prediction. This makes it possible to not only to model complex multi-modal survival outcomes but also to identify critical features and cross-modal interactions driving individual patient risk predictions. Details on graph learning with Graph Isomorphic Network (GIN) can be found in the Supplementary section E.

4.3 Model Training

Due to the complexity of the model, we use temporal pooling on the learned graphs from the GIN layers as described in the Supplementary section F, and we add a regularisation term to help prevent overfitting and support learning stability:

$$\mathcal{L}_{\text{reg}} = \lambda \sum_{(k,l) \in E} \|\mathbf{h}_k - \mathbf{h}_l\|^2 \tag{1}$$

where λ is a hyperparameter controlling the strength of the regularization, (k, l) represents an edge connecting nodes k and l, and $\mathbf{h}_k$, $\mathbf{h}_l$ are the feature representations of nodes k and l, respectively. Please note that E here represents the set of edges in the dynamically constructed graph, where each edge connects a pair of nodes (features).

For our right-censored setup, we need to estimate the joint distribution of the first event time and multiple competing events. We do so by directly estimating the negative log-likelihood of this distribution. For those patients who have suffered the specific event, we capture both the outcome and the time at which it occurs. For censored patients, we capture the censoring time conditioned on the measurements recorded prior to the censoring. If we assume $\widehat{a}_t = \hat{P}(T = \delta, \epsilon = \epsilon \mid X)$ represents the estimated probability of experiencing an event ϵ at time δ, this loss is defined as:

$$\begin{aligned} \mathcal{L}_{NLL} = -\sum_{i=1}^{N} \Bigg[& \mathbb{1}\left(\epsilon^i \neq \varnothing\right) \cdot \log\left(\frac{\widehat{a}^i_{\epsilon^i,\delta^i}}{1 - \sum_{\epsilon \neq \varnothing} \sum_{n \leq t^i_{ji}} \widehat{a}^i_{\epsilon,n}}\right) \\ & + \mathbb{1}\left(\epsilon^i = \varnothing\right) \cdot \log\left(1 - \sum_{\epsilon \neq \varnothing} \hat{F}_\epsilon\left(\delta^i \mid X^i\right)\right)\Bigg] \end{aligned} \tag{2}$$

The second component of the loss is a ranking loss designed to refine the model's ability to estimate cause-specific cumulative incidence functions (CIFs). Inspired by [16], this loss encourages the model to assign higher risk scores at the true event time for patients who experience the event earlier compared to those who survive longer. However, since patients' longitudinal measurements may begin at varying points in their clinical trajectory, direct comparisons based on absolute event times are often not meaningful. To address this, we compute relative times with respect to the most recent measurement. For a subject i, we define $s_i = \delta_i - t_i^{(J_i)}$, where δ_i is the event time and $t_i^{(J_i)}$ is the time of the last observation. We then define a pair (i, j) as an acceptable pair for event type ϵ if subject i experiences event ϵ at time s_i, while subject j does not experience any event until s_i, i.e., $s_j > s_i$. For these pairs, the model's predicted CIFs should satisfy:

$$\hat{F}_\epsilon\left(s_i + t_i^{(J_i)} \mid X_i\right) > \hat{F}_\epsilon\left(s_i + t_j^{(J_j)} \mid X_j\right)$$

ensuring that the estimated risk for subject i, who experienced the event earlier, is higher than that for subject j, who survived beyond that point. The ranking loss is computed over such acceptable pairs, allowing the model to learn consistent risk ordering across patients with heterogeneous measurement histories. For notational simplicity, let $R_\epsilon^{(i)} := \hat{F}_\epsilon\left(s_i + t_i^{(J_i)} \mid X_i\right)$ denote the predicted cumulative incidence for event type ϵ at the effective event time for subject i. The ranking loss is then:

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

where $M_{ij}^{\epsilon} = 1\left(\epsilon_i = \epsilon, s_i < s_j\right)$ indicates whether the pair (i, j) is acceptable for event type ϵ, and $\eta(\cdot)$ is a smooth loss function comparing the predicted CIFs. We adopt the formulation $\eta(a, b) = \exp\left(-\frac{a-b}{\sigma}\right)$, which penalises incorrect risk rankings in a margin-sensitive manner. For simplicity, we assume uniform weighting across event types with $\mu_\epsilon = \mu$ for all ϵ. This ranking term is integrated into the total loss to improve temporal discrimination across competing risks, particularly in the presence of censored and irregular longitudinal data.

Additionally, we define a structural similarity loss to ensure that the learned graph structure remains close to the original (previous timepoint) structure:

$$\mathcal{L}_{\text{struct}} = \mu\left(1 - \frac{\sum_{i,j} A_{ij} \cdot A'_{ij}}{\sqrt{\sum_{i,j} A_{ij}^2} \cdot \sqrt{\sum_{i,j} {A'}_{ij}^{2}}}\right) \tag{3}$$

where A is the original adjacency matrix, A' is the adjacency matrix after augmentation, and μ is a hyperparameter. The structural loss aids in the convergence of the adjacency matrix throughout learning. The final loss is the sum of the regularisation loss for complexity and overfitting adjustment, the joint negative log-likelihood of the competing risks, and a ranking loss to fine-tune the model to each cause accordingly, α, β, and γ are considered hyperparameters:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{reg}} + \alpha\mathcal{L}_{\text{NLL}} + \beta\mathcal{L}_{\text{rank}} + \gamma\mathcal{L}_{\text{struct}} \tag{4}$$

After the final graph embedding is created through graph expansion and overlaying hierarchical attention, G_s is passed through a multi-layer temporal graph isomorphism network (GIN) for graph learning. The output of the GIN is a graph which is pooled temporally to reduce the number of nodes with convolutional clustering and decrease the computational costs of the training. The reduced graph is then flattened and passed through E parallel multilayer perception subnetworks and a joint softmax layer to produce measures of probability of the estimated joint distribution of the event times and competing events. For a patient X, since the event times are discretised, the estimated CIF for a specific cause at time δ is:

$$\hat{F}_\epsilon\left(\delta \mid X\right) = \frac{\sum_{t_J < \delta \leq \Delta} \widehat{a}_{\epsilon,\delta}}{1 - \sum_{\epsilon \neq \varnothing} \sum_{n \leq t_J} \widehat{a}_{\epsilon,n}} \tag{5}$$

5 RESULTS

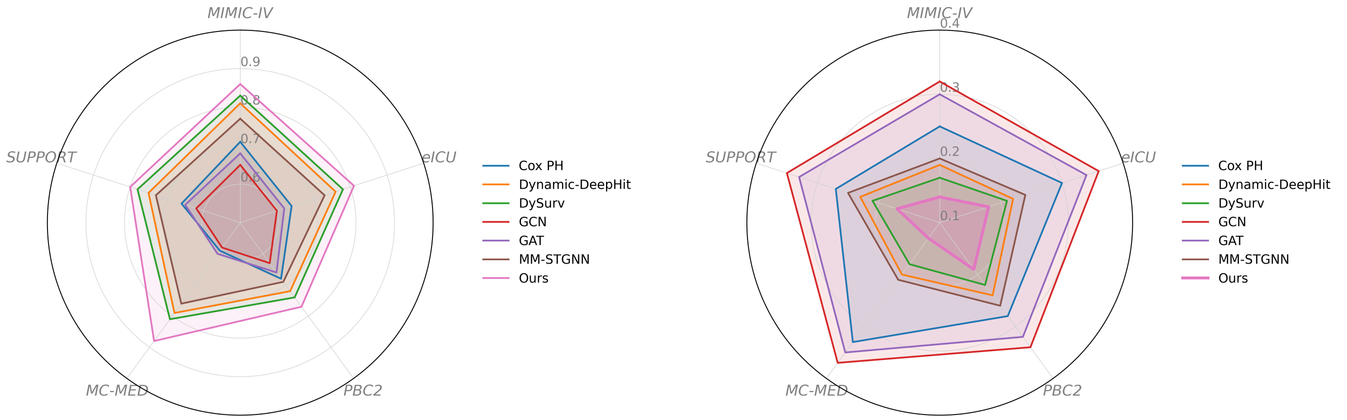
5.1 Model Comparison

We evaluate our proposed model against two groups of baselines under both single-risk and competing-risk survival analysis scenarios. The first group includes state-of-the-art statistical and deep learning models for survival analysis, comprising static models (Cox PH, DeepSurv, DeepHit) and dynamic models capable of processing longitudinal time-series data (Dynamic-DeepHit, DySurv). The second group consists of graph neural network models adapted for survival tasks, including GCN, GAT, GraphSAGE, TodyNet, DynaGraph, and MM-STGNN. For MM-STGNN, we replace the original imaging modality with textual embeddings from radiographic reports to match our input setup, and use the competing risk loss function for survival adaptation, similarly for MedGNN. For TodyNet and DynaGraph, we use the time-series and/or static modalities from the relevant datasets. Details on experimental settings and implementation choices are provided in Supplementary Section A.

