Medical Knowledge-Guided Fusion of Holistic Hospital Data for Tumor Survival Analysis

Jinquan Guan[1⋆], Qi Chen[3⋆], Zixi Zheng[1], Yongkang Cai[4], Zhiquan Huang[4], Jian Chen[1](✉), and Yutong Xie[2](✉)

[1] South China University of Technology, Guangzhou, China
[2] Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
[3] Australian Institute for Machine Learning, Adelaide University, Adelaide, Australia
[4] Sun Yat-sen Memorial Hospital, Guangzhou, China
ellachen@scut.edu.cn, yutong.xie678@gmail.com

Abstract. Effective survival analysis necessitates integrating holistic but heterogeneous hospital data, yet limited cohort sizes render purely data-driven fusion susceptible to learning spurious correlations. Existing methods fail to escape the data-driven paradigm as they rely on rigid architectural designs enforcing fixed interaction networks that limit modality scalability or incorporate medical knowledge as an additional input rather than leveraging it for explicit multimodal fusion guidance. To bridge these gaps, we propose the Medical Knowledge Guided Survival Analysis (MKGSurv) as a general multimodal framework. We introduce a Correlation-Aware Graph ATtention Network (CA-GAT) that utilizes medical priors derived from Large Language Models (LLMs) to steer cross-disciplinary interaction. We also design a Correlation-Guided Alignment (CGA) loss to regularize the latent space by explicitly aligning the learned features with quantified relevance score. Experiments and human studies on four TCGA cohorts and an in-house dataset show the effectiveness of MKGSurv. The code and model weights are publicly available at: https://github.com/guanjinquan/MKGSurv.

Keywords: Survival Analysis · Multimodal Fusion · Knowledge · LLM

1 Introduction

Tumor survival analysis is a crucial task in oncology, enabling precise risk stratification and personalized care [10]. Survival outcomes are determined by diverse factors spanning histopathology, genomic, clinical, and treatment data [12]. Specifically, genomics reflect molecular aberrations, histopathology reveals cellular morphology and tissue architecture, clinical data indicate physiological status, and treatment records define the initial intervention and resulting residual tumor status. Integrating these heterogeneous data facilitates a holistic understanding of tumor progression, enabling robust survival prediction [24].

⋆ J. Guan & Q. Chen contributed equally. Corresponding authors: J. Chen & Y. Xie.

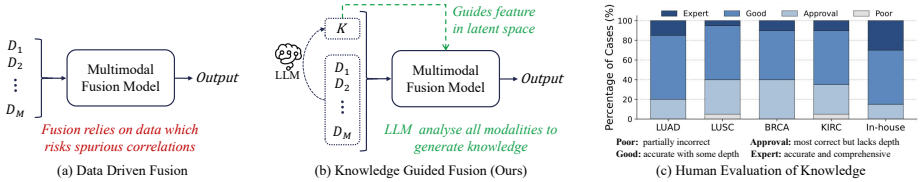

Fig. 1. (a) Traditional data-driven fusion. (b) Knowledge-guided fusion: leverages medical knowledge to derive explicit, quantitative cross-modal knowledge that steer multimodal integration. (c) Clinical evaluation by two clinicians of 4 and 20 years' experience on 100 randomly sampled patients (20 patients per cohort) confirms that the majority achieved approval or good ratings, while a subset was recognized as expert-level.

Deep learning has emerged as a promising method for modeling these complex survival patterns using medical data [14]. Early studies typically relied on a single modality, such as histopathology [27] or genomics [32]. However, unimodal modeling often fails to capture the full determinants of patient survival [2]. Many studies have explored multimodal survival analysis by integrating heterogeneous data to leverage complementary prognostic signals. Representative methods such as SurvPath [11], DIMAF [4], and MoME [30] focus on integrating histopathology with genomic profiles, whereas methods like CancerIDP [22] and AUTOSurv [13] combine genomic data with clinical variables. However, they are restricted to fixed modality combinations and therefore limit their scalability to the holistic and diverse data available in real-world clinical practice. To this end, multimodal fusion frameworks, ranging from general fusion methods such as FusionMamba [28] to survival-oriented models such as HGCN [8], HEALNet [7], and I2MoE [29], have improved cross-modal interaction to capture more informative predictive features from heterogeneous data.

