Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260182

Multimodal Graph-Based Model for Discrete-Time Survival Prediction in Liver Cancer

Hafsa AKEBLI[a] , Vincenzo DELLA MEA[a] and Kevin ROITERO[a]
[a] *University of Udine, Italy*
ORCiD ID: Hafsa Akebli https://orcid.org/0009-0002-3879-876X, Vincenzo Della Mea https://orcid.org/0000-0002-0144-3802, Kevin Roitero https://orcid.org/0000-0002-9191-3280

Abstract. Liver cancer is a leading cause of cancer mortality; hepatocellular carcinoma (HCC), its predominant form, requires accurate survival prediction to guide prognosis and treatment decisions. We propose a multimodal framework for discrete-time survival prediction using histopathological and clinical data from the TCGA-LIHC cohort. Patch-level features extracted from Whole-Slide Images using the UNI2-h foundation model were aggregated within each slide through a Graph Attention Network, while clinical variables were encoded with a sentence transformer. The resulting embeddings were fused through intermediate fusion strategies and processed by a multi-layer perceptron trained using a negative discrete-time log-likelihood loss to handle censored data. Across two clinically meaningful survival intervals, [0, 1) and [1, 5] years, multimodal fusion outperformed unimodal baselines, with concatenation achieving the highest overall performance (AUROC = 0.818). These findings highlight the complementary prognostic value of histology and clinical data for robust and interpretable survival modeling in HCC.

Keywords. Multimodal, Graph Neural Networks, Liver Cancer, Histopathology, Survival Prediction

1. Introduction

Liver cancer is the sixth most frequently diagnosed malignancy and the third leading cause of cancer-related deaths worldwide, with Hepatocellular Carcinoma (HCC) accounting for 75–85% of primary liver cancer cases [1]. Despite advances in diagnosis and therapy, HCC remains associated with poor prognosis, as most patients are diagnosed at intermediate or advanced stages when curative treatment options are limited [2].

Whole-Slide Images (WSIs) have enabled large-scale computational analysis of histopathology, allowing deep learning models to predict patient prognosis and survival outcomes [3]. However, histology alone cannot capture the full complexity of cancer progression. Genomic profiles, clinical data, and other imaging modalities (e.g., radiology) provide complementary information for survival prediction, motivating deep learning frameworks that integrate heterogeneous data sources for improved cancer prognosis [4,5], despite challenges arising from differences in data structure and scale. Discrete-time survival modeling provides a flexible alternative to traditional continuous time approaches by discretizing follow-up into clinically meaningful intervals, allowing

prognosis to be formulated as a multi-label classification task that naturally handles censored data and captures time-dependent risk dynamics within neural network frameworks [6,7].

In this work, we propose a multimodal framework for discrete-time survival prediction in HCC, where follow-up time is discretized into two clinically meaningful intervals. Patch-level features extracted from WSIs using the foundation model UNI2-h [8] are aggregated with a Graph Attention Network (GAT) [9] to obtain slide-level representations. Clinical data are encoded with a sentence transformer. The resulting modality specific embeddings are combined through an intermediate fusion module and processed by a Multi-Layer Perceptron (MLP) trained with a discrete-time log-likelihood loss. Experiments on the TCGA-LIHC cohort demonstrate that multimodal integration consistently outperforms unimodal models, highlighting the complementary prognostic value of histological and clinical information for liver cancer survival prediction.

2. Materials and Methods

2.1. Dataset

We used multimodal data from the TCGA-LIHC dataset, comprising WSIs and tabular clinical information for patients diagnosed with HCC. Each patient has one to three diagnostic WSIs acquired at 40 × magnification; when multiple slides are available, we retain the one with the largest tissue area. After filtering, 358 patients remain, including 232 censored and 126 uncensored cases (censoring rate: 64.8%). Clinical variables include demographic, diagnostic, treatment and pathological factors (e.g., age, gender, race/ethnicity, morphology, tumor status, primary diagnosis, AJCC pathologic T/N/M, and prior/synchronous malignancy). Clinical data were cleaned and preprocessed.

2.2. Problem Formulation

We formulate survival prediction as a discrete-time multi-label binary classification task, where for each patient the model predicts event occurrence (death) within two clinically motivated intervals, Interval 1: [0, 1) year and Interval 2: [1, 5] years. These intervals were defined to concentrate observed events in our cohort. Label counts per interval are reported in Table 1. Each patient i is represented by a feature vector $X_i \in \mathbb{R}^d$. For each interval $j \in \{1, 2\}$, define the binary label y_{ij} as:

$$y_{ij} = \begin{cases} 1, & \text{if patient } i \text{ dies within interval } j \\ 0, & \text{if patient } i \text{ is alive beyond interval } j \\ \text{excluded}, & \text{if patient } i \text{ is censored within interval } j \end{cases}$$

We set the interval boundaries at 1 and 5 years to align with established clinical milestones in HCC, where one-year survival reflects early mortality and five-year survival indicates long-term remission [10,11]. The Kaplan–Meier curve of our cohort (Fig. 1) shows that most deaths occur before five years, with only nine events thereafter, confirming the suitability of these thresholds. We adopt a discrete-time survival modeling framework [6,12], predicting event occurrence within these predefined intervals.