All experiments were conducted using PyTorch 3.8 and an NVIDIA A100 Tensor Core GPU. Evaluation is based on three complementary metrics: time-dependent concordance (including cause-specific concordance for competing risks), integrated Brier score (IBS), and integrated binomial log-likelihood (IBLL), as detailed in Supplementary Section B.

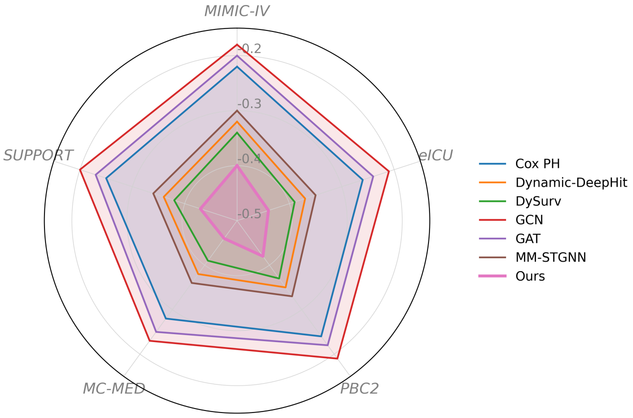
Figure 1 presents spider plots comparing our model against all baselines across five datasets in the single-risk setting. A complete set of comparisons across models and metrics is available in Supplementary Section G. As illustrated, our model consistently outperforms all baselines across different 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 compared to Dynamic-DeepHit and related methods indicate better generalisation and calibrated event probability estimates, avoiding the inflated concordance behaviour seen in prior ranking-based models.

We evaluate our proposed model against several advanced models applied to competing risk prediction, including Dynamic-DeepHit, MedGNN, and MM-STGNN. As shown in Table 1, our model outperforms all baselines across both the PBC2 and MC-MED datasets in terms of cause-specific time-dependent concordance. On PBC2, our model improves over MedGNN and MM-STGNN by 2.5 and 2.0 percentage points, respectively, in predicting death, with similar gains for liver transplant prediction. On MC-MED, the advantage is most notable in hospital admission prediction, where our model achieves 0.880 ± 0.010 concordance, outperforming MM-STGNN (0.798 ± 0.012) and MedGNN (0.791 ± 0.012). These results highlight the advantage of our dynamic and hierarchical graph formulation, which jointly learns intra- and inter-modality structure with attention-based interpretability, in contrast to fixed similarity graphs or late fusion approaches used in prior work.



(a) The time-dependent concordance index results across five datasets for the single-risk scenario (higher is better).

(b) The integrated Brier score results across five datasets for the single-risk scenario (lower is better).



(c) The integrated binomial log-likelihood results across five datasets for the single-risk scenario (lower is better).

Figure 1: Spiderplots of average model results (for five seeds) across our five datasets for three different common time-to-event metrics, namely a) time-dependent concordance, b) integrated Brier score, and c) integrated binomial log-likelihood. For MC-MED the event results highlighted here are for

Table 1: 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.
Dynamic-DeepHit	0.758	0.740	0.812	0.785	0.773
MedGNN	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**

5.2 Model Interpretability

To better understand the temporal dynamics of model decision-making, we visualise the interpretability weights associated with the top 10 time-series features across six time steps for each of the three competing outcomes in the MC-MED dataset. These weights, derived from the modality-specific interpretability matrices I^{TS}, are normalised and plotted as heatmaps in Figure 2. Each heatmap reveals how feature importance evolves across the patient's emergency department stay, with each time step corresponding to a discrete, regular interval in the model's prediction window. Higher weight magnitudes indicate a stronger influence on the model's loss during training, and thus on the predicted CIF for each outcome.

The learned interpretability matrices exhibit both temporal variability and sparsity, demonstrating the model's ability to identify and adapt to transient, yet informative, physiological signals. For example, features such as respiratory rate (RR) and triage vitals display early peaks in attention for the ED observation outcome, while renal markers such as creatinine and BUN grow more predictive over time for hospital admission. These patterns reflect clinically plausible trajectories where early vital sign abnormalities inform short-term disposition decisions, and longitudinal lab abnormalities guide inpatient escalation.

To further probe how the model fuses heterogeneous clinical data sources, we visualise the learned attention matrices across modalities in Figure 3. Each 4×4 matrix reflects the modality-level interpretability scores for the three competing risk outcomes in MC-MED. These are derived by averaging attention weights over modality-specific blocks from the final fused adjacency matrix. The diagonal entries quantify the within-modality contribution, while off-diagonal values indicate inter-modality influence. We observe strong attention weights within the time-series and radiography modalities across all outcomes, consistent with their dense temporal and semantic signal. Interestingly, ICU admission predictions exhibit high cross-modality attention from ICD codes to both TS and radiography, suggesting diagnostic history and clinical context are essential for identifying escalation risk. Conversely, observation outcome predictions emphasise more evenly distributed modality interactions, reflecting the complexity of these clinical profiles.

These visualisations improve the transparency of the model's decision logic, offering clinicians a window into what the model is paying attention to at different points in time. This can help identify clinically actionable biomarkers and facilitate more nuanced risk communication and triage planning in high-stakes ED settings. A more detailed exploration of these trends is provided in the Discussion section.

5.3 Model Evaluation

5.3.1 Ablation Studies. To assess the contribution of individual components within our proposed model framework, we conducted a comprehensive ablation study across both competing risk and single-risk survival settings. On the MC-MED dataset, which involves three competing outcomes (ICU admission, hospital admission, and in-ED observation), the full model consistently achieves the highest cause-specific concordance across all outcomes (Table 2), demonstrating the effectiveness of integrating architectural, loss, and input design choices. Removing the ranking loss or the structural stabilisation term leads to measurable drops in performance, highlighting their complementary roles in learning accurate event-time distributions and stabilising graph evolution, respectively. Excluding either the diagnostic history (ICD9/10 codes) or radiography embeddings similarly reduces performance, suggesting the value of multi-modal fusion. Furthermore, replacing the GIN module with a standard GCN results in consistent declines across outcomes, underscoring the benefit of structure-aware message passing in our dynamic graph encoder. To further validate these findings, we perform additional ablations on the MIMIC-IV and eICU datasets for ICU mortality prediction in the single-risk setting (Table 3). We observe that the same trends persist: removing ranking or structural losses reduces concordance, excluding modalities impairs predictive performance, and replacing GIN with GCN degrades model accuracy. These results demonstrate the robustness of the model's design choices across diverse clinical datasets and survival prediction regimes.

5.3.2 Calibration Experiments. In addition to ranking-based and discrimination metrics, we evaluate the probabilistic calibration of our model compared to Dynamic-DeepHit across two ICU datasets and three prediction horizons. The Supplementary contains calibration plots at 48h, 96h, and 144h for both the MIMIC-IV (top row) and eICU (bottom row) datasets, alongside bootstrap-based confidence intervals and reported Expected Calibration Error (ECE) for each model. In both cohorts, our model demonstrates closer alignment to the ideal calibration line and consistently lower ECE values. At the 48h prediction horizon, our model achieves an ECE of 0.024 on MIMIC-IV and 0.030 on eICU, compared to 0.068 and 0.082 for Dynamic-DeepHit, respectively. This performance trend persists across longer horizons as well, indicating that our model not only produces better discriminative rankings but also more calibrated survival probabilities. These results highlight the model's robustness across care settings and its suitability for clinical risk estimation tasks requiring reliable uncertainty quantification.