Despite architectural advancements, they rely on implicit inter-modal correlations learned in a data-driven manner to facilitate feature interactions, as shown in Fig. 1(a). As modality coverage expands, limited cohort sizes render purely data-driven fusion incapable of fully characterizing high-level semantic and biological synergies across heterogeneous modalities, fostering spurious correlations that impede effective fusion. With the rapid progress of large language models (LLMs), recent studies have begun to inject medical knowledge into survival analysis, yet they have not considered knowledge-guided general fusion over holistic data. For instance, KEMM [33] restricts knowledge injection to auxiliary inputs only for specific modalities without explicit cross-modal guidance. Consequently, these approaches remain constrained by data-driven fusion paradigms and cannot fully exploit knowledge informed cross-modal integration.

To bridge these gaps, we propose a novel Medical Knowledge-Guided Survival Analysis (**MKGSurv**) framework that advances multimodal survival modeling beyond purely data-driven fusion. MKGSurv organizes holistic patient data into four disciplines based on clinical data provenance and incorporates interdisciplinary relationships and prognostic insights derived from LLMs as knowledge

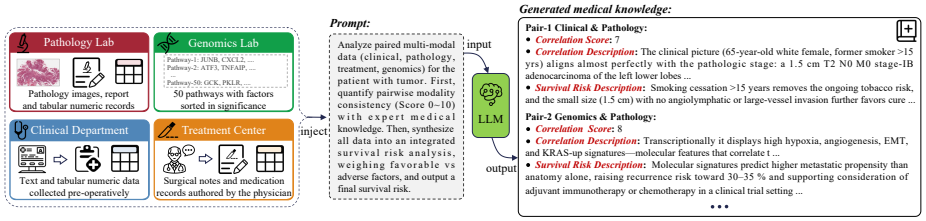

Fig. 2. Pipeline for medical knowledge generation. An LLM analyses multidisciplinary reports to generate knowledge of interdisciplinary correlation and prognostic insights.

to enable more informed multimodal integration (Fig. 1(b)), with clinician evaluation supporting its clinical validity (Fig. 1(c)). Leveraging this knowledge, MKGSurv guides multimodal representation learning at both the structural and objective levels. For structure, we introduce a Correlation-Aware Graph ATtention (CA-GAT) module, in which multimodal features are modeled as nodes, while LLM-derived descriptions are encoded as edges to explicitly guide cross-modal message passing and facilitate knowledge-informed interactions. For objective, we propose a Correlation-Guided Alignment (CGA) loss to regularize the latent representation by encouraging consistency between the learned interdisciplinary feature correlation and quantitative disciplinary relevance scores.

The contributions of this work are summarized as follows:

- We propose **MKGSurv**, a novel multimodal survival analysis framework that systematically integrates holistic hospital data and incorporates interdisciplinary knowledge to support structured multimodal modeling.
- We leverage the LLM to derive descriptions and quantified relevance scores, and develop a CA-GAT together with a CGA loss to respectively steer cross-disciplinary interaction and align multidisciplinary latent representations.
- We conducted extensive experiments on four TCGA cohorts and an in-house dataset, demonstrating MKGSurv's state-of-the-art performance, with human evaluation confirming the clinical validity of LLM-derived knowledge.

2 Methodology

We study the task of multimodal survival analysis. In this work, we organize holistic patient data into four disciplines, namely Pathology (P), Genomics (G), Clinical (C), and Treatment (T). Each discipline may comprise heterogeneous modalities, including images, genomic sequences, text or tabular variables. We denote the multimodal patient data as $\mathcal{M} = \{P, G, C, T\}$, and define a multimodal fusion model $F(\mathcal{M}) \rightarrow$ Risk to predict patient-specific survival risk.