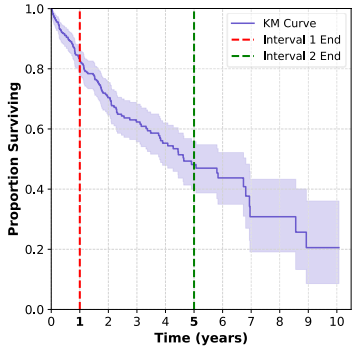

Figure 1. Kaplan–Meier survival curve of the TCGA-LIHC cohort ($N = 358$), with 1-year and 5-year thresholds defining Interval 1 and Interval 2.

Table 1. Label distribution across survival intervals, showing the number of deaths, survivors and censored patients.

Interval	Deaths	Survivors	Censored
Interval 1	55	251	52
Interval 2	62	40	149

2.3. Proposed Framework

The proposed framework, which integrates histology and clinical representations through a fusion module to predict discrete-time survival outcomes, is illustrated in Fig. 2.

2.3.1. Histology Graph Construction and Aggregation

Each WSI was tiled into non-overlapping 500 × 500 patches after background removal. Retained patches were embedded using UNI2-h foundation model [8], selected for its strong generalization in graph-based histology analysis [13]. Each WSI was represented as a patch-level graph, with nodes defined by the extracted features and edges defined by cosine k-nearest neighbors ($k = 5$). A four-layer GAT [9] aggregated node features via global mean pooling to obtain a 1024-dimensional slide representation.

2.3.2. Clinical Feature Extraction

The clinical dataset contained demographic, diagnostic, treatment and pathological variables. Each patient's record was converted into standardized text and encoded using the instruction-tuned *multilingual-e5-large-instruct* model [14]. Using the instruction "Given a patient's clinical profile, retrieve key information relevant to survival prediction.", we obtained 1024-dimensional embeddings for fusion.

2.3.3. Intermediate Fusion and Survival Prediction

Histology and clinical embeddings were fused using multiple intermediate fusion strategies. In concatenation, features were stacked along the feature dimension $X_{fusion} = X_{hist}|X_{clin}$. For scalar-weighted fusion, a learnable coefficient $\alpha \, \epsilon[0,1]$ balanced the modalities as $X_{fusion} = \alpha X_{hist} + (1-\alpha)X_{clin}$. Element-wise attention fusion used a sigmoid gate to assign per-dimension weights, $X_{fusion} = gX_{hist} + (1-g) \odot X_{clin}$, where $\odot$ denotes element-wise multiplication. The fused representation was passed to a three-layer

MLP trained with a negative discrete-time log-likelihood loss [15] to account for censored data and predict event occurrence within each time interval.

3. Experimental Results

3.1. Experimental Setup

Model performance was evaluated using AUROC, accuracy, and weighted F1-score per interval and overall. A stratified five-fold cross-validation study was conducted with z-score-normalized features. Embeddings were projected to 512 dimensions before fusion and trained with AdamW optimizer (lr = 10^{-4}, wd= 5×10^{-5}) for 600 epochs.

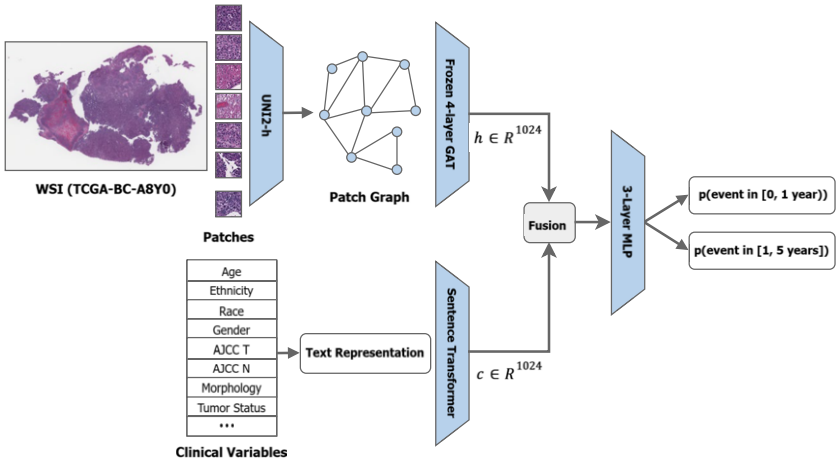

Figure 2. Overview of the proposed multimodal discrete-time survival framework: UNI2-h patch embeddings are aggregated by a frozen 4-layer GAT, clinical variables are encoded with a sentence transformer, and the two modality features are fused before a 3-layer MLP predicts event occurrence in [0, 1) year and [1, 5] years.

