DUALRE-SURV: DUAL-PATH COMPLEMENTARY FUSION OF REPRESENTATION AND EVIDENCE FOR MULTIMODAL CANCER SURVIVAL PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate survival prediction is vital for clinical decision. Fusing whole slide images (WSIs) and genomics data has become a key research direction, as they respectively capture the histomorphological and molecular characteristics of tumors. Existing methods typically employ feature concatenation, attention mechanisms, or cross-modal interactive learning for joint representation; however, they rarely explicitly preserve the prognostic evidence provided by different modalities at the decision level. When pathology and genomics yield inconsistent judgments regarding a patient's survival risk, modal conflicts may be propagated to the final prediction via the joint representation. To alleviate this issue, we propose DualRE-Surv, a dual-path complementary fusion of representation and evidence for multimodal cancer survival prediction. In addition to the representation path for learning pathology-genomics joint representations, we introduce a parallel evidence-aware path. This path employs Dirichlet Vacuity Estimation (DVE) to measure the sufficiency of modal evidence, utilizes Consensus-Complementarity Evidence Decomposition (CCED) to decouple cross-modal consensus evidence and complementary evidence, and enhances the discriminability of consensus evidence across different time intervals through Consensus Evidence Enhancement (CEE). Subsequently, Joint Evidential Survival (JES) integrates the two types of evidence and constructs the survival curve of the evidence path. The Representation-Evidence Survival Mixture Module further integrates the prediction results of the two paths, enabling the model to simultaneously utilize the cross-modal joint representation and the conflict-adjusted modal decision evidence. We evaluate the proposed method on eight TCGA cancer cohorts using five-fold cross-validation. DualRE-Surv achieves C-indices of 0.6477, 0.7192, 0.7318, 0.8114, 0.8120, 0.6104, 0.6830, and 0.6570 on HNSC, BLCA, BRCA, UCEC, KIRC, LUSC, SKCM, and STAD, respectively, outperforming existing methods. Furthermore, extensive validation experiments further verify the effectiveness of the overall DualRE-Surv framework and its key modules.

1 INTRODUCTION

The survival outcomes of cancer patients are jointly influenced by tumor histomorphology, molecular status, and the tumor microenvironment, exhibiting significant inter-individual heterogeneity. WSIs characterize tumor histomorphology and microenvironmental composition, while genomics data reflect gene expression and molecular pathways. Therefore, the effective integration of these two modalities has become a critical research challenge in multimodal cancer survival prediction (Gensheimer & Narasimhan (2019)). Due to the extremely high resolution of WSIs, existing methods typically partition them into patches, extract local features using pre-trained encoders, and then aggregate them via multiple instance learning (MIL), graph neural networks, or transformers (Lu et al. (2020)). DeepAttnMIL utilizes an attention mechanism to obtain patient-level prognostic representations (Yao et al. (2020)); and Patch-GCN models spatial relationships between tissue regions via graph convolutions (Chen et al. (2021a)); and TransMIL leverages transformers to capture correlations among different patches (Shao et al. (2021b)). To further incorporate genomics data, researchers have proposed various pathology-genomics fusion methods. Pathomic Fusion models

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cross-modal high-order interactions through a gating mechanism and kronecker product (Chen et al. (2020)); MCAT employs co-attention to learn the correspondences between genomics representations and patches (Chen et al. (2021b)) and MOTCat constrains cross-modal matching via optimal transport (Xu & Chen (2023)). These methods have advanced pathology and genomics fusion, enhancing the representational capacity for multimodal survival prediction.

Despite significant progress in multimodal representation learning, existing methods typically employ feature concatenation, attention mechanisms, or cross-modal interactive learning for joint representation, they rarely explicitly preserve the prognostic evidence provided by different modalities at the decision level. When pathology and genomics yield inconsistent judgments regarding a patient's survival risk, modal conflicts may be propagated to the final prediction via the joint representation. To alleviate this issue, we propose DualRE-Surv, a dual-path complementary fusion of representation and evidence for multimodal cancer survival prediction. In addition to the Representation Path for learning pathology-genomics joint representations, the model introduces a parallel Evidence-Aware Path. This path employs Dirichlet Vacuity Estimation (DVE) to measure the sufficiency of modal evidence, utilizes Consensus-Complementarity Evidence Decomposition (CCED) to separate cross-modal consensus evidence and complementary evidence, and enhances the discriminability of consensus evidence across different time intervals through Consensus Evidence Enhancement (CEE). Subsequently, Joint Evidential Survival integrates the two types of evidence and constructs the survival curve of the evidence path. The Representation-Evidence Survival Mixture Module further integrates the prediction results of the two paths, enabling the model to simultaneously utilize the cross-modal joint representation and the conflict-adjusted modal decision evidence. The main contributions of this paper are summarized as follows:

- We propose a novel framework called dual-path complementary fusion of representation and evidence for multimodal cancer survival prediction, which consists of the Input Module, the Representation Path, the Evidence-Aware Path, and the Representation-Evidence Survival Mixture Module.
- We propose an Evidence-Aware Path comprising DVE, CCED, CEE, and JES, which respectively perform evidence sufficiency assessment, consensus–complementarity evidence decomposition, consensus evidence enhancement, and joint evidential survival modeling.
- We systematically evaluate DualRE-Surv on eight TCGA cancer cohorts (HNSC, BLCA, BRCA, UCEC, KIRC, LUSC, SKCM, and STAD). The results demonstrate that this method achieves excellent survival prediction performance. Furthermore, extensive validation experiments further verify the effectiveness of the overall DualRE-Surv framework and its key modules.

2 METHOD

DualRE-Surv consists of the Input Module, the Representation Path, the Evidence-Aware Path, and the Representation–Evidence Survival Mixture Module. The framework is shown in Figure 1. The Input Module encodes WSI and genomics data separately. The Representation Path learns cross-modal interactions and constructs a survival curve from the joint representation. The Evidence-Aware Path maps modalitie survival logits into the evidence space and models prognostic evidence through DVE, CCED, CEE, and JES, producing an evidence-based survival curve. Finally, the Representation–Evidence Survival Mixture Module combines the two survival curves to obtain the final survival estimate.

2.1 INPUT MODULE

Pathology Input. Each WSI is divided into multiple patches, from which a pretrained UNI encoder extracts features. Given a WSI containing N_w patches, its initial feature matrix is denoted as $\mathbf{P}_w \in \mathbb{R}^{N_w \times 1024}$. To reduce the feature dimensionality, a multilayer perceptron (MLP) projects the original features into a lower-dimensional space, yielding the patch embedding matrix:

$$\mathbf{P}_w = f_{\mathrm{MLP}}(\mathbf{P}_w) \in \mathbb{R}^{N_w \times 256} \tag{1}$$

Genomics Input. The genomics data are reorganized into pathway-level representations according to biological priors. The raw feature vector of the p-th pathway is denoted as $\mathbf{g}_{o,p} \in \mathbb{R}^{d_p}$, where

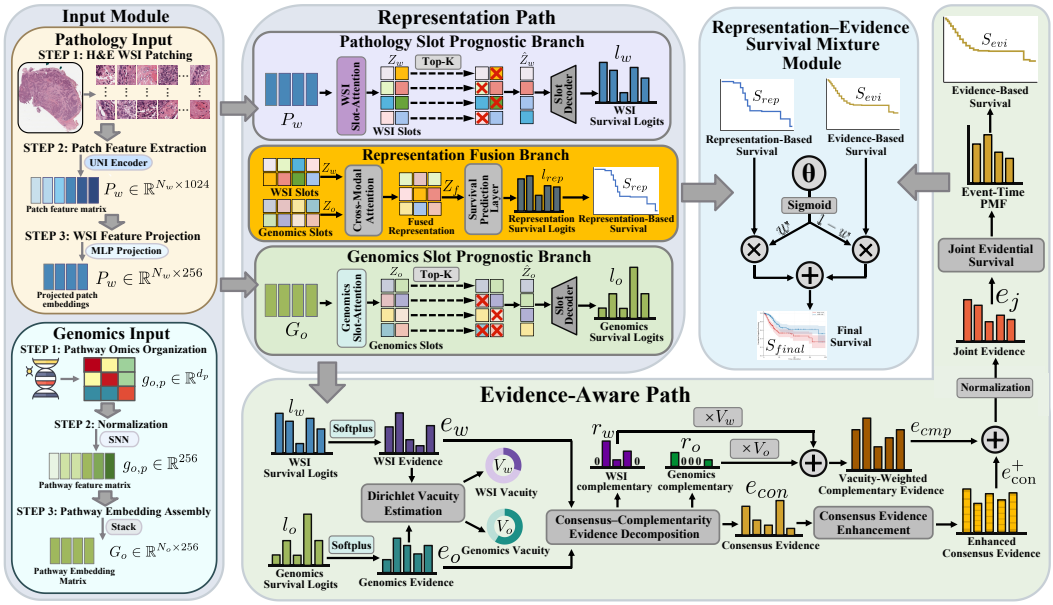

Figure 1: Overall framework of DualRE-Surv.