2.1 Medical Knowledge Generation

We leverage the rich knowledge of LLMs to generate interdisciplinary correlations and prognostic insights as medical priors, as shown in Fig. 2. First, we transform

raw patient data into textual reports. For pathology, we use diagnostic pathology reports. For genomics, we concatenate gene names ranked by their expression levels in descending order. Textual and tabular variables from the clinical and treatment disciplines are directly serialized into descriptive statements. We assemble these inputs into a comprehensive patient report for LLM analysis. Then, conditioned on the patient context, the LLM assesses correlation strengths and produces three outputs for all discipline pairs: (1) quantitative relevance scores that measure pairwise cross-disciplinary correlations, (2) structured textual explanations describing interdisciplinary relationships, and (3) survival-oriented reasoning that articulates risk-related clinical insights.

2.2 Embeddings Extraction

Following prior studies [4,11,20], we employ modality-specific methods to extract embeddings $X_i \in \mathbb{R}^{N_i \times D}$, where discipline $i \in \mathcal{M} = \{P, G, C, T\}$ and N_i denotes the number of features. We project all embeddings to a unified dimension $D = 512$ using distinct two-layer MLP projectors with widths 1024 and 512.

We construct discipline-level embeddings as follows. (1) For pathology, we collect histopathology images along with pathology reports and tissue tabular records. Each slide is segmented into non-overlapping 256×256 patches. Patch embeddings are extracted using UNI [1] and aggregated via PANTHER [21] into $N_{pimg} = 128$ patient-specific cluster-level prototypes. These image prototypes are concatenated with ClinicalBERT-encoded pathology-report embeddings and MLP-projected tissue tabular embeddings to form $X_P \in \mathbb{R}^{N_p \times D}$. (2) For genomics, we map transcriptomic profiles to $N_g = 50$ Hallmark pathway [11] features and obtain $X_G \in \mathbb{R}^{N_g \times D}$. (3) For clinical data, we retain demographic and diagnostic variables and exclude follow-up variables to ensure all inputs are available before the first discharge. The clinical text is encoded by a frozen ClinicalBERT [9]. The tabular variables and textual features are projected and concatenated to obtain $X_C \in \mathbb{R}^{N_c \times D}$. (4) For treatment, we use surgical notes and tabular therapeutic records restricted to the initial regimen during the index hospitalization. The surgical notes are encoded by a frozen ClinicalBERT [9]. The resulting textual features and tabular variables are projected and concatenated to construct $X_T \in \mathbb{R}^{N_t \times D}$. (5) For LLM-derived descriptions, we encode descriptions separately using frozen ClinicalBERT [9] and concatenate their embeddings to get $E_{i,j} \in \mathbb{R}^{N_e \times D}$ for discipline pair $\{i, j\}$.

2.3 MKGSurv Framework

Our framework sequentially deploys Intra-Discipline, Inter-Discipline, and Fusion modules as shown in Fig. 3. This clinically motivated design leverages intra-discipline consistency while modeling complex cross-disciplinary interactions.

First, we introduce an Intra-Discipline Module that performs local fusion among features within each discipline. For embeddings $X_i \in \{X_P, X_G, X_C, X_T\}$, we apply a shared transformer encoder block [25] to propagate information within the same discipline, thereby improving representation consistency. This

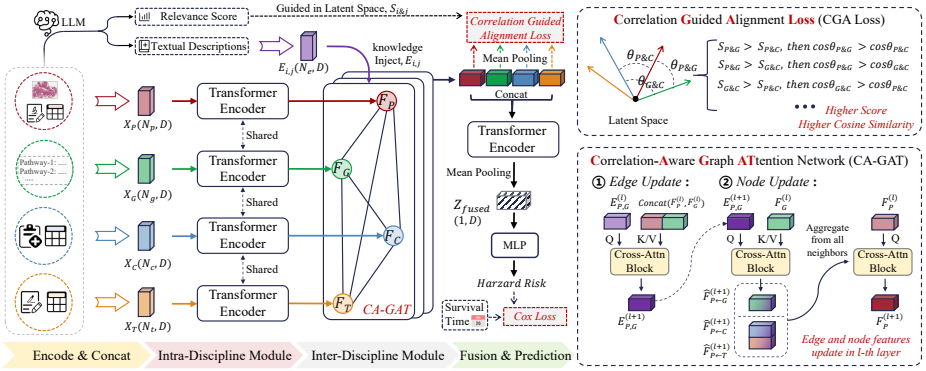

Fig. 3. Overview of MKGSurv. It processes multidisciplinary data via intra- and inter-discipline modules, employing CGA loss to regularize embeddings for survival analysis.

produces fine-grained embeddings $F_i \in \mathbb{R}^{N_i \times D}$ that encode intra-discipline interactions and facilitate effective subsequent cross-disciplinary fusion.