6 DISCUSSION

Most dynamic survival models, including Dynamic-DeepHit, overlook the structural and temporal heterogeneity inherent in multimodal EHR data. Existing GNN-based approaches either rely on static graph assumptions or treat modalities as independent streams, failing to model how relationships between features evolve over time or across data types. Our approach introduces a novel framework that constructs modality-specific spatio-temporal graphs and fuses them via hierarchical cross-modality attention, enabling interpretable and dynamic estimation of cause-specific risk in competing outcomes. Unlike prior multi-modal methods that use early fusion or fixed graph topologies, our model learns both intra- and intermodality structure end-to-end, offering a flexible and expressive mechanism to capture clinical complexity.

Extensive benchmarking across five datasets shows our model consistently outperforms strong baselines in discrimination, calibration, and interpretability. The combination of ranking loss, temporal structural regularisation, and modality-aware graph attention provides stable graph learning and improves temporal risk ranking. Our interpretability framework uncovers clinically meaningful insights at both feature and modality levels, such as rising renal risk markers and diagnostic history relevance for ICU escalation. These results establish a new direction for temporal graph learning in healthcare—one that is multimodal, interpretable, and tailored to the high-stakes setting of competing risks.

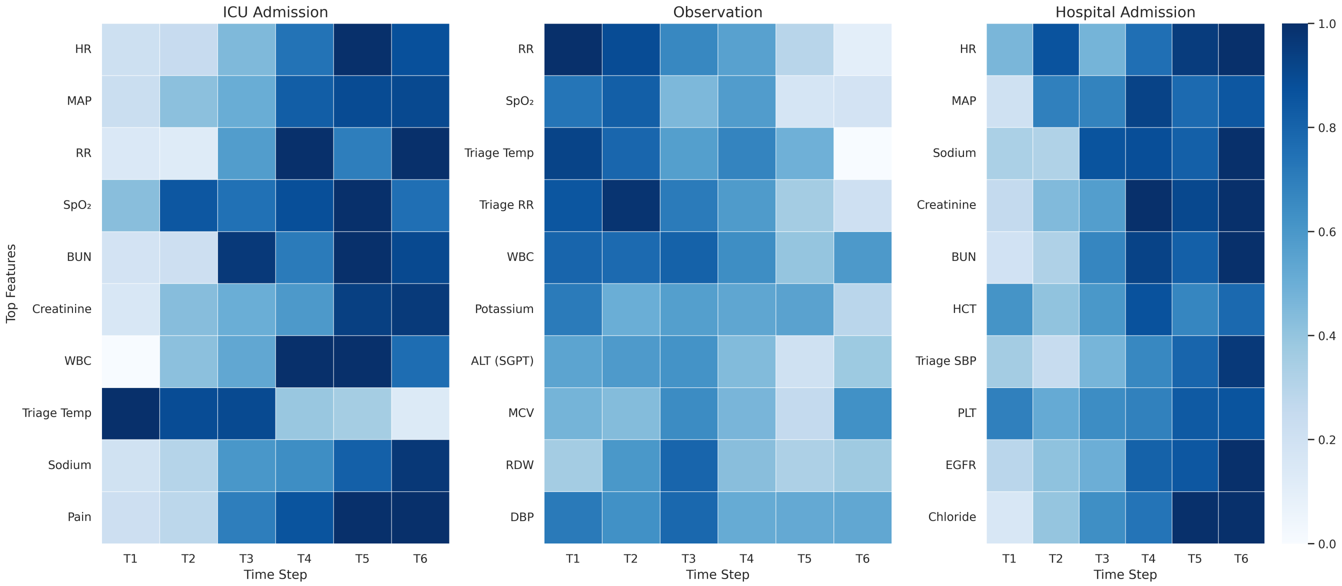


Figure 2: Temporal interpretability heatmaps of the top 10 time-series features across 6 time steps for each competing outcome in MC-MED. Scores from I^{TS} highlight local and global predictive contributions.

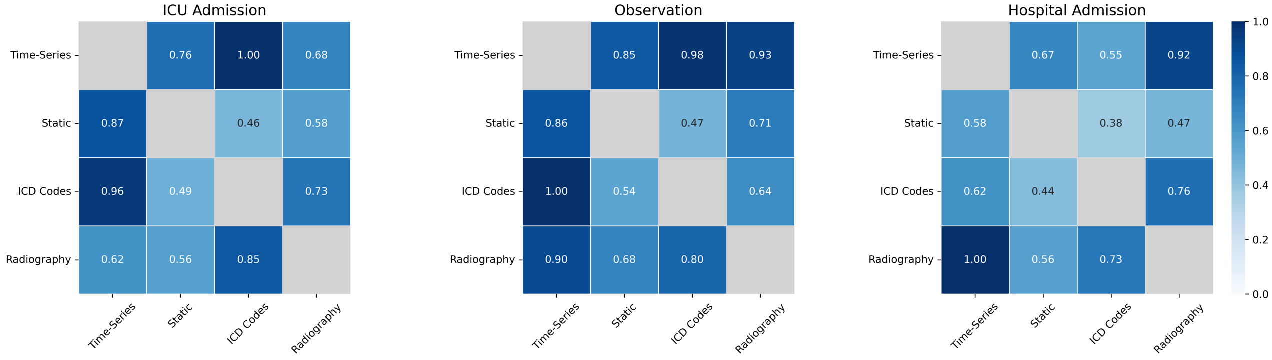


Figure 3: Modality-level attention weights for ICU admission, ED observation, and hospital admission. Each 4×4 heatmap shows intra- and cross-modality contributions from the fused attention matrix.

Table 2: 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 (ICU admission, discharge, in-hospital observation), averaged over 5 seeds. Higher is better.

Model Variant	ICU Admission	Hospital Admission	Observation
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)	0.799 ± 0.011	0.831 ± 0.014	0.774 ± 0.016
w/o Radiography Embeddings	0.805 ± 0.009	0.785 ± 0.012	0.775 ± 0.015
GCN backbone	0.811 ± 0.008	0.813 ± 0.011	0.782 ± 0.013
Ours (Full Model)	**0.827 ± 0.007**	**0.880 ± 0.010**	**0.797 ± 0.012**

Table 3: Ablation study on ICU mortality prediction using MIMIC-IV and eICU. We report the time-dependent concordance index ($C^{\text{td}}_{\text{ind}}$) for the single-risk setting, averaged over 5 seeds. Higher is better.

Model Variant	**MIMIC-IV**	**eICU**
w/o Ranking Loss	0.814 ± 0.011	0.801 ± 0.014
w/o Structural Loss	0.817 ± 0.010	0.805 ± 0.012
w/o History (ICD9/10)	0.812 ± 0.012	0.798 ± 0.015
w/o Radiography Embeddings	0.816 ± 0.010	0.800 ± 0.013
GCN backbone	0.820 ± 0.009	0.807 ± 0.012
Ours (Full Model)	**0.860 ± 0.007**	**0.810 ± 0.010**

A Code and Data

A.1 Data Description

The pre-processing pipeline for MIMIC-IV was based on previously published workflows, and eICU was based in part on work done by [25, 29]. 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.

The most relevant feature distributions for eICU, MIMIC-IV, and MC-MED can be found in Table 4. The list of all features are in tables 6, 5 7, and 8.

A.2 Code, Benchmark Models, and Training Details

Sample data and code implementations can be found here: https://anonymous.4open.science/r/Multi-Modal-Graph-A1BC/README.md. The repository includes implementations for all baseline and proposed models, with a provided `requirements.txt` specifying package dependencies and versions.

We evaluate our model against two groups of baselines for both single-risk and competing-risk survival analysis tasks. The first group comprises classical and deep survival models: Cox Proportional Hazards (Cox PH), DeepSurv, DeepHit, Dynamic-DeepHit, and DySurv. The second group includes graph-based survival models: GCN, GAT, GraphSAGE, TodyNet, DynaGraph, MedGNN, and MM-STGNN. For MM-STGNN, we replace the original imaging modality with embeddings from radiographic reports to match our input setup and adapt it with a competing-risk loss function. The same adaptation is made for MedGNN. For TodyNet and DynaGraph, we retain only the time-series and/or static modalities as supported by each dataset.

All models were implemented in PyTorch and trained on a single NVIDIA V100 GPU with 50 GB RAM. Data loading and training were fully reproducible using fixed seeds: 42, 1992, 1709, 250, and

Table 4: Summary of demographics and clinical variables across three datasets: eICU, MIMIC-IV, and MC-MED. MIMIC-IV was used exclusively as an external validation set.