d_p is the number of genes contained in that pathway. Because the dimensionality d_p varies across pathways, a SNN maps each pathway feature into a shared low-dimensional embedding space:

$$\mathbf{g}_{o,p} = f_{\mathrm{SNN}}(\mathbf{g}_{o,p}) \in \mathbb{R}^{256} \tag{2}$$

The embeddings of all N_o pathways are stacked to form the patient-level genomics feature matrix:

$$\mathbf{G}_o = \mathrm{Stack}\left(\mathbf{g}_{o,1}, \mathbf{g}_{o,2}, \ldots, \mathbf{g}_{o,N_o}\right) \in \mathbb{R}^{N_o \times 256} \tag{3}$$

2.2 Representation Path

The Representation Path consists of the Pathology Slot Prognostic Branch, the Genomics Slot Prognostic Branch, and the Representation Fusion Branch. It models pathology–genomics interactions in the feature space for survival prediction. Inspired by SlotSPE (Zhang et al. (2026b)), the two modality branches employ slot attention to learn structured representations and generate survival logits. The fusion branch integrates these representations to construct a representation-based survival curve.

Pathology Slot Prognostic Branch. Given the pathology features $\mathbf{P}_w$ produced by the Input Module, this branch employs WSI slot-attention to aggregate a large number of patch features and compress them into a fixed number of pathology slots:

$$\mathbf{Z}_w = \mathrm{SlotAttn}_w(\mathbf{P}_w) \tag{4}$$

where $\mathbf{Z}_w$ denotes the set of pathology slots. The pathology slot representations are further selected according to their importance scores:

$$\widehat{\mathbf{Z}}_w = \mathrm{TopK}(\mathbf{Z}_w) \tag{5}$$

where $\widehat{\mathbf{Z}}_w$ denotes the set of selected pathology slots. The selected slots are then mapped by a slot decoder to the WSI survival logits over K time intervals:

$$\mathbf{l}_w = \mathrm{Decoder}_w\left(\widehat{\mathbf{Z}}_w\right) \in \mathbb{R}^K \tag{6}$$

where $\mathbf{l}_w$ denotes the prognostic predictions of the pathology modality over the time intervals.

Genomics Slot Prognostic Branch. The Genomics Slot Prognostic Branch adopts a structure similar to that of the Pathology Slot Prognostic Branch. The genomics features $\mathbf{G}_o$ produced by the Input Module are first processed by genomics slot attention to obtain the set of genomics slots $\mathbf{Z}_o$. After Top-K selection and slot decoder processing, the branch generates the genomics survival logits $\mathbf{l}_o$.

Representation Fusion Branch. The pathology slots $\mathbf{Z}_w$ and genomics slots $\mathbf{Z}_o$ interact through cross-modal attention to produce the joint representation:

$$\mathbf{Z}_f = \mathrm{CrossAttn}\left(\mathbf{Z}_w, \mathbf{Z}_o\right) \tag{7}$$

The survival prediction layer maps $\mathbf{Z}_f$ to the representation-based survival logits:

$$\mathbf{l}_{\mathrm{rep}} = f_{\mathrm{surv}}(\mathbf{Z}_f) \in \mathbb{R}^K \tag{8}$$

The conditional hazard probability for the k-th time interval is defined as $h_{\mathrm{rep}}(k) = \sigma(l_{\mathrm{rep},k})$. The representation-based survival curve is then computed as:

$$S_{\mathrm{rep}}(k) = \prod_{j=1}^{k} \left[1 - h_{\mathrm{rep}}(j)\right], \qquad k = 1, 2, \ldots, K \tag{9}$$

2.3 Evidence-Aware Path

In addition to constructing the representation-based survival curve, the two modality slot prognostic branches in the Representation Path separately output the WSI and genomics survival logits, $\mathbf{l}_w, \mathbf{l}_o \in \mathbb{R}^K$. Taking these logits as input, the Evidence-Aware Path models WSI and genomics prognostic evidence at the decision level through DVE, CCED, CCE, and JES, and constructs an evidence-aware survival curve. Its detailed structure is shown in Figure 2.

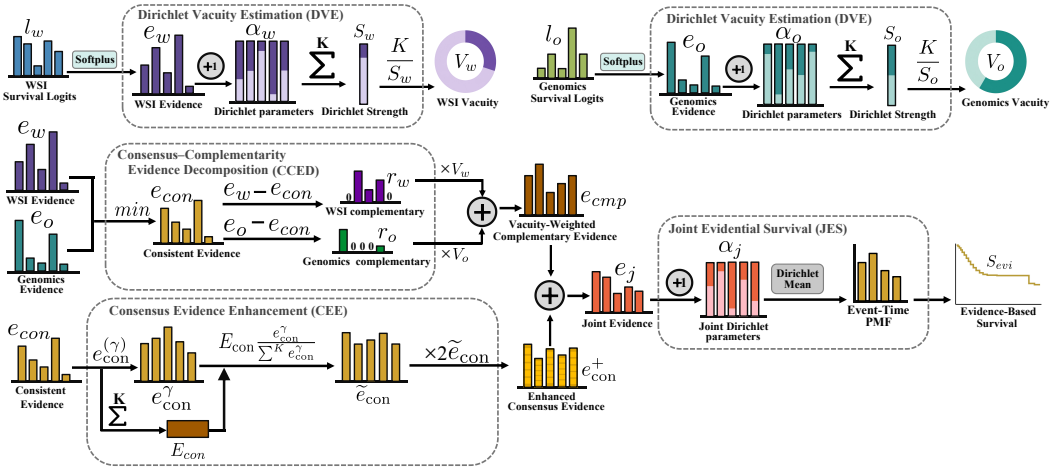

Figure 2: Detailed structure of the Evidence-Aware Path.

Dirichlet Vacuity Estimation (DVE). For each modality $m \in \{w, o\}$, the WSI and genomics survival logits are mapped to nonnegative evidence using the softplus function:

$$\mathbf{e}_m = \log\left(1 + \exp(\mathbf{l}_m)\right), \qquad m \in \{w, o\} \tag{10}$$

where $\mathbf{e}_w$ and $\mathbf{e}_o$ denote the nonnegative evidence vectors of the WSI and genomics. By incorporating a unit dirichlet prior, the dirichlet parameters are obtained as $\boldsymbol{\alpha}_m = \mathbf{e}_m + \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^K$ is an all-ones vector. The unit dirichlet prior ensures that the dirichlet parameter remains positive even when no additional evidence is assigned to a particular time interval. The dirichlet strength of each modality is then obtained by summing its parameters over the K time intervals:

$$S_m = \sum_{k=1}^{K} \alpha_{m,k} = K + \sum_{k=1}^{K} e_{m,k}, \qquad m \in \{w, o\} \tag{11}$$

where $\alpha_{m,k}$ and $e_{m,k}$ denote the dirichlet parameter and evidence value of modality m for the k-th time interval, respectively. S_m denotes the total amount of evidence provided by modality m across all time intervals. A larger S_m indicates that the modality provides more sufficient evidence. Accordingly, the vacuity of each modality is defined as:

$$V_m = \frac{K}{S_m}, \qquad m \in \{w, o\} \tag{12}$$

V_m is larger when the total evidence of a modality is low and gradually decreases as the total evidence increases.

Consensus–Complementarity Evidence Decomposition (CCED). Pathology and genomics may provide similar prognostic evidence within the same time interval, while also offering distinct evidential support. To decouple these two components, we compare the evidence vectors of the two modalities element-wise and define their minimum as the consensus evidence:

$$\mathbf{e}_{\text{con}} = \min(\mathbf{e}_w, \mathbf{e}_o) \tag{13}$$

where $\min(\cdot)$ denotes the element-wise minimum operation. The portions exceeding the consensus evidence are defined as the complementary evidence of the corresponding modalities:

$$\mathbf{r}_w = \mathbf{e}_w - \mathbf{e}_{\text{con}} \tag{14}$$

$$\mathbf{r}_o = \mathbf{e}_o - \mathbf{e}_{\text{con}} \tag{15}$$

where $\mathbf{r}_w$ and $\mathbf{r}_o$ denote the complementary evidence provided by pathology and genomics beyond the consensus component, respectively. Further explanation and analysis are provided in Appendix B.