Secondly, in the Inter-Discipline Module, we propose a Correlation-Aware Graph Attention (CA-GAT) to model dependencies across disciplines. Inspired by [5,26], we define a graph where the disciplinary features $F_i \in \{F_P, F_G, F_C, F_T\}$ serve as node feature and LLM-generated medical knowledge $E_{i,j}$ serves as the initial edge features between disciplines i and j. In each CA-GAT layer, edge features first attend to node features to align textual knowledge with multimodal representations. Let $\mathcal{F}(Q, K, V) = \text{FFN}(\text{Cross-Attn}(Q, K, V))$ denote a cross-attention block with a feed-forward network, edge update is defined as:

$$E_{i,j}^{(l+1)} = \mathcal{F}\Big(Q = E_{i,j}^{(l)},\ K = V = \text{Concat}\Big(F_i^{(l)}, F_j^{(l)}\Big)\Big), \tag{1}$$

where Concat denotes the concatenation operation, l means the l-th CA-GAT layer. Subsequently, the updated edge feature guides knowledge conditioned message passing from node j to node i is defined as $\hat{F}_{i\leftarrow j}$, computing by:

$$\hat{F}_{i\leftarrow j} \ = \ \mathcal{F}\Big(Q = E_{i,j}^{(l+1)},\ K = V = F_j^{(l)}\Big). \tag{2}$$

We then aggregate features from neighboring nodes and update the i-th node:

$$F_i^{(l+1)} \ = \ \mathcal{F}\Big(Q = F_i^{(l)},\ K = V = \text{Concat}_{j\in\{\mathcal{P},\mathcal{G},\mathcal{C},\mathcal{T}\}\&j\neq i}\Big(\hat{F}_{i\leftarrow j}\Big)\Big). \tag{3}$$

This interaction ensures that disciplinary features are mutually refined by both domain-specific knowledge and cross-disciplinary information.

Finally, the updated features from all disciplines are concatenated and processed by a transformer encoder block, and then mean pooling is used to aggregate the embeddings into a unified representation Z_{fused}. Notably, our framework can dynamically address missing modalities via attention masking.

2.4 Training Objective

The model is optimized with a joint objective that combines the Cox partial likelihood loss [3] and our proposed Correlation-Guided Alignment (CGA) loss. We adopt the Cox loss, denoted as $\mathcal{L}_{cox}$, to learn survival risk prediction from the fused representation Z_{fused}. We also introduce LLM-derived relevance scores as medical priors and use them to regularize the latent feature space. Specifically, the CGA Loss, denoted as $\mathcal{L}_{cga}$, encourages the learned multimodal representations to adhere to medically meaningful relationships. We apply a patient-wise softmax to the LLM-derived relevance scores over all valid disciplinary pairs to obtain the target distribution $S_{\text{LLM}} = \{S_{P\&G}, S_{P\&T}..., S_{C\&T}\}$. Let $S'_{\text{Model}} = \{cos(\theta_{P\&G}), cos(\theta_{P\&T})..., cos(\theta_{C\&T})\}$ denote the pairwise cosine similarities between the discipline-pooled embeddings produced by the model. The CGA Loss minimizes the Kullback–Leibler (KL) divergence [16] between the LLM prior and the model-induced distribution over cross-disciplinary relationships:

$$\mathcal{L}_{cga} \;=\; \mathcal{D}_{KL}(S_{\text{LLM}} \parallel \text{Softmax}(S'_{\text{Model}}/\tau))\,, \tag{4}$$

where τ represents a learnable temperature parameter. By aligning these two distributions, $\mathcal{L}_{cga}$ pulls disciplines with strong medical relevance closer in the latent space to promote supportive evidence integration. Simultaneously, it also distinguishes weakly correlated disciplines to preserve possible complementary information. The overall training objective is $\mathcal{L}_{total} = \mathcal{L}_{cox} + \lambda\,\mathcal{L}_{cga}$.