3.2. Multimodal Survival Performance

Results are summarized in Table 2. The clinical-only model was stronger than histology-only overall (AUROC = 0.778 vs. 0.669) and in Interval 2 (AUROC = 0.772 vs. 0.617), indicating that clinical variables better capture long-term outcomes. Conversely, histology was more informative for early survival, reaching higher Interval 1 F1 and accuracy (F1 = 0.772, Acc = 0.752) than clinical-only. At the overall level and in Interval 1, all multimodal fusion strategies outperformed unimodal baselines. Among them, concatenation consistently achieved the highest performance across all metrics and intervals (AUROC = 0.818), confirming its robustness. Scalar-weighted fusion matched its short-term accuracy (Acc = 0.784), while element-wise attention fusion reached comparable overall performance (AUROC = 0.798) with more balanced results across intervals. In Interval 2, the clinical-only model remained competitive, indicating that adaptive fusion may be less effective for long-term prediction. Performance was generally lower in this interval due to the higher proportion of censored patients. The interval-based formulation also enhances interpretability by distinguishing early from late mortality risks.

Table 2. Mean metrics across 5-fold cross-validation for discrete-time survival prediction at overall and interval levels.

Modality	Overall			Interval 1			Interval 2		
	AUROC	F1	Acc	AUROC	F1	Acc	AUROC	F1	Acc
Histology Only	0.669	0.691	0.682	0.721	0.772	0.752	0.617	0.610	0.611
Clinical Only	0.778	0.717	0.705	0.783	0.712	0.680	0.772	0.722	0.731
Element-wise Attention Fusion	0.798	0.742	0.736	0.816	0.773	0.755	0.781	0.711	0.718
Scalar-weighted Fusion	0.793	0.749	0.744	0.826	0.798	**0.784**	0.761	0.700	0.703
Concatenation Fusion	**0.818**	**0.777**	**0.771**	**0.848**	**0.805**	**0.784**	**0.788**	**0.749**	**0.759**

4. Conclusion and Directions

This work shows the promise of multimodal learning for survival prediction in oncology, where histology and clinical features offer complementary prognostic value. By modeling survival in discrete-time, our framework enables interpretable risk stratification that distinguishes early from long-term outcomes in HCC. Future extensions integrating pathology reports or genomic data could further enhance performance and interpretability.

Acknowledgements

This work was partially supported by the Next-Generation EU (PNRR: Piano Nazionale di Ripresa e Resilienza - Missione 4 Componente 2, D.M. 118/2023)

References

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2024;74(3):229-63. Available from: https://doi.org/10.3322/caac.21834.

[2] Yan Q, song Sun Y, An R, et al. Application and progress of the detection technologies in hepatocellular carcinoma. Genes & Diseases. 2023;10(5):1857-69.

[3] Lee M. Recent Advancements in Deep Learning Using Whole Slide Imaging for Cancer Prognosis. Bioengineering. 2023;10(8). Available from: https://www.mdpi.com/2306-5354/10/8/897.

[4] Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Scientific Reports. 2021 06;11(1):13505. Available from: https://doi.org/10.1038/s41598-021-92799-4.

[5] Yang H, Yang M, Chen J, et al. Multimodal deep learning approaches for precision oncology: a comprehensive review. Briefings in Bioinformatics. 2025 01;26(1):bbae699. Available from: https://doi.org/10.1093/bib/bbae699.

[6] Suresh K, Severn C, Ghosh D. Survival prediction models: an introduction to discrete-time modeling. BMC Medical Research Methodology. 2022;22(1):207. Available from: https://doi.org/10.1186/s12874-022-01679-6.

[7] Kvamme H, Borgan O. Continuous and Discrete-Time Survival Prediction with Neural Networks. Lifetime Data Analysis. 2021;27(4):710-36. Available from: https://doi.org/10.1007/s10985-021-09532-6.

[8] Chen RJ, Ding T, Lu MY, et al. Towards a General-Purpose Foundation Model for Computational Pathology. Nature Medicine. 2024.

[9] Veli˘ckovi´c P, Cucurull G, Casanova A, et al. Graph Attention Networks. International Conference on Learning Representations. 2018. Available from: https://openreview.net/forum?id=rJXMpikCZ

[10] Vogel A, Meyer T, Sapisochin G, et al. Hepatocellular carcinoma. Lancet. 2022;400(10360):1345-62.

[11] Galle PR, Forner A, Llovet JM, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. Journal of Hepatology. 2018;69(1):182-236.

[12] Tutz G, Schmid M. Modeling Discrete Time-to-Event Data. Cham: Springer; 2016. Available from: https://doi.org/10.1007/978-3-319-28158-2.

[13] Akebli H, Mea VD, Roitero K. Benchmarking Feature Extractors for Prostate Cancer Detection Using Graph Attention Networks: A Focus on Foundation Models. In: 2025 IEEE 13th International Conference on Healthcare Informatics (ICHI); 2025. p. 685-6.

[14] Wang L, Yang N, Huang X, et al. Multilingual E5 Text Embeddings: A Technical Report. arXiv preprint arXiv:240205672. 2024.

[15] Zadeh SG, Schmid M. Bias in Cross-Entropy-Based Training of Deep Survival Networks. IEEE Trans-actions on Pattern Analysis and Machine Intelligence. 2021;43(9):3126-37.