Consensus Evidence Enhancement (CEE). To emphasize the time intervals receiving stronger consensus support, we first compute the total amount of consensus evidence:

$$E_{\text{con}} = \sum_{k=1}^{K} e_{\text{con},k} \tag{16}$$

An element-wise power transformation is then applied to the consensus evidence:

$$\mathbf{e}_{\text{con}}^{(\gamma)} = \mathbf{e}_{\text{con}}^{\gamma}, \qquad \gamma = 1.25 \tag{17}$$

where γ denotes the power transformation. To preserve the total amount of the original consensus evidence, the transformed evidence is renormalized using E_{con}:

$$\widetilde{\mathbf{e}}_{\text{con}} = E_{\text{con}} \frac{\mathbf{e}_{\text{con}}^{(\gamma)}}{\left\| \mathbf{e}_{\text{con}}^{(\gamma)} \right\|_1} \tag{18}$$

Because the consensus evidence is jointly supported by both pathology and genomics, the normalized result is scaled by a factor of two to obtain the enhanced consensus evidence $\mathbf{e}_{\text{con}}^{+} = 2\widetilde{\mathbf{e}}_{\text{con}}$. Further explanation and analysis are provided in Appendix C.

Joint Evidential Survival (JES). After CCED, $\mathbf{r}_w$ and $\mathbf{r}_o$ represent the complementary evidence provided by pathology and genomics, respectively. The complementary evidence is weighted by the vacuity of the corresponding modality and then combined to obtain the vacuity-weighted complementary evidence:

$$\mathbf{e}_{\text{cmp}} = V_w \mathbf{r}_w + V_o \mathbf{r}_o \tag{19}$$

where $\mathbf{e}_{\text{cmp}} \in \mathbb{R}^K_{\geq 0}$ denotes the vacuity-weighted complementary evidence, while V_w and V_o denote the vacuity of the pathology and genomics modalities, respectively. When a modality provides less overall evidence, a larger V_m preserves more of its remaining complementary evidence. When the evidence is more abundant, a smaller V_m imposes a stronger constraint on its complementary evidence, preventing high-magnitude evidence from a single modality from dominating the joint result. The enhanced consensus evidence and vacuity-weighted complementary evidence are then combined to form the joint evidence:

$$\mathbf{e}_j = \mathbf{e}_{\text{con}}^{+} + \mathbf{e}_{\text{cmp}} \tag{20}$$

A unit prior is added to the joint evidence to obtain the joint dirichlet parameters $\boldsymbol{\alpha}_j = \mathbf{e}_j + \mathbf{1}$. The mean of the joint dirichlet distribution is used as the event-time probability mass function:

$$\mathbf{p}_{\text{evi}} = \frac{\boldsymbol{\alpha}_j}{\|\boldsymbol{\alpha}_j\|_1} \tag{21}$$

where $p_{\text{evi}}(k)$ denotes the probability that the event occurs within the k-th time interval. By accumulating the event-time probabilities, the evidence-aware survival curve is obtained as:

$$S_{\text{evi}}(k) = 1 - \sum_{t=1}^{k} p_{\text{evi}}(t), \qquad k = 1, 2, \ldots, K \tag{22}$$

Further explanation and analysis are provided in Appendix D.

2.4 Representation–Evidence Survival Mixture Module

This module dynamically fuses representation-based (S_{rep}) and evidence-based (S_{evi}) survival predictions. A learnable parameter θ is passed through a sigmoid activation to generate an adaptive weight w. The final survival prediction is then computed as a weighted sum:

$$S_{\text{final}} = wS_{\text{rep}} + (1 - w)S_{\text{evi}} \tag{23}$$

Further explanation and analysis are provided in Appendix E.

2.5 Loss Function

DualRE-Surv is trained using the discrete-time survival negative log-likelihood. For patient i, let $y_i \in \{1, \ldots, K\}$ denote the interval containing the observed time and $c_i \in \{0, 1\}$ the censoring status, where $c_i = 1$ indicates censoring. Given the predicted survival curve S_i, the probability of an event occurring in interval y_i is $S_i(y_i-1)-S_i(y_i)$. The patient-level survival negative log-likelihood is therefore defined as:

$$\mathcal{L}_{\text{NLL}}^{(i)}(S) = -(1 - c_i) \log[S_i(y_i - 1) - S_i(y_i)] - (1 - \alpha)c_i \log S_i(y_i) \tag{24}$$

where α controls the loss weight of censored samples. The loss is averaged over the N patients:

$$\mathcal{L}_{\text{NLL}}(S) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\text{NLL}}^{(i)}(S) \tag{25}$$

Survival supervision is applied to the final fused output and the unimodal WSI and genomics outputs:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{NLL}}(S_{\text{final}}), \qquad \mathcal{L}_{\text{uni}} = \mathcal{L}_{\text{NLL}}(S_w) + \mathcal{L}_{\text{NLL}}(S_o) \tag{26}$$

The overall survival prediction loss is: $\mathcal{L}_{\text{surv}} = \mathcal{L}_{\text{final}} + \mathcal{L}_{\text{uni}}$.

3 Experiment

3.1 Datasets and Evaluation Metrics

We evaluate DualRE-Surv on eight cancer cohorts from The Cancer Genome Atlas (TCGA) (Weinstein et al. (2013)): HNSC ($n = 438$), BLCA ($n = 381$), BRCA ($n = 1,046$), UCEC ($n = 488$), KIRC ($n = 488$), LUSC ($n = 460$), SKCM ($n = 409$), and STAD ($n = 366$). The data for each patient include a WSI and genomics data. Disease-specific survival (DSS) is used as the prediction endpoint. Survival time is recorded in months, while event and censoring status are determined from clinical follow-up information. We use the C-index (Harrell Jr et al., 1996) as the primary metric and Kaplan-Meier curves (Kaplan & Meier, 1958) with the log-rank test for risk-stratification evaluation.

3.2 Experiment details

All experiments are implemented in PyTorch (Paszke et al. (2019)) and conducted on a single NVIDIA 5090 GPU. The network is trained using the Adam optimizer (Kingma & Ba (2014)) with an initial learning rate of 2×10^{-4}, a weight decay of 1×10^{-5}, and a batch size of 32. Each fold is trained for 30 epochs with a random seed of 42. Further details are provided in the Appendix F.

3.3 Comparative experiment

We compare DualRE-Surv with both unimodal and multimodal methods. MLP (Haykin (1994)), SNN (Klambauer et al. (2017)), and SNNTrans (Klambauer et al. (2017)) use only genomics data, whereas ABMIL (Ilse et al. (2018)), TransMIL (Shao et al. (2021b)), CLAM-SB (Lu et al. (2020)), and DPsurv (Xing et al. (2025)) use only WSIs. PORPOISE (Chen et al. (2022)), MCAT (Chen et al. (2021b)), MOTCat (Xu & Chen (2023)), CMTA (Zhou & Chen (2023)), SurvPath (Jaume et al. (2024)), PIBD (Zhang et al. (2024)), MMP (Song et al. (2024)), LD-CVAE (Zhou et al. (2025)), SurMoE (Zhang et al. (2026a)), and SlotSPE (Zhang et al. (2026b)) use both WSIs and genomics data. All methods are evaluated using the same five-fold data splits. Table 1 reports the mean C-index across the eight datasets, the complete results are provided in the Appendix G. DualRE-Surv achieves the highest mean C-index on all datasets, reaching 0.6477, 0.7192, 0.7318, 0.8114, 0.8120, 0.6104, 0.6830, and 0.6570 on HNSC, BLCA, BRCA, UCEC, KIRC, LUSC, SKCM, and STAD.

Table 1: Comparison of mean C-index performance on eight TCGA datasets. G. and P. denote genomics and pathology. **Bold** and underlined indicate the best and second-best results.