3 Experiments

3.1 Experimental Settings

We evaluated our approach on four public cancer cohorts from The Cancer Genome Atlas [1] (TCGA) and one in-house cohort. The public cohorts include Lung Adenocarcinoma (LUAD, n=438), Lung Squamous Cell Carcinoma (LUSC, n=452), Breast Invasive Carcinoma (BRCA, n=868), and Kidney Renal Clear Cell Carcinoma (KIRC, n=340). The in-house cohort consisted of 1,325 patients with Oral Squamous Cell Carcinoma and was approved by the institutional ethics committee (Approval No. SYSKY-2024-440-01). As genomic data were unavailable for the in-house cohort due to electronic system limitations, we treated this as a missing modality scenario. Prediction target was overall survival (OS) for the TCGA cohorts and tumor recurrence-free survival for the in-house cohort.

We compared our framework against three categories of models: (1) unimodal baselines (pathology or genomics); (2) pathology-gene specific methods; and (3) general multimodal fusion models. Models utilized identical feature encoders and prediction head while keeping core fusion mechanism for fairness. Performance was evaluated by the concordance index (C-Index) [6]. During training, we optimized all models using AdamW [18] with an initial learning rate of 5×10^{-5}, cosine annealing, and a weight decay of 1×10^{-4} for 60 epochs on an NVIDIA

[1] https://portal.gdc.cancer.gov

Table 1. Performance comparison on five datasets under tumor-specific (P+G) and holistic fusion (P+G+C+T) settings. Results are reported as 5-fold mean±std.

Setting	Model	LUAD	LUSC	BRCA	KIRC	In House	Mean
P	PANTHER	0.5540±0.0461	0.5315±0.0554	0.6319±0.0237	0.6733±0.0804	0.7842±0.0834	-
G	MLP	0.5669±0.0530	0.5198±0.0212	0.6399±0.0613	0.6880±0.0836	-	-
P+G	SurvPath (CVPR'24)	0.6188±0.0763	0.5856±0.0519	0.6512±0.0572	**0.7132±0.0813**	-	0.6422
	MOME (MICCAI'24)	0.6031±0.0387	0.5749±0.0641	**0.6647±0.0555**	0.6979±0.0791	-	0.6352
	DIMAF (MICCAI'25)	0.6033±0.0405	0.5667±0.0614	0.6473±0.0525	0.7048±0.0650	-	0.6305
	MKGSurv (Ours)	**0.6348±0.0612**	**0.5893±0.0582**	0.6580±0.0626	0.7044±0.0979	-	**0.6466**
P+G+C+T	Transformer	0.6117±0.0240	0.6016±0.0307	0.6566±0.0405	0.7335±0.0786	0.8053±0.0650	0.6817
	HGCN (TMI'23)	0.6513±0.0482	0.6363±0.0429	0.7163±0.0737	0.7910±0.0517	0.8193±0.0761	0.7228
	I2MoE (ICML'25)	0.6171±0.0462	0.6367±0.0421	0.7200±0.0486	0.7839±0.0384	0.8237±0.0675	0.7163
	HealNet (NeurIPS'24)	0.6643±0.0492	0.6694±0.0402	0.7316±0.0823	0.7954±0.0363	0.8227±0.0790	0.7367
	MKGSurv (Ours)	**0.6880±0.0544**	**0.6729±0.0417**	**0.7361±0.0587**	**0.8083±0.0198**	**0.8296±0.0551**	**0.7470**

Table 2. Ablation study on the framework architecture components.