Attributes	**eICU (N = 82,155)**	**MIMIC-IV (N = 71,935)**	**MC-MED (N = 23,128)**
Age (mean ± SD)	67.2 ± 12.4	74.1 ± 13.4	58.3 ± 18.9
Sex (male)	64.5%	57.1%	46.7%
Ethnicity (Caucasian)	77.3%	71.1%	40.2%
Ethnicity (African American)	10.6%	8.0%	6.5%
Ethnicity (Hispanic)	3.7%	3.1%	18.6%
Ethnicity (Asian)	1.6%	2.0%	13.3%
Lactate (mmol/L)	2.5 ± 2.3	2.0 ± 1.5	2.4 ± 2.1
SBP (mmHg)	120.0 ± 16.3	126.3 ± 18.8	129.2 ± 17.2
Glucose (mg/dL)	147.3 ± 56.7	136.5 ± 49.3	138.4 ± 51.0
WBC ($\times 10^9$/L)	15.1 ± 9.3	10.6 ± 7.4	9.9 ± 6.1
RDW (%)	15.0 ± 2.0	14.4 ± 2.1	14.1 ± 2.0
Urea Nitrogen (mg/dL)	22.8 ± 13.4	22.8 ± 17.0	24.1 ± 16.8
Bicarbonate (mmol/L)	24.8 ± 4.4	23.3 ± 3.1	23.9 ± 4.2
Mortality	12.0%	9.7%	2.3% (ICU admission)

Table 5: Features extracted from the MIMIC-IV database. The features include demographic data collected for all patients, ICU unit-specific information like the type of unit, hospital information, vital signs, and biochemical measurements.

Static Variables			
Feature	*Type*	*Feature*	*Type*
Sex	binary	Admission Type	categorical
Age	integer	Insurance	categorical
ICU Type	categorical	Ethnicity	categorical
Time-series Variables			
Feature	*Type*	*Feature*	*Type*
Anion Gap	continuous	WBC	continuous
Weight	continuous	Temperature	continuous
SBP	continuous	DBP	continuous
Sodium	continuous	Respiratory Rate	continuous
RBC	continuous	Prothrombin Time PT	continuous
Prothrombin Time INR	continuous	Potassium	continuous
Platelets	categorical	Phosphorous	continuous
Phosphate	continuous	Partial Thromboplastin Time	continuous
Oxygen Saturation	continuous	MCGC	continuous
Magnesium	continuous	Hemoglobin	continuous
Hematocrit	continuous	Heart Rate	continuous
Glucose	continuous	Chloride	continuous
Creatinine	continuous	Calcium	continuous
BUN	continuous	Bicarbonate	continuous
Vent	binary	Vaso	binary
Adenosine	binary	Dobutamine	binary
Dopamine	binary	Epinephrine	binary
Isuprel	binary	Milrinone	binary
Norepinephrine	binary	Phenylepinephrine	binary
Vasopressin	binary	Colloid	binary
Crystalloid	binary	Intervention Duration	binary

213. Validation performance was used for early stopping and final model selection.

The Adam optimizer was used throughout, with a default learning rate of 0.001 unless otherwise noted. For benchmark models, we

Table 6: Features extracted from the eICU database. The features include demographic data collected for all patients, ICU unit-specific information like type and number of beds, hospital information like regional location and teaching status, vital signs including respiratory rate and blood pressure, and biochemical measurements including troponin and levels of potassium and protein in the blood.

Feature	**Type**	**Feature**	**Type**
Sex	binary	Unit Stay Type	categorical
Age	integer	Num Beds Category	categorical
Height	continuous	Region	categorical
Weight	continuous	Teaching Status	binary
Ethnicity	categorical	Physician Speciality	categorical
Unit Type	categorical	Unit Type	categorical
Unit Admit Source	categorical	Mechanical Ventilation	binary
Unit Visit Number	categorical		
Time-series Variables			
Feature	*Type*	*Feature*	*Type*
		Base Excess	continuous
-basos	continuous	FiO2	continuous
-eos	continuous	HCO3	continuous
-monos	continuous	Hct	continuous
-polys	continuous	Hgb	continuous
ALT	continuous	MCH	continuous
AST	continuous	MCHC	continuous
BUN	continuous	MCV	continuous
O2 Sat (%)	continuous	MPV	continuous
PT-INR	continuous	PT	continuous
RBC	continuous	PTT	continuous
RDW	continuous	WBC	continuous
Alkaline ph.	continuous	Albumin	continuous
Bedside Glucose	continuous	Anion Gap	continuous
Calcium	continuous	Bicarbonate	continuous
Creatinine	continuous	Glucose	continuous
Lactate	continuous	Magnesium	continuous
pH	continuous	paCO2	continuous
paO2	continuous	Phosphate	continuous
Platelets	continuous	Potassium	continuous
Sodium	continuous	Bilirubin	continuous
Protein	continuous	Troponin - I	continuous
Urinary s. Gravity	continuous	mean BP	continuous
SBP	continuous	DBP	continuous

used a batch size of 32 for eICU, MIMIC-IV, and MC-MED datasets. For our model, we searched over {8, 13, 32, 64}, with 32 found optimal. Epochs were set to 10 for eICU, 11 for MIMIC-IV, and 10 for MC-MED. Each full run on MIMIC-IV takes approximately 23 minutes.

We performed grid search on MIMIC-IV to tune hyperparameters. Details of hyperparameter ranges and selected values are given below in Table 9.

B Metrics

In this section, we switch sample notation from superscripts to subscripts, i.e., $\mathbf{X}^i$ becomes $\mathbf{X}_i$ for patient i. Since our model predicts full event-time distributions under competing risks and right-censoring, rather than single-time binary labels, classical classification metrics are insufficient. We rely on metrics designed specifically for survival analysis, including both ranking-based and calibration-based scores.

The most widely used metric in survival modelling is the *concordance index* (C_{ind}), which estimates the probability that, for a randomly selected comparable pair of patients, the model correctly ranks their relative event times [15]. While straightforward under proportional hazards assumptions, where risk rankings are time-invariant, we instead use the time-dependent extension $C_{\text{ind}}^{\text{td}}$ that

Table 7: Features extracted from the MC-MED database. The features include demographic data, triage variables, vital signs, biochemical measurements, as well as ICD diagnosis codes and radiography embeddings.

Static Variables			
Feature	*Type*	*Feature*	*Type*
Sex	binary	Ethnicity	categorical
Age	integer	Triage Acuity Score	ordinal
Time-series Variables			
Feature	*Type*	*Feature*	*Type*
HR	continuous	RR	continuous
SBP	continuous	DBP	continuous
SpO2	continuous	Temp	continuous
Pain Score	continuous	Perfusion Index	continuous
1min HRV	continuous	5min HRV	continuous
BUN	continuous	Creatinine	continuous
Sodium	continuous	Potassium	continuous
Chloride	continuous	Bicarbonate	continuous
Calcium	continuous	Corrected Calcium	continuous
Albumin	continuous	Globulin	continuous
Total Protein	continuous	Bilirubin (Total)	continuous
AST (SGOT)	continuous	ALT (SGPT)	continuous
Alkaline Phosphatase	continuous	Anion Gap	continuous
Hemoglobin	continuous	Hematocrit	continuous
MCV	continuous	MCH	continuous
MCHC	continuous	RBC	continuous
WBC	continuous	Platelet Count	continuous
EGFR (no race)	continuous	Glucose	continuous
Triage Vital Signs (HR, RR, SBP, DBP, Temp, SpO2)	continuous	–	–
ICD Codes	multi-hot vectors (500 most frequent codes)		
Radiography Embeddings	Dense vector representations (768-dim embeddings)		

accommodates time-varying survival predictions [1]. Formally:

$$C_{\text{ind}}^{\text{td}} = \mathbb{P}\left\{\hat{F}(t_i \mid \mathbf{X}_i) > \hat{F}(t_i \mid \mathbf{X}_j) \mid t_i < t_j, \epsilon_i \neq \varnothing\right\} \tag{6}$$

where $\hat{F}(t \mid \mathbf{X})$ denotes the predicted cumulative incidence (for any event) at time t. Only uncensored events ($\epsilon_i \neq \varnothing$) are considered for the evaluation.