Method	Modality		C-index							
	G.	P.	HNSC	BLCA	BRCA	UCEC	KIRC	LUSC	SKCM	STAD
MLP	✓		0.5674	0.6570	0.6443	0.7054	0.7643	0.5307	0.6320	0.5485
SNN	✓		0.5987	0.6080	0.6368	0.7570	0.7322	0.5681	0.6159	0.5278
SNNTrans	✓		0.5896	0.6254	0.6482	0.7432	0.7082	0.5612	0.6288	0.5857
DualRE-Surv	✓		0.6402	0.6704	0.6792	0.7822	0.7906	0.5935	0.6624	0.6298
ABMIL		✓	0.6213	0.6421	0.6490	0.7794	0.7367	0.5275	0.6405	0.6036
TransMIL		✓	0.6055	0.6677	0.6411	0.7800	0.7751	0.5563	0.6098	0.5983
CLAM-SB		✓	0.6005	0.6704	0.6681	0.7836	0.6908	0.5173	0.6431	0.6145
DPsurv		✓	0.5763	0.6227	0.6661	0.6358	0.6942	0.5322	0.5978	0.6043
DualRE-Surv		✓	0.6336	0.6667	0.6806	0.7917	0.7847	0.6064	0.6593	0.6471
PORPOISE	✓	✓	0.5998	0.6398	0.6541	0.7953	0.7149	0.5988	0.6659	0.6239
MCAT	✓	✓	0.6125	0.6679	0.6812	0.7986	0.8017	0.5764	0.6580	0.6348
MOTCat	✓	✓	0.6137	0.6896	0.7008	0.7809	0.7868	0.5868	0.6577	0.6308
CMTA	✓	✓	0.6249	0.6783	0.7196	0.8003	0.7915	0.5760	<u>0.6802</u>	0.6417
SurvPath	✓	✓	0.6036	0.6935	0.6754	0.7991	0.7595	0.5901	0.6594	0.6296
PIBD	✓	✓	0.6368	0.6821	0.6812	0.7812	0.7787	0.5895	0.6531	<u>0.6501</u>
MMP	✓	✓	0.6225	0.7033	<u>0.7212</u>	0.7654	0.8046	0.5793	0.6653	0.6377
LD-CVAE	✓	✓	0.6198	0.6850	0.7135	<u>0.8010</u>	0.7954	<u>0.6038</u>	0.6688	0.6469
SurMoE	✓	✓	0.6334	0.6689	0.7000	0.7582	0.7660	0.5964	0.6572	0.6295
SlotSPE	✓	✓	<u>0.6418</u>	<u>0.7061</u>	0.7083	0.7975	<u>0.8078</u>	0.5775	0.6694	0.6337
DualRE-Surv	✓	✓	**0.6477**	**0.7192**	**0.7318**	**0.8114**	**0.8120**	**0.6104**	**0.6830**	**0.6570**

3.4 Ablation experiment

To evaluate the Evidence-Aware Path, we conduct four full-modality ablations by removing the entire path, DVE, CCED, and CEE, respectively. Table 2 reports the mean C-index, with complete results provided in the Appendix H. Removing the entire path or any of its components reduces performance across all eight cohorts, confirming their contributions to the model.

Table 2: Ablation study of DualRE-Surv on eight TCGA cohorts (mean C-index).

Variant	C-index							
	HNSC	BLCA	BRCA	UCEC	KIRC	LUSC	SKCM	STAD
w/o Evidence-Aware Path	0.6379	0.7005	0.7029	0.8016	0.7902	0.5829	0.6703	0.6418
w/o DVE	0.6397	0.7159	0.7250	0.8040	0.7997	0.6038	0.6732	0.6504
w/o CCED	0.6413	0.7137	0.7001	0.8023	0.8099	0.6077	0.6799	0.6366
w/o CEE	0.6467	0.7087	0.7307	0.8082	0.8105	0.5988	0.6765	0.6547
DualRE-Surv	**0.6477**	**0.7192**	**0.7318**	**0.8114**	**0.8120**	**0.6104**	**0.6830**	**0.6570**

3.5 Survival risk stratification analysis

Patients were stratified into high- and low-risk groups based on the median predicted risk. Kaplan–Meier curves (Kaplan & Meier (1958)) and stratified log-rank tests (Peto & Peto (1972)) were used to assess the separation between the two groups. DualRE-Surv achieved statistically significant separation across all eight TCGA cohorts and yielded smaller stratified log-rank p-values than SlotSPE and LD-CVAE in every cohort. Figure 3 shows a comparison on HNSC, and the complete results are provided in the Appendix I.

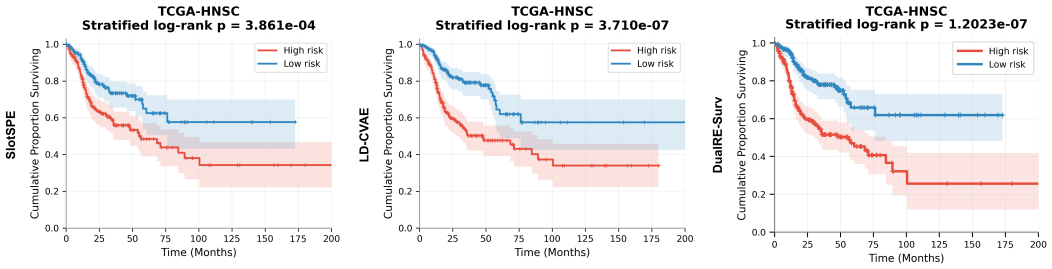

Figure 3: Comparison of Kaplan-Meier risk stratification results on HNSC.

3.6 Modality Conflict Analysis

To compare the three prediction paths under modality conflict, we measure patient-level conflict using the discrepancy between the event-time distributions predicted from pathology and genomics:

$$C = \frac{\mathrm{JS}(\mathbf{p}_w, \mathbf{p}_o)}{\log 2} \sqrt{(1 - V_w)(1 - V_o)} \tag{27}$$

where $\mathrm{JS}(\cdot)$ denotes the Jensen–Shannon divergence, which measures the discrepancy between the event-time distributions predicted by the two modalities; $\mathbf{p}_w$ and $\mathbf{p}_o$ are the event-time distributions predicted by the two modalities. Patients are divided into low-conflict, medium-conflict, and high-conflict groups according to their conflict scores. We compare the Representation Path, Evidence-Aware Path, and final prediction using the patient-level discrete-time survival NLL. Partial results are shown in Figure 4. Across all eight cohorts, the final prediction achieves a lower mean NLL than the Representation Path in every conflict group. A lower NLL indicates a higher predicted probability of the observed event interval for uncensored patients or of survival up to the censoring time for censored patients, demonstrating the effectiveness of integrating the two paths under different levels of modality conflict. Detailed definitions and complete results are provided in Appendix J.

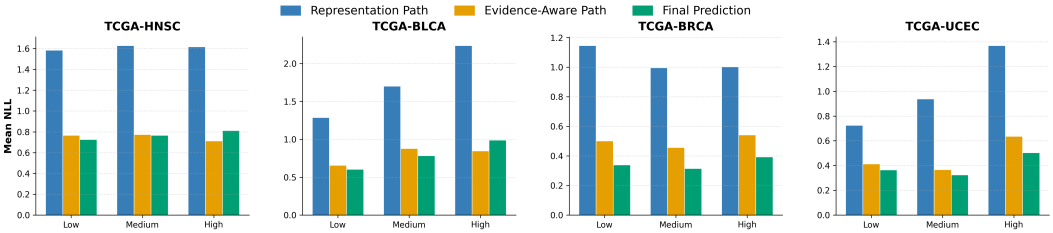

Figure 4: Comparison of the mean NLLs of the Representation Path, Evidence-Aware Path, and final prediction on HNSC, BLCA, BRCA, UCEC.

3.7 Conflict Perturbation Analysis

To assess the effectiveness of the Evidence-Aware Path under modality conflict, we keep the pathology evidence unchanged and progressively replace the genom genomics evidence with evidence selected from a patient exhibiting a substantially different event-time distribution. For patient i, the perturbed genomics evidence is constructed as:

$$\tilde{\mathbf{e}}_{o,i}^{(\lambda)} = (1 - \lambda)\mathbf{e}_{o,i} + \lambda \mathbf{e}_{o,d(i)}, \qquad \lambda \in \{0, 0.25, 0.5, 0.75, 1\} \tag{28}$$

where $d(i)$ denotes the selected patient and λ controls the perturbation strength. We compare the full evidential mechanism with direct evidence addition and a variant without vacuity weighting. Partial experimental results are shown in Figure 5, while further details and complete results are provided in Appendix K. As shown in the left panels, the mean conflict score increases with λ, indicating that the perturbation progressively enlarges the discrepancy between pathology and genomics evidence. The right panels show that the survival-curve deviations increase with the perturbation strength under all three evidence combination strategies. However, the full evidential mechanism consistently

produces smaller deviations than direct evidence addition and the variant without vacuity weighting.

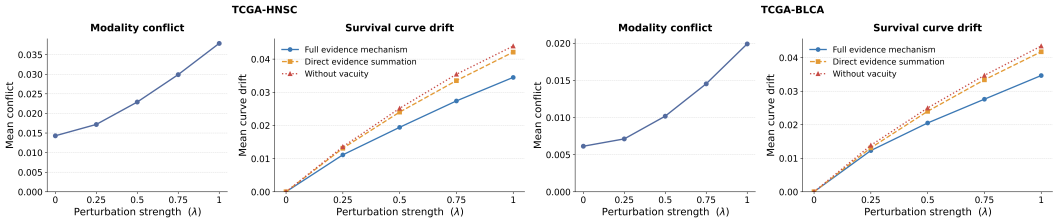

Figure 5: Conflict perturbation results on HNSC and BLCA.

3.8 Robustness Analysis under Input Perturbations

To evaluate the robustness of DualRE-Surv to input perturbations, we apply Gaussian and dropout noise to the pathology and genomics features at five noise. We compare DualRE-Surv with SlotSPE using the mean absolute risk change relative to the clean input, where a smaller value indicates greater robustness. Partial experimental results are shown in Figure 6. DualRE-Surv consistently exhibits smaller changes than SlotSPE, demonstrating its robustness against noise perturbations. Detailed experimental settings and complete results are provided in the Appendix L.