Intra-Discipline Module	Inter-Discipline Module	LUAD	LUSC	BRCA	KIRC	In House	Mean
×	×	0.6162±0.0715	0.6416±0.0299	0.6941±0.0167	0.7635±0.0466	0.8136±0.0744	0.7058
×	✓	0.6569±0.0544	0.6579±0.0385	0.7328±0.0414	0.8028±0.0262	0.8161±0.0777	0.7313
✓	×	0.6505±0.0410	0.6604±0.0421	0.7123±0.0673	0.7650±0.0450	0.8232±0.0698	0.7223
✓	✓	**0.6880±0.0544**	**0.6729±0.0417**	**0.7361±0.0587**	**0.8083±0.0198**	**0.8296±0.0551**	**0.7470**

Table 3. Ablation study on medical knowledge guidance and CGA loss.

CA-GAT Edge	CGA Loss	LUAD	LUSC	BRCA	KIRC	In House	Mean
×	✓	0.6309±0.0742	0.5937±0.0520	0.6942±0.0646	0.7462±0.0689	0.8232±0.0581	0.6976
✓	×	0.6632±0.0549	0.6496±0.0458	0.7074±0.0310	0.7750±0.0610	0.8069±0.0913	0.7204
Random	Random	0.5873±0.0527	0.5643±0.0620	0.6364±0.0335	0.6558±0.0995	0.7978±0.0912	0.6483
Random	×	0.6328±0.0455	0.6339±0.0342	0.6972±0.0920	0.7489±0.0751	0.8129±0.0841	0.7051
✓	✓	**0.6880±0.0544**	**0.6729±0.0417**	**0.7361±0.0587**	**0.8083±0.0198**	**0.8296±0.0551**	**0.7470**

RTX 3090 GPU. For MKGSurv, we used three CA-GAT layers and set the CGA loss weight to $\lambda = 7$. Following [7], we employed GeGLU [19] activation in the model. We used the Kimi-K2 [23] to generate medical knowledge. All datasets were evaluated using 5-fold cross-validation with a 6:2:2 train-val-test split.

3.2 Comparison with SOTA Methods

Table 1 compares the performance of methods under three configurations. Overall, unimodal baselines (P or G) consistently underperform multimodal approaches, indicating that complementary cross-modality information is crucial for accurate survival prediction. In the P+G setting, MKGSurv achieves the best C-index on LUAD and LUSC, while remaining competitive on BRCA and KIRC. MKGSurv attains the highest average C-index of 0.6466 across the four TCGA cohorts, slightly outperforming the second-best SurvPath (0.6422). These results suggest that the proposed knowledge-guided fusion effectively exploits the synergy between pathology and genomics. Performance improvements are more evident in the P+G+C+T setting, where information from multiple disciplines is jointly integrated. MKGSurv achieves the highest C-index, yielding the best overall mean C-index of 0.7470. This corresponds to an absolute improvement over the strongest baseline, HealNet (0.7367). A one-sided one-sample t-test fur-

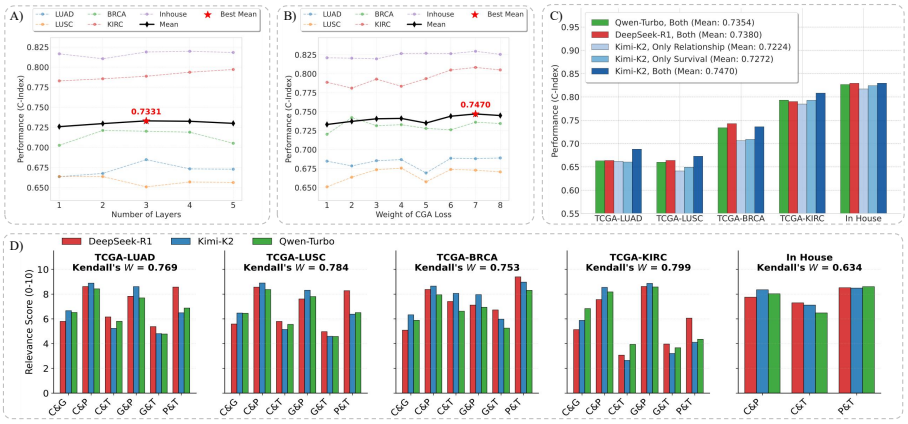

Fig. 4. Ablation study on hyperparameters and LLM-generated knowledge. **A**: Effect of the number of CA-GAT layers (L) with $\lambda = 1$. **B**: Effect of the loss weight (λ) with $L = 3$. **C**: Comparison of LLMs for knowledge generation and content ablations (relationships only, risk descriptions only, and their combination), showing that both descriptions are necessary. **D**: LLM-derived relevance scores for each discipline pair.

ther confirms that MKGSurv significantly outperforms the second-best method HealNet ($p = 0.0402 < 0.05$). Overall, these results support the benefit of incorporating medical knowledge into multimodal fusion and indicate that MKGSurv generalizes across different cancer types and data configurations.