In the competing risks setting, each patient may experience one of multiple mutually exclusive event types. Therefore, concordance must be evaluated separately for each event type to assess how well the model ranks time-to-event predictions per cause. The cause-specific time-dependent concordance index for event ϵ is defined as:

$$C_{\text{ind},\epsilon}^{\text{td}} = \mathbb{P}\left\{\hat{F}_\epsilon(t_i \mid \mathbf{X}_i) > \hat{F}_\epsilon(t_i \mid \mathbf{X}_j) \mid t_i < t_j, \epsilon_i = \epsilon\right\} \tag{7}$$

where $\hat{F}_\epsilon(t \mid \mathbf{X})$ is the predicted cumulative incidence function (CIF) for event type ϵ. Only pairs where patient i experienced event ϵ before patient j are considered.

We report both per-cause $C_{\text{ind},\epsilon}^{\text{td}}$ and the macro-averaged concordance across all non-censoring events:

$$\bar{C}_{\text{ind}}^{\text{td}} = \frac{1}{E}\sum_{\epsilon=1}^{E} C_{\text{ind},\epsilon}^{\text{td}} \tag{8}$$

where E is the total number of event types.

The *Brier Score* measures the squared difference between the predicted and true survival probabilities at a given timepoint, similar to mean squared error, and ranges from 0 (best) to 1 (worst) [7, 14]. To adjust for censoring, we use the inverse probability of censoring weighted (IPCW) Brier Score:

$$\text{BS}_{\text{IPCW}}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{S}(t \mid \mathbf{X}_i)^2 \mathbb{1}\{t_i \leq t, \epsilon_i \neq \varnothing\}}{\hat{G}(t_i^-)} + \frac{(1 - \hat{S}(t \mid \mathbf{X}_i))^2 \mathbb{1}\{t_i > t\}}{\hat{G}(t)}\right] \tag{9}$$

where $\hat{S}(t \mid \mathbf{X}_i)$ is the predicted survival probability, and $\hat{G}(t)$ is the Kaplan–Meier estimate of the censoring distribution.

The *Integrated Brier Score (IBS)* averages the Brier Score over time:

$$\text{IBS} = \frac{1}{t_{\max}}\int_0^{t_{\max}} \text{BS}_{\text{IPCW}}(t)\, dt \tag{10}$$

where $t_{\max}$ is the maximum observed time.

Table 8: Summary of features used in the PBC2 and SUPPORT datasets. PBC2 includes liver disease-specific biomarkers and clinical indicators, while SUPPORT comprises general clinical, physiological, and comorbidity-related variables for hospitalized patients.

Category	**PBC2**		**SUPPORT**	
	Feature	*Type*	*Feature*	*Type*
Demographics	Age	continuous	Age	continuous
	Sex	binary	Sex	binary
			Race	categorical
			DNR Status (Day 1)	binary
Clinical	Edema	categorical	Coma Score	ordinal
	Ascites	binary	Cancer	binary
	Hepatomegaly	binary	CHF	binary
	Spiders (vascular lesions)	binary	Cirrhosis	binary
Biomarkers / Labs	Albumin	continuous	Serum Albumin	continuous
	Bilirubin	continuous	Serum Sodium	continuous
	ALP	continuous	Serum Potassium	continuous
	AST (SGOT)	continuous	Serum Creatinine	continuous
	Prothrombin Time	continuous	Hematocrit	continuous
	Platelets	continuous	WBC Count	continuous
			PaO_2	continuous
Vitals			Systolic BP	continuous
			Heart Rate	continuous
			Respiratory Rate	continuous
			Temperature	continuous
			Glasgow Coma Scale	continuous

Table 9: Hyperparameter search ranges for all models. Bold values indicate optimal parameters found via grid search on the validation set. All models were trained using the Adam optimiser.

Model	**Batch Size**	**Learning Rate**	**Dropout**	**Graph Layers**	**MLP Layers**
Our Model	8, 16, **32**, 64	0.01, 0.001, **0.0001**	0.5, **0.7**, 0.9	**2**, 3, 4	**2**, 4, 6
DeepSurv	32, **64**, 128	0.01, 0.001, **0.0001**	0.5, **0.7**, 0.9	–	2, **4**, 6
DeepHit	32, 64, **128**	0.01, 0.001, **0.0001**	0.5, **0.7**, 0.9	–	2, **4**, 6
Dynamic-DeepHit	32, 64, **128**	0.01, 0.001, **0.0001**	0.5, **0.7**, 0.9	–	2, **4**, 6
DySurv	**32**, 64, 128	**0.001**, 0.0001	0.5, **0.7**, 0.9	–	2, **4**, 6
GCN	32, 64, **128**	0.01, 0.001, **0.0001**	**0.5**, 0.7, 0.9	2, **3**, 4	2, **4**, 6
GAT	32, 64, **128**	0.01, 0.001, **0.0001**	**0.5**, 0.7, 0.9	2, **3**, 4	2, **4**, 6
GraphSAGE	32, 64, **128**	**0.001**, 0.0001	0.5, **0.7**, 0.9	2, **3**, 4	2, **4**, 6
TodyNet	32, **64**, 128	**0.001**, 0.0001	0.5, **0.7**, 0.9	**2**, 3, 4	2, **4**, 6
DynaGraph	32, 64, **128**	**0.0001**, 0.001	0.5, **0.7**, 0.9	**2**, 3, 4	2, **4**, 6
MedGNN	32, 64, **128**	**0.0001**, 0.001	0.5, **0.7**, 0.9	2, 3, **4**	2, **4**, 6
MM-STGNN	**32**, 64, 128	**0.0001**, 0.001	0.5, **0.7**, 0.9	**2**, 3, 4	2, **4**, 6

Finally, we assess probabilistic calibration using the IPCW-adjusted *binomial log-likelihood* (BLL), which measures the accuracy of predicted survival probabilities as probabilistic forecasts:

$$\mathrm{BLL}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\log(1-\hat{S}(t \mid \mathbf{X}_i))\mathbb{1}\{t_i \le t, \epsilon_i \ne \varnothing\}}{\hat{G}(t_i)} + \frac{\log(\hat{S}(t \mid \mathbf{X}_i))\mathbb{1}\{t_i > t\}}{\hat{G}(t)}\right] \tag{11}$$

Its integrated version, the *Integrated Binomial Log-Likelihood (IBLL)*, is:

$$\mathrm{IBLL} = \frac{1}{t_{\max}}\int_0^{t_{\max}} \mathrm{BLL}(t)\, dt \tag{12}$$

All metrics are computed on the held-out test set. Time integrals are approximated using numerical integration over 100 uniformly spaced timepoints, consistent with prior work [15]. Together, $C^{\mathrm{td}}_{\mathrm{ind}}$, IBS, and IBLL provide complementary perspectives on model performance, evaluating both temporal discrimination and calibration under competing risks.

C Interpretability

To enable transparency in predictions, we integrate a pseudo-attention mechanism during graph construction. Each graph at time slice s is associated with a learnable interpretability matrix $I_s \in \mathbb{R}^{d \times d}$, which captures the relative contribution of nodes and edges to the loss. These matrices are initialised uniformly and updated via backpropagation based on their influence on the model's output.

For a node v, we define its importance score by combining its individual gradient saliency and the average importance of its connected edges:

$$I_v = \alpha \cdot \|\nabla_{h_v} L\| + (1 - \alpha) \cdot \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} \|\nabla_{e_{v,u}} L\|, \tag{13}$$

where $\alpha \in [0, 1]$ balances node and edge contributions, h_v is the embedding of node v, $e_{v,u}$ is the edge from v to u, and L is the task loss. This formulation allows us to attribute model behaviour to individual clinical variables and their temporal interactions.

These interpretability matrices are computed independently for each modality (e.g., vitals, demographics, ICD codes, radiography) and later fused, preserving hierarchical structure. The fused matrix informs both graph construction and attention-weighted pooling, offering a consistent mechanism for identifying influential features over time.

In contrast to traditional post-hoc methods, our integrated approach ensures interpretability is aligned with the model's learned representations and dynamically adapts across training epochs. This is critical in healthcare, where understanding temporal feature salience can support trust and accountability in clinical decision-making.

D Convergence of the Hierarchical Interpretability Matrix

Definitions and Assumptions

Let $G_t = (V, E_t, I_t)$ denote the graph at epoch t, where:

- V is the set of nodes (e.g., patient-specific features across modalities).
- E_t is the set of edges determined by temporal and modality-level relationships.
- I_t is the fused interpretability matrix at epoch t, derived from hierarchical attention across modalities and time.