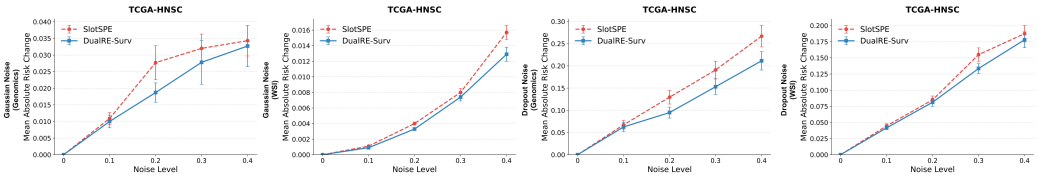

Figure 6: Robustness comparison between SlotSPE and DualRE-Surv on HNSC.

3.9 Computational Overhead Analysis

We compare the computational overhead of DualRE-Surv with other methods on HNSC. As shown in Table 3, DualRE-Surv has 31.535M trainable parameters and requires 4.7 hours for training and 0.091 seconds per inference. Its computational cost is close to SlotSPE and substantially lower than LD-CVAE. Although it is larger than CMTA, DualRE-Surv requires less training time and achieves better survival prediction across the evaluated cohorts.

Table 3: Computational overhead comparison on the HNSC cohort.

Method	Parameters	Training time (h)	Inference time (s)
CMTA	5.01M	6.1	0.027
LD-CVAE	209.799M	10.9	0.194
SlotSPE	31.535M	4.6	0.087
DualRE-Surv	31.535M	4.7	0.091

4 Conclusion

This study presents DualRE-Surv to address the effect of decision discrepancies between pathology and genomics modalities on multimodal survival prediction. The Representation Path learns their joint representation, while the Evidence-Aware Path explicitly models modality-specific prognostic evidence at the decision level. The survival curves produced by the two paths are then combined to obtain the final prediction. Experiments on eight TCGA cancer cohorts show that DualRE-Surv achieves the highest C-index on all cohorts. Furthermore, extensive validation experiments further verify the effectiveness of the overall DualRE-Surv framework and its key modules.

AI USE STATEMENT

Generative AI tools were used to assist with English translation, language editing, and the refinement and presentation of parts of the mathematical derivations. All AI-assisted text, equations, and derivations were independently reviewed and verified by the authors. Generative AI was not used to generate datasets, conduct experiments, produce the reported results, or determine the conclusions of this work. The authors take full responsibility for the content of this paper.

ETHICS STATEMENT

This study uses de-identified data from The Cancer Genome Atlas (TCGA). All data were obtained and used in accordance with the applicable data-access and data-use policies. No new participants were recruited, and no interventions or additional patient data collection were conducted. The proposed method is intended for research purposes and should not be used for clinical decision-making without further external and prospective validation. Because model performance may vary across institutions and patient populations, broader validation is necessary before potential clinical application.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce this work in the main paper and the appendices, including the cohort construction and data preprocessing procedures, model architecture, learning objectives, implementation details, hyperparameter settings and evaluation protocols. Complete experimental results, additional analyses, and detailed theoretical derivations are also included in the appendices. The TCGA cohorts used in this study are identified to facilitate independent evaluation.

REFERENCES

Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. 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.

Richard J. Chen, Ming Y. Lu, Muhammad Shaban, Chengkuan Chen, Tiffany Y. Chen, Drew F. K. Williamson, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks, 2021a. URL https://arxiv.org/abs/2107.13048.

Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *2021 IEEE/CVF international conference on computer vision (ICCV)*, pp. 3995–4005. IEEE, 2021b.

Richard J Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Jana Lipkova, Zahra Noor, Muhammad Shaban, Maha Shady, Mane Williams, Bumjin Joo, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer cell*, 40(8):865–878, 2022.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024.

Travers Ching, Xun Zhu, and Lana X Garmire. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology*, 14(4): e1006076, 2018.

Kexin Ding, Mu Zhou, Dimitris N. Metaxas, and Shaoting Zhang. Pathology-and-genomics multimodal transformer for survival outcome prediction, 2023. URL https://arxiv.org/abs/2307.11952.

Michael F Gensheimer and Balasubramanian Narasimhan. A scalable discrete-time survival model for neural networks. *PeerJ*, 7:e6257, 2019.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Jie Hao, Youngsoon Kim, Tejaswini Mallavarapu, Jung Hun Oh, and Mingon Kang. Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. *BMC medical genomics*, 12(Suppl 10):189, 2019.

Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice hall PTR, 1994.

Ling Huang, Yucheng Xing, Qika Lin, Su Ruan, and Mengling Feng. Esurvfusion: An evidential multimodal survival fusion model based on gaussian random fuzzy numbers. *arXiv preprint arXiv:2412.01215*, 2024.

Ling Huang, Yucheng Xing, Swapnil Mishra, Thierry Denœux, and Mengling Feng. Evidential time-to-event prediction with calibrated uncertainty quantification. *International Journal of Approximate Reasoning*, 181:109403, 2025.

Zhi Huang, Xiaohui Zhan, Shunian Xiang, Travis S Johnson, Bryan Helm, Christina Y Yu, Jie Zhang, Paul Salama, Maher Rizkalla, Zhi Han, et al. Salmon: survival analysis learning with multi-omics neural networks on breast cancer. *Frontiers in genetics*, 10:166, 2019.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pp. 2127–2136. Pmlr, 2018.

Guillaume Jaume, Anurag Vaidya, Richard J Chen, Drew FK Williamson, Paul Pu Liang, and Faisal Mahmood. Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In *2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11579–11590. IEEE, 2024.

Shuai Jiang, Arief A Suriawinata, and Saeed Hassanpour. Mhattnsurv: Multi-head attention for survival prediction using whole-slide pathology images. *Computers in biology and medicine*, 158:106883, 2023.

Songhan Jiang, Zhengyu Gan, Linghan Cai, Yifeng Wang, and Yongbing Zhang. Multimodal cross-task interaction for survival analysis in whole slide pathological images, 2024. URL `https://arxiv.org/abs/2406.17225`.

Songhan Jiang, Linghan Cai, Zhengyu Gan, Yifeng Wang, Guo Tang, and Yongbing Zhang. Uncertainty-aware survival analysis with dirichlet distribution for multi-scale pathology and genomics. *IEEE Transactions on Medical Imaging*, 2025.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in neural information processing systems*, 30, 2017.

Ruiqing Li, Xingqi Wu, Ao Li, and Minghui Wang. Hfbsurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. *Bioinformatics*, 38(9):2587–2594, 2022.

Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images, 2020. URL `https://arxiv.org/abs/2004.09666`.

Hui Luo, Jiashuang Huang, Hengrong Ju, Tianyi Zhou, and Weiping Ding. Multimodal multi-instance evidence fusion neural networks for cancer survival prediction. *Scientific Reports*, 15(1): 10470, 2025.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Richard Peto and Julian Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2):185–198, 1972.

Olivier B Poirion, Zheng Jing, Kumardeep Chaudhary, Sijia Huang, and Lana X Garmire. Deepprog: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome medicine*, 13(1):112, 2021.

Wei Shao, Tongxin Wang, Zhi Huang, Zhi Han, Jie Zhang, and Kun Huang. Weakly supervised deep ordinal cox model for survival prediction from whole-slide pathological images. *IEEE Transactions on Medical Imaging*, 40(12):3739–3747, 2021a.

Wei Shao, YangYang Shi, Daoqiang Zhang, JunJie Zhou, and Peng Wan. Tumor micro-environment interactions guided graph learning for survival analysis of human cancers from whole-slide pathological images. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11694–11703. IEEE, 2024a.

Zhimin Shao, Weibei Dou, and Yu Pan. Dual-level deep evidential fusion: Integrating multimodal information for enhanced reliable decision-making in deep learning. *Information Fusion*, 103: 102113, 2024b.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. Transmil: Transformer based correlated multiple instance learning for whole slide image classification, 2021b. URL https://arxiv.org/abs/2106.00908.

Andrew H Song, Richard J Chen, Guillaume Jaume, Anurag J Vaidya, Alexander S Baras, and Faisal Mahmood. Multimodal prototyping for cancer survival prediction. *arXiv preprint arXiv:2407.00224*, 2024.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Xingqi Wu, Yi Shi, Minghui Wang, and Ao Li. Camr: cross-aligned multimodal representation learning for cancer survival prediction. *Bioinformatics*, 39(1):btad025, 2023.

Yucheng Xing, Ling Huang, Jingying Ma, Ruping Hong, Jiangdong Qiu, Pei Liu, Kai He, Huazhu Fu, and Mengling Feng. Dpsurv: Dual-prototype evidential fusion for uncertainty-aware and interpretable whole-slide image survival prediction. *arXiv preprint arXiv:2510.00053*, 2025.

Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp. 21184–21194. IEEE, 2023.

Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis*, 65:101789, October 2020. ISSN 1361-8415. doi: 10.1016/j.media.2020.101789. URL http://dx.doi.org/10.1016/j.media.2020.101789.

Andrew Zhang, Guillaume Jaume, Anurag Vaidya, Tong Ding, and Faisal Mahmood. Accelerating data processing and benchmarking of ai models for pathology. *arXiv preprint arXiv:2502.06750*, 2025.

Wei Zhang, Wenxin Xu, Tong Chen, Collin Sakal, and Xinyue Li. Integrating images and genomics for multi-modal cancer survival analysis via mixture of experts. *Information Fusion*, 126:103521, 2026a.

Yilan Zhang, Yingxue Xu, Jianqi Chen, Fengying Xie, and Hao Chen. Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction. In *International Conference on Learning Representations*, volume 2024, pp. 36854–36876, 2024.

Yilan Zhang, Li Nanbo, Changchun Yang, Jürgen Schmidhuber, and Xin Gao. Structural prognostic event modeling for multimodal cancer survival analysis, 2026b. URL https://arxiv.org/abs/2512.01116.

Fengtao Zhou and Hao Chen. Cross-modal translation and alignment for survival analysis. In *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp. 21428–21437. IEEE, 2023.

Junjie Zhou, Jiao Tang, Yingli Zuo, Peng Wan, Daoqiang Zhang, and Wei Shao. Robust multimodal survival prediction with conditional latent differentiation variational autoencoder. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10384–10393. IEEE, 2025.

APPENDIX

A RELATED WORK

Unimodal Cancer Survival Prediction. Unimodal cancer survival prediction primarily uses WSIs or genomics data as input. BDOCOX introduces ordinal constraints and survival-ranking regularization into a weakly supervised Cox model (Shao et al. (2021a)). Patch-GCN organizes patches into spatial graphs and models tissue structures using graph convolution (Chen et al. (2021a)). MHAttnSurv uses multi-head attention to extract prognostic features from different morphological regions (Jiang et al. (2023)), whereas TMEGL explicitly models spatial relationships between cells and tissue regions in the tumor microenvironment (Shao et al. (2024a)). Cox-nnet employs neural networks to learn nonlinear representations associated with Cox risk from high-dimensional gene expression data (Ching et al. (2018)). Cox-PASNet incorporates gene–pathway associations as network connectivity constraints to improve model interpretability (Hao et al. (2019)). SALMON aggregates expression profiles through gene co-expression modules and combines them with other molecular indicators for survival analysis (Huang et al. (2019)). DeepProg uses autoencoders to extract latent multi-omics features and performs patient risk stratification through survival-related feature selection and ensemble learning (Poirion et al. (2021)). Although these methods model patient prognosis from either tissue morphology or molecular states, a single modality cannot fully capture the multilevel biological characteristics of tumors.

Multimodal Cancer Survival Prediction. Multimodal cancer survival prediction aims to jointly leverage WSIs and genomics data to characterize patient prognosis from the perspectives of tissue morphology and molecular states. Early methods performed multimodal fusion through feature concatenation or tensor operations. PORPOISE integrates histological and molecular information for pan-cancer survival prediction while providing cross-modal interpretations (Chen et al. (2022)). HFBSurv progressively combines pathology and gene expression data using a hierarchical factorized bilinear structure, thereby reducing the parameter complexity of high-order fusion (Li et al. (2022)). CAMR decomposes multimodal representations into modality-invariant and modality-specific components and employs distribution alignment to reduce representational discrepancies (Wu et al. (2023)). Recent studies have increasingly focused on fine-grained cross-modal interactions. MOTCat constrains cross-modal matching via optimal transport (Xu & Chen (2023)). PathOmics adopts unsupervised pretraining to learn intrinsic associations between pathology and multi-omics data (Ding et al. (2023)). MCTI introduces tumor subtype classification to assist pathology region learning and employs transfer-guided attention for cross-task and cross-modal interaction (Jiang et al. (2024)). SlotSPE uses slot attention to compress pathology and genomics inputs into patient-specific structured prognostic events, enabling compact modeling of both intra-modal and inter-modal relationships (Zhang et al. (2026b)). These methods primarily focus on joint representations of pathology and omics, rarely explicitly modeling the prognostic evidence provided by different modalities at the decision level to mitigate the impact of modality conflicts on the final prediction.

Evidence Learning. Evidential learning provides a new perspective for characterizing the degree of predictive support and uncertainty. TMC constructs dirichlet opinions for individual views and combines multi-view evidence using Dempster–Shafer theory, demonstrating that predictions from different data sources can be represented and fused in the decision space (Han et al. (2022)). DDEF performs evidence fusion at both the basic belief assignment and multimodal levels to improve the accuracy and reliability of multi-source decisions (Shao et al. (2024b)). Evidential learning has also been introduced into cancer survival prediction. ETTE develops an evidential time-to-event model using Gaussian random fuzzy numbers to estimate epistemic and aleatoric uncertainties in event-time prediction (Huang et al. (2025)). EsurvFusion separately estimates survival outcomes and their reliability for each modality, and then fuses reliability-adjusted unimodal predictions at the decision level (Huang et al. (2024)). M2EF-NNs employs Dempster–Shafer theory to dynamically adjust modality weights in multimodal survival prediction (Luo et al. (2025)). UMSA uses Dirichlet distributions to model predictive uncertainty and integrates multiscale pathology images with genomics data (Jiang et al. (2025)).

B Analysis of Consensus–Complementarity Evidence Decomposition

As defined in Equation 10, $\mathbf{e}_w$ and $\mathbf{e}_o$ are the nonnegative evidence vectors of WSI and genomics, respectively. Let $e_{w,k}$ and $e_{o,k}$ denote the pathology and genomics evidence in the k-th time interval, respectively. Since $e_{w,k}$ and $e_{o,k}$ are nonnegative evidence values defined over the same time interval, the evidence jointly supported by the two modalities should not exceed the evidence provided by either modality. Their minimum is the largest nonnegative value satisfying this constraint. Therefore, for any time interval k, we define the minimum of the pathology and genomics evidence as the consensus evidence:

$$e_{\mathrm{con},k} = \min(e_{w,k}, e_{o,k}) \tag{29}$$

It follows that:

$$0 \le e_{\mathrm{con},k} \le e_{w,k}, \qquad 0 \le e_{\mathrm{con},k} \le e_{o,k} \tag{30}$$

The evidence exceeding this consensus component is defined as the modality-specific complementary evidence:

$$r_{w,k} = \max(e_{w,k} - e_{o,k}, 0), \qquad r_{o,k} = \max(e_{o,k} - e_{w,k}, 0) \tag{31}$$

Both $\mathbf{r}_w$ and $\mathbf{r}_o$ are nonnegative vectors. For any time interval k, at most one of $r_{w,k}$ and $r_{o,k}$ is nonzero, so $r_{w,k} r_{o,k} = 0$. Moreover, the sum of the complementary evidence from the two modalities satisfies:

$$\mathbf{r}_w + \mathbf{r}_o = |\mathbf{e}_w - \mathbf{e}_o| \tag{32}$$

CCED therefore decomposes the pathology and genomics evidence into a jointly supported consensus component and modality-specific complementary components while preserving the original modality evidence:

$$\mathbf{e}_w = \mathbf{e}_{\mathrm{con}} + \mathbf{r}_w, \qquad \mathbf{e}_o = \mathbf{e}_{\mathrm{con}} + \mathbf{r}_o \tag{33}$$

The sum of the pathology and genomics evidence is:

$$\mathbf{e}_w + \mathbf{e}_o = 2\mathbf{e}_{\mathrm{con}} + \mathbf{d} \tag{34}$$

where $\mathbf{d} = \mathbf{r}_w + \mathbf{r}_o = |\mathbf{e}_w - \mathbf{e}_o|$ represents the non-consensus evidence between the two modalities, and $D = \|\mathbf{d}\|_1$ denotes its total amount. Under direct evidence summation, all non-consensus evidence of magnitude D enters the combined result. By explicitly separating consensus and complementary evidence, CCED enables the two components to be handled separately in subsequent stages.