3.3 Ablation Studies

We validate the core architectural components of MKGSurv in Table 2. Removing the Intra-Discipline Module reduces the overall C-Index to 0.7313, while removing the Inter-Discipline Module causes a larger drop to 0.7223. Disabling both modules yields the lowest performance of 0.7058, supporting the necessity of hierarchical within-discipline modeling and cross-disciplinary interaction learning. Table 3 analyzes the contributions of knowledge guidance and the CGA loss. Randomly initializing both edge features and supervision yields the lowest C-index, confirming that performance gains stem from informative knowledge rather than noise. In the absence of edge guidance, CA-GAT reverts to unguided node-level cross-attention and achieves a C-index of 0.6976 using only the CGA loss. Conversely, utilizing guided edges without CGA loss reaches 0.7204. Notably, randomly initialized edges without CGA loss achieve only 0.7051, still below LLM-guided edges without CGA loss, indicating that the gain comes from informative LLM priors rather than the structure alone. The integration of both components achieves optimal performance, demonstrating the complementary benefits of explicit edge guidance and latent alignment. We further analyze the sensitivity to key hyperparameters and LLM-derived knowledge. Fig. 4A–B shows that performance is stable over a reasonable range of differ-

ent hyperparameters. Fig. 4C demonstrates robustness across LLMs [17, 23, 31], with Kimi-K2 exhibiting the best performance and confirming that integrating disciplinary correlation and survival risk descriptions is essential for effective interaction. Clinician evaluation validates the reliability of knowledge in Fig. 1C and Fig. 4D visualizes the distribution of average relevance scores for discipline pairs and shows consistent rankings across LLMs, with Kendall [15]'s $W > 0.60$.

4 Conclusion

In this work, we present MKGSurv, a general multimodal fusion framework of holistic data for tumor survival analysis that leverages LLM-derived medical priors to bridge gaps among heterogeneous data in limited cohorts. By integrating the proposed CA-GAT and CGA loss, our approach explicitly steers cross-disciplinary interactions to align with established medical knowledge, thereby mitigating the susceptibility of purely data-driven fusion to spurious correlations. Extensive validation on four TCGA cohorts and an in-house dataset confirms that MKGSurv achieves state-of-the-art prognostic accuracy and robustness. Ultimately, this work offers a reliable, knowledge-guided fusion paradigm for translating complex hospital data into effective clinical decision support.

Acknowledgments. This research was funded by the National Natural Science Foundation of China (Grant No. 62376099).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature medicine **30**(3), 850–862 (2024)
2. Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F.: Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging **41**(4), 757–770 (2020)
3. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
4. Eijpe, A., Lakbir, S., Erdal Cesur, M., Oliveira, S.P., Abeln, S., Silva, W.: Disentangled and interpretable multimodal attention fusion for cancer survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 117–127. Springer (2025)
5. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y.: Hypergraph neural networks. Proceedings of the AAAI Conference on Artificial Intelligence **33**, 3558–3565 (2019)
6. Harrell Jr, F.E., Lee, K.L., Mark, D.B.: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine **15**(4), 361–387 (1996)