Let $I_t = f(I_t^{\text{dyn}}, I_t^{\text{static}}, I_t^{\text{ICD}}, I_t^{\text{rad}})$ be the fused matrix at epoch t, computed from the attention matrices associated with each modality through a learnable aggregation function $f(\cdot)$.

Convergence Definition: The hierarchical interpretability matrix converges if:

$$\lim_{t \to \infty} \|I_t - I_{t-1}\|_F = 0,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Assumptions:

(1) Each modality-specific attention matrix $I_t^{(m)}$ is optimized via backpropagation with a differentiable loss function $\mathcal{L}$ and follows gradient descent updates.
(2) The fused interpretability matrix I_t is differentiable with respect to all $I_t^{(m)}$, and the fusion function $f(\cdot)$ is smooth (e.g., weighted sum or attention-based).
(3) The learning rate η_t satisfies $\eta_t \to 0$, $\sum_{t=1}^{\infty} \eta_t = \infty$, and $\sum_{t=1}^{\infty} \eta_t^2 < \infty$.

Proof of Convergence

We begin with the gradient update rule for each modality-specific interpretability matrix:

$$I_t^{(m)} = I_{t-1}^{(m)} - \eta_t \nabla_{I^{(m)}} \mathcal{L}(I_{t-1}^{(m)}).$$

Assuming $f(\cdot)$ is differentiable and linear or Lipschitz-continuous, the fused matrix evolves as:

$$I_t = f(I_t^{\text{dyn}}, I_t^{\text{static}}, I_t^{\text{ICD}}, I_t^{\text{rad}}).$$

Applying the multivariate Taylor expansion of $\mathcal{L}$ at $I_{t-1}^{(m)}$, for each modality:

$$\mathcal{L}(I_t^{(m)}) \approx \mathcal{L}(I_{t-1}^{(m)}) - \eta_t \|\nabla_{I^{(m)}} \mathcal{L}(I_{t-1}^{(m)})\|_F^2 + \frac{1}{2} \eta_t^2 \|H^{(m)}\|_F \|\nabla_{I^{(m)}} \mathcal{L}(I_{t-1}^{(m)})\|_F^2,$$

where $H^{(m)}$ is the Hessian of $\mathcal{L}$ with respect to $I^{(m)}$. Assuming bounded Hessians and the learning rate conditions, the second-order term vanishes faster, so:

$$\lim_{t \to \infty} \|\nabla_{I^{(m)}} \mathcal{L}(I_{t-1}^{(m)})\|_F = 0.$$

Since each modality-specific matrix $I_t^{(m)}$ converges and $f(\cdot)$ is continuous, their composition $I_t = f(I_t^{(m)})$ also converges. Hence:

$$\lim_{t \to \infty} \|I_t - I_{t-1}\|_F = 0.$$

Conclusion: Under standard assumptions on smoothness of the loss function and learning rates, the fused interpretability matrix constructed through hierarchical attention converges. This ensures the interpretability weights used in the model stabilise, making the model's explanations consistent and trustworthy over training epochs.

E Graph Learning with GIN

Continuing from the Methods, we first construct a fused graph representation for each sample i by concatenating the adjacency matrix $A^{(i)}$ with the corresponding modality-level interpretability matrix $I^{(i)}$:

$$G^{(i)} = \left(A^{(i)} \parallel I^{(i)}\right) \tag{14}$$

where $\parallel$ denotes concatenation along the feature dimension. This fusion integrates both structural and attention-based relational information across modalities and time.

To extract higher-order node representations from these fused graphs, we employ a multi-layer Graph Isomorphism Network (GIN), which has been shown to be effective at capturing complex node interactions and topological patterns [26]. The GIN layers iteratively update node embeddings through neighborhood aggregation, capturing temporal and multimodal dependencies in the graph structure. Formally, for a k-layer GIN, the node embeddings are updated as:

$$h_v^{(k)} = \mathrm{MLP}^{(k)} \left((1 + \epsilon^{(k)}) \cdot h_v^{(k-1)} + \sum_{u \in \mathcal{N}(v)} h_u^{(k-1)} \right) \tag{15}$$

where $h_v^{(k)}$ is the embedding of node v at layer k, $\mathcal{N}(v)$ denotes the neighbors of node v, $\epsilon^{(k)}$ is a learnable scalar, and $\text{MLP}^{(k)}$ is a multi-layer perceptron specific to the k-th layer.

The final node embeddings $h^{(K)}$ from the last GIN layer are then passed into the temporal pooling module described in Section F, which compresses these embeddings hierarchically while preserving essential temporal dynamics. This pipeline allows the model to produce rich, temporally-aware, and topologically-grounded representations for downstream competing risks survival prediction.

F Temporal Pooling

After obtaining spatio-temporal embeddings from the hierarchical GIN module, we apply a learnable temporal pooling mechanism to compress the temporal graph representation while preserving salient temporal and relational structures across time. Instead of flattening the node-time embeddings, which risks losing meaningful temporal dependencies, we adopt a hierarchical 2D convolution-based pooling strategy inspired by [21]. This approach enables the model to learn soft assignments of temporal nodes into clusters, producing compact graph representations suitable for downstream survival prediction.

Formally, given an input tensor $X^l \in \mathbb{R}^{N^l \times d}$ at the l-th pooling layer (where N^l is the number of temporal graph nodes and d the embedding dimension), we apply a 2D convolution over the temporal axis (treated as channels), producing an output $X^{l+1} \in \mathbb{R}^{N^{l+1} \times d}$:

$$X^{l+1} = \sum_{j=0}^{N^l - 1} W(N^{l+1}, j) \star X^l + b(N^{l+1})$$

where W are learnable convolution weights, b is a bias term, and $\star$ denotes cross-correlation. This operation yields new cluster-level node embeddings while reducing the node count from N^l to N^{l+1}.

To propagate structural information through the hierarchy, we reconstruct lower-level adjacency matrices based on the learned soft clustering. Let $W^l \in \mathbb{R}^{N^{l+1} \times N^l \times 1 \times k}$ be the 2D convolutional weights reshaped into a matrix $M^l \in \mathbb{R}^{N^{l+1} \times N^l}$ using a learnable vector $V^l \in \mathbb{R}^{1 \times k}$ such that $M^l = W^l \cdot V^l$. The coarsened adjacency matrix is then computed as:

$$A^{l+1} = M^l A^l M^{l\top} \in \mathbb{R}^{N^{l+1} \times N^{l+1}}$$

This allows the model to maintain a graph structure at each level of abstraction, where A^{l+1} captures the weighted connectivity between temporal clusters. Crucially, both node and edge representations evolve jointly and hierarchically, enabling end-to-end optimisation of both temporal resolution and relational importance. The final pooled representations X^{l+1} and A^{l+1} are passed to cause-specific multi-layer perceptions for competing risk survival prediction.

G Full Model Evaluation Results

We provide a comprehensive assessment of all evaluated models across three standard metrics in survival analysis: cause-specific time-dependent concordance index (C-index), Integrated Brier Score (IBS), and Integrated Binomial Log-Likelihood (IBLL). These metrics respectively measure the discriminative performance, calibration, and probabilistic accuracy of survival models.

Our proposed model achieves consistently superior performance across all five datasets (MIMIC-IV, eICU, PBC2, MC-MED, SUPPORT) when compared to both traditional survival methods (e.g., Cox PH), temporal deep learning models (e.g., DeepHit, Dynamic-DeepHit, DySurv), and graph-based approaches (e.g., GCN, GAT, GraphSAGE, MM-STGNN, MedGNN, DynaGraph, TodyNet). This robustness across diverse settings underscores the effectiveness of our model's dynamic, multi-modal graph architecture with hierarchical attention and temporal pooling.

Figure 4a shows a spider plot of concordance scores across datasets, clearly illustrating our model's consistent edge in discriminative power. Supplementary Figures 4b and 4c provide the corresponding plots for IBS and IBLL metrics, revealing strong calibration and generalisation performance across datasets as well.

In all metrics, performance gains are especially pronounced on complex, multi-modal datasets like MIMIC-IV and MC-MED, where our model benefits from its ability to jointly represent static, temporal, diagnostic, and radiographic modalities. These improvements are also evident in smaller datasets like PBC2 and SUPPORT, demonstrating the model's adaptability to varying cohort sizes and clinical settings.