C Analysis of Consensus Evidence Enhancement

Let the total amount of consensus evidence be:

$$E_{\mathrm{con}} = \|\mathbf{e}_{\mathrm{con}}\|_1 = \sum_{k=1}^{K} e_{\mathrm{con},k} \tag{35}$$

CEE applies an element-wise power transformation to the consensus evidence:

$$\mathbf{e}_{\mathrm{con}}^{(\gamma)} = \mathbf{e}_{\mathrm{con}}^{\gamma}, \qquad \gamma > 1 \tag{36}$$

The transformed evidence is then normalized according to the original total amount of consensus evidence:

$$\widetilde{\mathbf{e}}_{\mathrm{con}} = E_{\mathrm{con}} \frac{\mathbf{e}_{\mathrm{con}}^{(\gamma)}}{\left\|\mathbf{e}_{\mathrm{con}}^{(\gamma)}\right\|_1} \tag{37}$$

When $E_{\mathrm{con}} = 0$, we set $\widetilde{\mathbf{e}}_{\mathrm{con}} = \mathbf{0}$. When $E_{\mathrm{con}} > 0$, the normalized evidence satisfies:

$$\|\widetilde{\mathbf{e}}_{\mathrm{con}}\|_1 = E_{\mathrm{con}} \tag{38}$$

Because consensus evidence is supported by both pathology and genomics, CEE scales the normalized consensus evidence by a factor of two $\mathbf{e}_{\mathrm{con}}^{+} = 2\widetilde{\mathbf{e}}_{\mathrm{con}}$ and therefore $\|\mathbf{e}_{\mathrm{con}}^{+}\|_1 = 2E_{\mathrm{con}}$. This

equals the total amount of the consensus component, $2\mathbf{e}_{\text{con}}$, obtained by directly summing the evidence from the two modalities. Thus, CEE does not increase the total amount of jointly supported evidence but adjusts its relative distribution across time intervals. For any two intervals satisfying $e_{\text{con},a} > e_{\text{con},b} > 0$, the ratio of their transformed evidence is:

$$\frac{\widetilde{e}_{\text{con},a}}{\widetilde{e}_{\text{con},b}} = \left(\frac{e_{\text{con},a}}{e_{\text{con},b}}\right)^{\gamma} \tag{39}$$

Since $\gamma > 1$ and $\frac{e_{\text{con},a}}{e_{\text{con},b}} > 1$, the following inequality holds:

$$\left(\frac{e_{\text{con},a}}{e_{\text{con},b}}\right)^{\gamma} > \frac{e_{\text{con},a}}{e_{\text{con},b}} \tag{40}$$

Therefore, the power transformation increases the relative difference between intervals with strong and weak consensus evidence. While preserving the total amount of consensus evidence, CEE increases the relative contribution of strongly supported intervals and reduces that of weakly supported intervals, thereby redistributing the evidence across time intervals.

D Analysis of Joint Evidential Survival

The preceding analysis shows that CCED decomposes the pathology and genomics evidence into consensus and complementary components. CEE then produces the enhanced consensus evidence while preserving its total amount $\mathbf{e}^{+}_{\text{con}} = 2\widetilde{\mathbf{e}}_{\text{con}}$. JES scales the complementary evidence from the two modalities using the vacuities estimated by DVE:

$$\mathbf{e}_{\text{cmp}} = V_w \mathbf{r}_w + V_o \mathbf{r}_o \tag{41}$$

Since $\mathbf{r}_w \geq \mathbf{0}$, $\mathbf{r}_o \geq \mathbf{0}$, and $0 < V_w, V_o \leq 1$, it follows that:

$$\mathbf{0} \leq \mathbf{e}_{\text{cmp}} \leq \mathbf{r}_w + \mathbf{r}_o = \mathbf{d} \tag{42}$$

Let the total amount of the scaled complementary evidence be $E_{\text{cmp}} = \|\mathbf{e}_{\text{cmp}}\|_1$. Since all evidence vectors are nonnegative, we have:

$$\begin{aligned} E_{\text{cmp}} &= V_w \|\mathbf{r}_w\|_1 + V_o \|\mathbf{r}_o\|_1 \\ &\leq \|\mathbf{r}_w\|_1 + \|\mathbf{r}_o\|_1 \\ &= D \end{aligned} \tag{43}$$

where $D = \|\mathbf{d}\|_1$ denotes the total amount of non-consensus evidence between the two modalities. JES combines the enhanced consensus evidence and the scaled complementary evidence to form the joint evidence:

$$\mathbf{e}_j = \mathbf{e}^{+}_{\text{con}} + \mathbf{e}_{\text{cmp}} \tag{44}$$

The total amount of joint evidence is:

$$\|\mathbf{e}_j\|_1 = 2E_{\text{con}} + E_{\text{cmp}} \tag{45}$$

where $E_{\text{con}} = \|\mathbf{e}_{\text{con}}\|_1$. In contrast, directly summing the pathology and genomics evidence gives $\|\mathbf{e}_w + \mathbf{e}_o\|_1 = 2E_{\text{con}} + D$. Since $E_{\text{cmp}} \leq D$, it follows that:

$$\|\mathbf{e}_j\|_1 \leq \|\mathbf{e}_w + \mathbf{e}_o\|_1 \tag{46}$$

Both combination strategies contain the same amount of consensus evidence, $2E_{\text{con}}$, and differ only in the amount of non-consensus evidence entering the joint result. Direct summation retains all non-consensus evidence of magnitude D, whereas JES scales it to E_{cmp}. It preserves complementary information while limiting the amount of non-consensus evidence incorporated into the joint result.

D.1 Upper Bound on the Shift in the Event-Time Distribution

To isolate the effect of complementary evidence on the event-time distribution, we remove the complementary component from the joint evidence and retain only the enhanced consensus evidence, based on which a reference Dirichlet distribution is defined. Its parameter is $\boldsymbol{\alpha}_{\text{con}} = \mathbf{e}^{+}_{\text{con}} + \mathbf{1}$. Let $E_{\text{con}} = \|\mathbf{e}_{\text{con}}\|_1$. According to the definition of CEE, the enhanced consensus evidence satisfies:

$$\left\|\mathbf{e}^{+}_{\text{con}}\right\|_1 = 2E_{\text{con}} \tag{47}$$

The strength of the reference Dirichlet distribution is therefore:

$$B = \|\boldsymbol{\alpha}_{\text{con}}\|_1 = 2E_{\text{con}} + K \tag{48}$$

The reference Dirichlet distribution characterizes the uncertainty over possible event-time probability vectors. To obtain a deterministic event-time prediction, we use its expectation. Accordingly, the predicted event-time distribution is given by the Dirichlet mean, $\mathbf{p}_{\text{con}} = \frac{\boldsymbol{\alpha}_{\text{con}}}{B}$. After incorporating the scaled complementary evidence, the joint Dirichlet parameter becomes $\boldsymbol{\alpha}_{\text{evi}} = \boldsymbol{\alpha}_{\text{con}} + \mathbf{e}_{\text{cmp}}$. Since $E_{\text{cmp}} = \|\mathbf{e}_{\text{cmp}}\|_1$, the strength of the joint Dirichlet distribution is $B + E_{\text{cmp}}$. Therefore, the corresponding predicted event-time distribution is defined as:

$$\mathbf{p}_{\text{evi}} = \frac{\boldsymbol{\alpha}_{\text{con}} + \mathbf{e}_{\text{cmp}}}{B + E_{\text{cmp}}} \tag{49}$$

The difference between the joint and reference event-time distributions can be defined as:

$$\mathbf{p}_{\text{evi}} - \mathbf{p}_{\text{con}} = \frac{\mathbf{e}_{\text{cmp}}}{B + E_{\text{cmp}}} - \frac{E_{\text{cmp}}\boldsymbol{\alpha}_{\text{con}}}{B(B + E_{\text{cmp}})} \tag{50}$$

Applying the triangle inequality and using $\|\mathbf{e}_{\text{cmp}}\|_1 = E_{\text{cmp}}$ and $\|\boldsymbol{\alpha}_{\text{con}}\|_1 = B$, we obtain:

$$\begin{aligned} \|\mathbf{p}_{\text{evi}} - \mathbf{p}_{\text{con}}\|_1 &\leq \frac{\|\mathbf{e}_{\text{cmp}}\|_1}{B + E_{\text{cmp}}} + \frac{E_{\text{cmp}}\,\|\boldsymbol{\alpha}_{\text{con}}\|_1}{B(B + E_{\text{cmp}})} \\ &= \frac{E_{\text{cmp}}}{B + E_{\text{cmp}}} + \frac{E_{\text{cmp}}}{B + E_{\text{cmp}}} \\ &= \frac{2E_{\text{cmp}}}{B + E_{\text{cmp}}} \end{aligned} \tag{51}$$

Therefore, the total variation distance between the two event-time distributions satisfies:

$$\text{TV}\left(\mathbf{p}_{\text{evi}}, \mathbf{p}_{\text{con}}\right) = \frac{1}{2}\|\mathbf{p}_{\text{evi}} - \mathbf{p}_{\text{con}}\|_1 \leq \frac{E_{\text{cmp}}}{B + E_{\text{cmp}}} \tag{52}$$

This bound shows that the distributional shift induced by the complementary evidence is controlled by its total amount E_{cmp}. A larger Dirichlet strength B associated with the enhanced consensus evidence reduces the relative influence of the complementary evidence on the event-time distribution.