7. Hemker, K., Simidjievski, N., Jamnik, M.: Healnet: Multimodal fusion for heterogeneous biomedical data. Advances in Neural Information Processing Systems **37**, 64479–64498 (2024)
8. Hou, W., Lin, C., Yu, L., Qin, J., Yu, R., Wang, L.: Hybrid graph convolutional network with online masked autoencoder for robust multimodal cancer survival prediction. IEEE transactions on medical imaging **42**(8), 2462–2473 (2023)
9. Huang, K., Altosaar, J., Ranganath, R.: Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342 (2019)
10. Hui, D., Paiva, C.E., Del Fabbro, E.G., Steer, C., Naberhuis, J., van de Wetering, M., Fernández-Ortega, P., Morita, T., Suh, S.Y., Bruera, E., et al.: Prognostication in advanced cancer: update and directions for future research. Supportive Care in Cancer **27**(6), 1973–1984 (2019)
11. Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11579–11590 (2024)
12. Jee, J., Fong, C., Pichotta, K., Tran, T.N., Luthra, A., Waters, M., Fu, C., Altoe, M., Liu, S.Y., Maron, S.B., et al.: Automated real-world data integration improves cancer outcome prediction. Nature **636**(8043), 728–736 (2024)
13. Jiang, L., Xu, C., Bai, Y., Liu, A., Gong, Y., Wang, Y.P., Deng, H.W.: Autosurv: interpretable deep learning framework for cancer survival analysis incorporating clinical and multi-omics data. NPJ precision oncology **8**(1), 4 (2024)
14. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**(1), 24 (2018)
15. Kendall, M.G., Smith, B.B.: The problem of m rankings. Annals of Mathematical Statistics **10**(3), 275–287 (1939)
16. Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics **22**(1), 79–86 (1951)
17. Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al.: Deepseek-v3 technical report. arXiv preprint arXiv:2412.19437 (2024)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Shazeer, N.: Glu variants improve transformer. arXiv preprint arXiv:2002.05202 (2020)
20. Soenksen, L.R., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., Wiberg, H.M., Li, M.L., Fuentes, I., Bertsimas, D.: Integrated multimodal artificial intelligence framework for healthcare applications. NPJ digital medicine **5**(1), 149 (2022)
21. Song, A.H., Chen, R.J., Ding, T., Williamson, D.F., Jaume, G., Mahmood, F.: Morphological prototyping for unsupervised slide representation learning in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11566–11578 (2024)
22. Sun, B., Chen, L.: Interpretable deep learning for improving cancer patient survival based on personal transcriptomes. Scientific Reports **13**(1), 11344 (2023)
23. Team, K., Bai, Y., Bao, Y., Chen, G., Chen, J., Chen, N., Chen, R., Chen, Y., Chen, Y., Chen, Y., et al.: Kimi k2: Open agentic intelligence. arXiv preprint arXiv:2507.20534 (2025)
24. Vale-Silva, L.A., Rohr, K.: Long-term cancer survival prediction using multimodal deep learning. Scientific Reports **11**(1), 13505 (2021)

25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018)
27. Wulczyn, E., Steiner, D.F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., Mermel, C.H., Chen, P.H.C., Liu, Y., Stumpe, M.C.: Deep learning-based survival prediction for multiple cancer types using histopathology images. PloS one **15**(6), e0233678 (2020)
28. Xie, X., Cui, Y., Tan, T., Zheng, X., Yu, Z.: Fusionmamba: Dynamic feature enhancement for multimodal image fusion with mamba. Visual Intelligence **2**(1), 37 (2024)
29. Xin, J., Yun, S., Peng, J., Choi, I., Ballard, J.L., Chen, T., Long, Q.: I2MoE: Interpretable multimodal interaction-aware mixture-of-experts. In: International Conference on Machine Learning. pp. 68870–68888. PMLR (2025)
30. Xiong, C., Chen, H., Zheng, H., Wei, D., Zheng, Y., Sung, J.J., King, I.: Mome: Mixture of multimodal experts for cancer survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 318–328. Springer (2024)
31. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)
32. Yousefi, S., Amrollahi, F., Amgad, M., Dong, C., Lewis, J.E., Song, C., Gutman, D.A., Halani, S.H., Velazquez Vega, J.E., Brat, D.J., et al.: Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Scientific reports **7**(1), 1–11 (2017)
33. Zhao, C., Xu, Y., Zhou, F., Wang, Y., Chen, H.: Llm-driven knowledge enhancement for multimodal cancer survival prediction. arXiv preprint arXiv:2512.14594 (2025)