H Calibration Plots

Results of calibrations can be seen in Figure 5.

References

[1] Laura Antolini, Patrizia Boracchi, and Elia Biganzoli. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine* 24, 24 (2005), 3927–3944.

[2] Peter C Austin, Douglas S Lee, and Jason P Fine. 2016. Introduction to the analysis of survival data in the presence of competing risks. *Circulation* 133, 6 (2016), 601–609.

[3] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems* 33 (2020), 17804–17815.

[4] Heloísa Oss Boll, Ali Amirahmadi, Mirfarid Musavian Ghazani, Wagner Ourique de Morais, Edison Pignaton de Freitas, Amira Soliman, Farzaneh Etminani, Stefan Byttner, and Mariana Recamonde-Mendoza. 2024. Graph neural networks for clinical risk prediction based on electronic health records: A survey. *J. Biomed. Informatics* 151 (2024), 104616.

[5] Juan Du, Dajian Zeng, Zhao Li, Jingxuan Liu, Mingqi Lv, Ling Chen, Dan Zhang, and Shouling Ji. 2022. An interpretable outcome prediction model based on electronic health records and hierarchical attention. *International Journal of Intelligent Systems* 37, 6 (2022), 3460–3479.

[6] Wei Fan, Jingru Fei, Dingyu Guo, Kun Yi, Xiaozhuang Song, Haolong Xiang, Hangting Ye, and Min Li. 2025. MedGNN: Towards Multi-resolution Spatiotemporal Graph Learning for Medical Time Series Classification. *arXiv preprint arXiv:2502.04515* (2025).

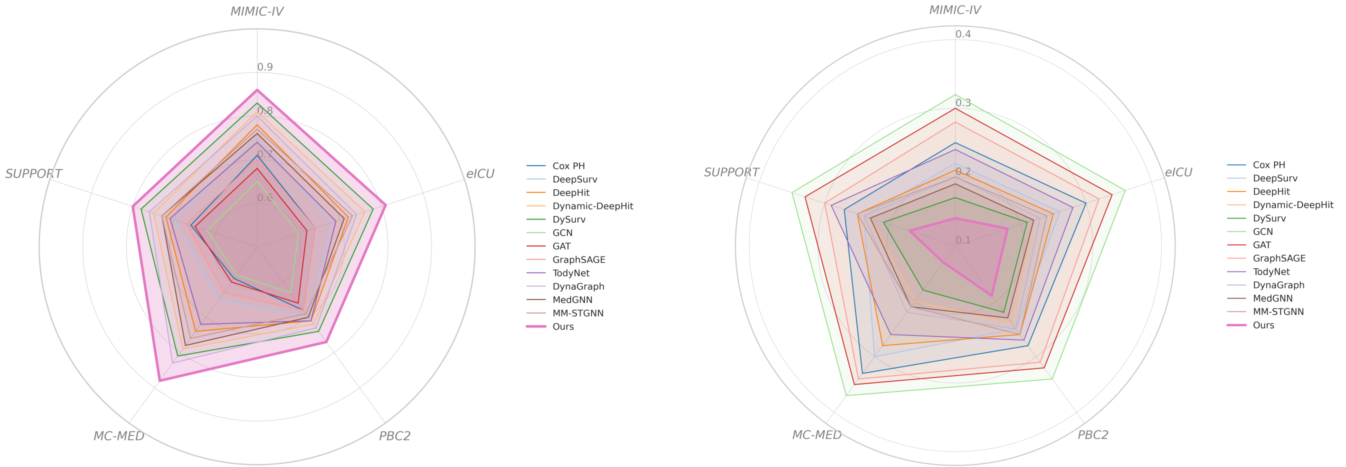
[7] Stephane Fotso. 2018. Deep neural networks for survival analysis based on a multi-task framework. *arXiv preprint arXiv:1801.05512* (2018).

[8] Rebecca Giddings, Anabel Joseph, Thomas Callender, Sam M Janes, Mihaela Van der Schaar, Jessica Sheringham, and Neal Navani. 2024. Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. *The Lancet Digital Health* 6, 2 (2024), e131–e144.

[9] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

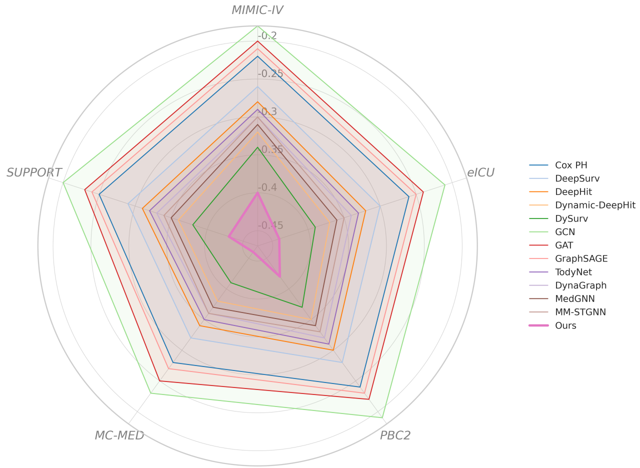
[10] Na Hong, Chun Liu, Jianwei Gao, Lin Han, Fengxiang Chang, Mengchun Gong, and Longxiang Su. 2022. State of the art of machine learning–enabled clinical decision support in intensive care units: literature review. *JMIR medical informatics* 10, 3 (2022), e28781.

[11] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee, and Matthew P Lungren. 2020. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine* 3, 1 (2020), 136.



(a) Cause-specific concordance index (higher is better).

(b) Integrated Brier Score (IBS; lower is better).



(c) Integrated binomial log-likelihood (IBLL; lower is better).

Figure 4: Performance of all models across five datasets (MIMIC-IV, eICU, PBC2, MC-MED, SUPPORT) measured using (a) cause-specific concordance index for the most common cause, (b) integrated Brier score (IBS), and (c) integrated binomial log-likelihood (IBLL). Our model consistently achieves the best performance across metrics and datasets.

[12] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. 2018. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology* 18, 1 (2018), 1–12.

[13] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[14] Håvard Kvamme and Ørnulf Borgan. 2019. The brier score under administrative censoring: Problems and solutions. *arXiv preprint arXiv:1912.08581* (2019).

[15] Håvard Kvamme, Ørnulf Borgan, and Ida Scheel. 2019. Time-to-event prediction with neural networks and Cox regression. *Journal of machine learning research* 20, 129 (2019), 1–30.

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

[17] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. 2018. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[18] Jianing Li, Thomas H Scheike, and Mei-Jie Zhang. 2015. Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals. *Lifetime data analysis* 21, 2 (2015), 197–217.

[19] Yikuan Li, Ramsey M Wehbe, Faraz S Ahmad, Hanyin Wang, and Yuan Luo. 2022. Clinical-longformer and clinical-bigbird: Transformers for long clinical sequences. *arXiv preprint arXiv:2201.11838* (2022).

[20] Yikuan Li, Ramsey M Wehbe, Faraz S Ahmad, Hanyin Wang, and Yuan Luo. 2023. A comparative study of pretrained language models for long clinical text. *Journal of the American Medical Informatics Association* 30, 2 (2023), 340–347.

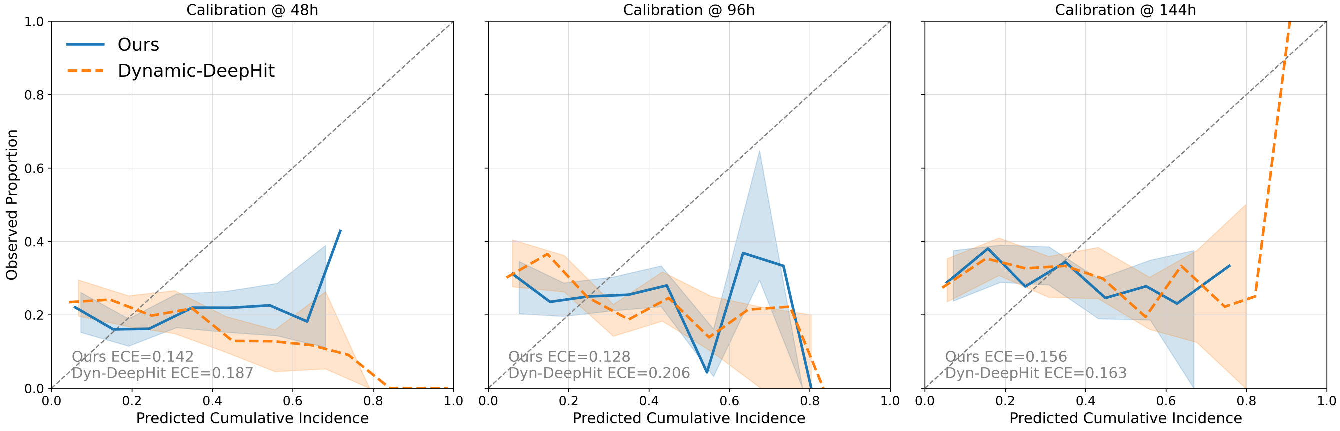
[21] Huaiyuan Liu, Xianzhang Liu, Donghua Yang, Zhiyu Liang, Hongzhi Wang, Yong Cui, and Jun Gu. 2023. TodyNet: Temporal Dynamic Graph Neural Network for Multivariate Time Series Classification. *arXiv preprint arXiv:2304.05078* (2023).

[22] Huaiyuan Liu, Donghua Yang, Xianzhang Liu, Xinglei Chen, Zhiyu Liang, Hongzhi Wang, Yong Cui, and Jun Gu. 2024. Todynet: temporal dynamic graph neural network for multivariate time series classification. *Information Sciences* 677 (2024), 120914.

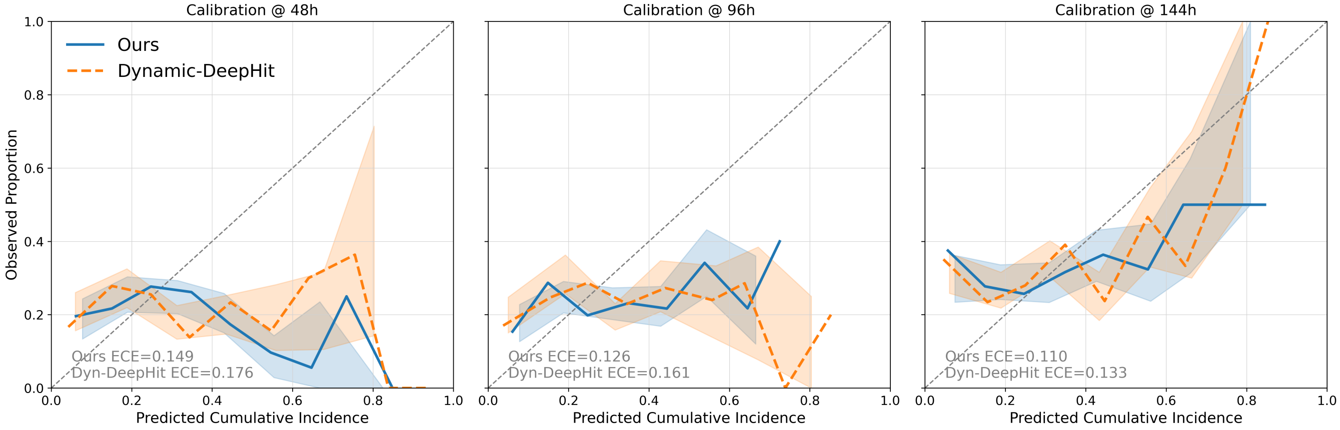
[23] Torben Martinussen. 2022. Causality and the Cox regression model. *Annual Review of Statistics and Its Application* 9, 1 (2022), 249–259.

[24] Munib Mesinovic, Soheila Molaei, Peter Watkinson, and Tingting Zhu. 2025. DynaGraph: Interpretable Multi-Label Prediction from EHRs via Dynamic Graph Learning and Contrastive Augmentation. *arXiv preprint arXiv:2503.22257* (2025).

[25] Munib Mesinovic, Peter Watkinson, and Tingting Zhu. 2024. DySurv: dynamic deep learning model for survival analysis with conditional variational inference. *Journal of the American Medical Informatics Association* (2024), ocae271.



(a) Calibration curves on MIMIC-IV dataset (ICU mortality at 48h, 96h, 144h).



(b) Calibration curves on eICU dataset (ICU mortality at 48h, 96h, 144h).

Figure 5: Calibration performance of our model versus Dynamic-DeepHit across multiple horizons and datasets. Panels (a) and (b) show results for MIMIC-IV and eICU, respectively. Each panel depicts calibration curves at 48h, 96h, and 144h prediction horizons. Curves are LOWESS-smoothed, and shaded regions indicate 80% bootstrap confidence intervals. Expected Calibration Error (ECE) is reported within each subplot. our model consistently yields better-aligned curves and lower ECE values, demonstrating improved calibration across ICU datasets.

[26] Ciyuan Peng, Jiayuan He, and Feng Xia. 2024. Learning on multimodal graphs: A survey. *arXiv preprint arXiv:2402.05322* (2024).

[27] Jiajun Qiu, Yao Hu, Li Li, Abdullah Mesut Erzurumluoglu, Ingrid Braenne, Charles Whitehurst, Jochen Schmitz, Jatin Arora, Boris Alexander Bartholdy, Shrey Gandhi, et al. 2025. Deep representation learning for clustering longitudinal survival data from electronic health records. *Nature Communications* 16, 1 (2025), 2534.

[28] Md Mahmudur Rahman, Koji Matsuo, Shinya Matsuzaki, and Sanjay Purushotham. 2021. Deeppseudo: Pseudo value based deep learning models for competing risk analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 479–487.

[29] Emma Rocheteau, Pietro Liò, and Stephanie Hyland. 2021. Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit. In *Proceedings of the conference on health, inference, and learning*. 58–68.

[30] Jörg Schilcher, Alva Nilsson, Oliver Andlid, and Anders Eklund. 2024. Fusion of electronic health records and radiographic images for a multimodal deep learning prediction model of atypical femur fractures. *Computers in Biology and Medicine* 168 (2024), 107704.

[31] Alice S Tang, Sarah R Woldemariam, Silvia Miramontes, Beau Norgeot, Tomiko T Oskotsky, and Marina Sirota. 2024. Harnessing EHR data for health research. *Nature Medicine* 30, 7 (2024), 1847–1855.

[32] Siyi Tang, Amara Tariq, Jared A Dunnmon, Umesh Sharma, Praneetha Elugunti, Daniel L Rubin, Bhavik N Patel, and Imon Banerjee. 2023. Predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks. *IEEE Journal of Biomedical and Health Informatics* 27, 4 (2023), 2071–2082.

[33] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[34] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 753–763.

[35] Haoyan Xu, Ziheng Duan, Yueyang Wang, Jie Feng, Runjian Chen, Qianru Zhang, and Zhongbin Xu. 2021. Graph partitioning and graph neural network based hierarchical graph matching for graph similarity computation. *Neurocomputing* 439 (2021), 348–362.

[36] Yishu Xue and Elizabeth D Schifano. 2017. Diagnostics for the Cox model. *Communications for Statistical Applications & Methods* 24, 6 (2017).

[37] Daochen Zha, Kwei-Herng Lai, Kaixiong Zhou, and Xia Hu. 2022. Towards similarity-aware time-series classification. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. SIAM, 199–207.

[38] Xinlu Zhang, Shiyang Li, Zhiyu Chen, Xifeng Yan, and Linda Ruth Petzold. 2023. Improving medical predictions by irregular multimodal electronic health records modeling. In *International Conference on Machine Learning*. PMLR, 41300–41313.

[39] Zhongheng Zhang, Jaakko Reinikainen, Kazeem Adedayo Adeleke, Marcel E Pieterse, and Catharina GM Groothuis-Oudshoorn. 2018. Time-varying covariates and coefficients in Cox regression models. *Annals of translational medicine* 6, 7 (2018), 121.

[40] Zihan Zhang, Lei Shi, and Ding-Xuan Zhou. 2024. Classification with deep neural networks and logistic loss. *Journal of Machine Learning Research* 25, 125 (2024), 1–117.

[41] Weicheng Zhu and Narges Razavian. 2021. Variationally regularized graph-based representation learning for electronic health records. In *Proceedings of the conference on health, inference, and learning*. 1–13.