D.2 Comparison with the Case Without Vacuity Scaling

With the enhanced consensus evidence held fixed, all complementary evidence enters the joint result when vacuity scaling is not applied:

$$\mathbf{e}^{\text{nv}}_{\text{cmp}} = \mathbf{r}_w + \mathbf{r}_o = \mathbf{d} \tag{53}$$

where $\mathbf{e}^{\text{nv}}_{\text{cmp}}$ denotes the complementary evidence without vacuity scaling, and $\left\|\mathbf{e}^{\text{nv}}_{\text{cmp}}\right\|_1 = D$. The corresponding event-time probability mass function is:

$$\mathbf{p}_{\text{nv}} = \frac{\boldsymbol{\alpha}_{\text{con}} + \mathbf{d}}{B + D} \tag{54}$$

Following a derivation similar to that of Equation 50, we obtain:

$$\text{TV}\left(\mathbf{p}_{\text{nv}}, \mathbf{p}_{\text{con}}\right) \leq \frac{D}{B + D} \tag{55}$$

Since $E_{\text{cmp}} \leq D$ and the function $f(x) = \frac{x}{B+x}$ is monotonically increasing for $x \geq 0$ and $B > 0$, it follows that:

$$\frac{E_{\text{cmp}}}{B + E_{\text{cmp}}} \leq \frac{D}{B + D} \tag{56}$$

Thus, with the enhanced consensus evidence held fixed, the theoretical upper bound on the distributional change induced by non-consensus evidence under vacuity scaling is no greater than that obtained without scaling.

D.3 Upper Bound on the Shift in the Evidence-Path Survival Curve

The survival curve produced by the Evidence-Aware Path is obtained by accumulating the event-time distribution:

$$S_{\mathrm{evi}}(k) = 1 - \sum_{t=1}^{k} p_{\mathrm{evi}}(t) \tag{57}$$

The survival curve corresponding to the consensus reference distribution is defined as:

$$S_{\mathrm{con}}(k) = 1 - \sum_{t=1}^{k} p_{\mathrm{con}}(t) \tag{58}$$

For any discrete time interval k, the difference between the two survival probabilities is bounded as:

$$\begin{aligned} |S_{\mathrm{evi}}(k) - S_{\mathrm{con}}(k)| &= \left| \sum_{t=1}^{k} [p_{\mathrm{con}}(t) - p_{\mathrm{evi}}(t)] \right| \\ &\leq \mathrm{TV}\left(\mathbf{p}_{\mathrm{evi}}, \mathbf{p}_{\mathrm{con}}\right) \\ &\leq \frac{E_{\mathrm{cmp}}}{B + E_{\mathrm{cmp}}}. \end{aligned} \tag{59}$$

Without vacuity scaling, let the corresponding survival curve be denoted by $S_{\mathrm{nv}}(k)$. It satisfies:

$$|S_{\mathrm{nv}}(k) - S_{\mathrm{con}}(k)| \leq \frac{D}{B + D} \tag{60}$$

Together with Equation 56, this shows that the upper bound on the shift of the Evidence-Aware Path survival curve induced by non-consensus evidence under vacuity scaling is no greater than the corresponding bound without scaling. Overall, CCED separates consensus from non-consensus evidence, CEE redistributes the consensus evidence across time intervals, and DVE and JES regulate the amount of non-consensus evidence incorporated into the joint prediction.

E Analysis of the Conflict Response of Dual-Path Fusion

The Evidence-Aware Path does not alter the feature interactions within the Representation Path. Instead, it produces an independent decision-level survival curve. To analyze the response of their fusion to modality conflict, we consider the predictions before and after introducing the conflict. Under the original input, let the survival curves produced by the Representation Path and the Evidence-Aware Path be $S_{\mathrm{rep}}(k)$ and $S_{\mathrm{evi}}(k)$, respectively. After introducing modality conflict, the corresponding curves are denoted by $S'_{\mathrm{rep}}(k)$ and $S'_{\mathrm{evi}}(k)$. Their responses to the conflict at time interval k are defined as:

$$\Delta_{\mathrm{rep}}(k) = \left|S'_{\mathrm{rep}}(k) - S_{\mathrm{rep}}(k)\right|, \qquad \Delta_{\mathrm{evi}}(k) = |S'_{\mathrm{evi}}(k) - S_{\mathrm{evi}}(k)| \tag{61}$$

The final survival curves before and after introducing modality conflict are given by:

$$\begin{aligned} S_{\mathrm{final}}(k) &= \omega S_{\mathrm{rep}}(k) + (1-\omega) S_{\mathrm{evi}}(k), \\ S'_{\mathrm{final}}(k) &= \omega S'_{\mathrm{rep}}(k) + (1-\omega) S'_{\mathrm{evi}}(k) \end{aligned} \tag{62}$$

where $\omega = \sigma(\theta)$ is a learnable fusion weight. The response of the final survival curve at interval k is defined as:

$$\Delta_{\mathrm{final}}(k) = |S'_{\mathrm{final}}(k) - S_{\mathrm{final}}(k)| \tag{63}$$

Substituting the two final survival curves and applying the triangle inequality yields:

$$\begin{aligned} \Delta_{\mathrm{final}}(k) &= \left|\omega \left[S'_{\mathrm{rep}}(k) - S_{\mathrm{rep}}(k)\right] + (1-\omega) \left[S'_{\mathrm{evi}}(k) - S_{\mathrm{evi}}(k)\right]\right| \\ &\leq \omega \left|S'_{\mathrm{rep}}(k) - S_{\mathrm{rep}}(k)\right| + (1-\omega) \left|S'_{\mathrm{evi}}(k) - S_{\mathrm{evi}}(k)\right| \\ &= \omega \Delta_{\mathrm{rep}}(k) + (1-\omega) \Delta_{\mathrm{evi}}(k) \end{aligned} \tag{64}$$

Thus, the response magnitude of the final prediction is bounded by a convex combination of the response magnitudes of the two paths. If $\Delta_{\mathrm{evi}}(k) \leq \Delta_{\mathrm{rep}}(k)$, then:

$$\begin{aligned} \Delta_{\mathrm{final}}(k) &\leq \omega \Delta_{\mathrm{rep}}(k) + (1-\omega) \Delta_{\mathrm{evi}}(k) \\ &\leq \Delta_{\mathrm{rep}}(k) \end{aligned} \tag{65}$$

Therefore, when the Evidence-Aware Path is less responsive to the introduced conflict than the Representation Path, dual-path fusion reduces the sensitivity of the final prediction relative to using the Representation Path alone. As shown in Appendix A, the Evidence-Aware Path separates consensus and complementary evidence, enhances the consensus component, and applies vacuity-based scaling to regulate the influence of non-consensus evidence on the event-time distribution and the resulting survival curve. Dual-path fusion then incorporates this decision-level prediction into the final output, preventing modality conflict within the Representation Path from determining the final prediction without adjustment. In this way, the Evidence-Aware Path preserves cross-modal prognostic information while providing an additional decision path that is explicitly regulated with respect to modality conflict.

F EXPERIMENT DETAILS

All experiments are implemented in PyTorch (Paszke et al. (2019)) and conducted on a single NVIDIA 5090 GPU. The network is trained using the Adam optimizer (Kingma & Ba (2014)) with an initial learning rate of 2×10^{-4}, a weight decay of 1×10^{-5}, and a batch size of 32. Each fold is trained for 30 epochs with the random seed fixed at 42. For the pathology modality, we follow the Trident (Zhang et al. (2025)) preprocessing pipeline by dividing each WSI into 256×256 patches and extracting 1,024-dimensional features using UNI encoder (Chen et al. (2024)) and the extracted features are projected to 256 dimensions through a linear mapping. The genomics data are organized according to biological pathways and mapped to 256 dimensions. Both the pathology and genomics branches use eight slots, with the number of slot attention iterations set to 10. The Top-K selection ratio is set to 0.25, retaining the two slots with the highest responses from each modality for survival prediction. The power exponent used in CCE is set to $\gamma = 1.25$.

G COMPARATIVE EXPERIMENT

To evaluate the survival prediction performance of the proposed method, we compare it with both unimodal and multimodal methods. MLP applies a multilayer perceptron directly to genomics features for survival prediction (Haykin (1994)); SNN uses a self-normalizing neural network to learn stable representations from high-dimensional genomics data (Klambauer et al. (2017)), and SNNTrans extends SNN with a self-attention module to model dependencies among genomics features.(Klambauer et al. (2017)). These methods use only genomics data. In contrast, several methods rely solely on WSIs: ABMIL uses learnable attention pooling to aggregate patch-level features into a slide-level representation (Ilse et al. (2018)), TransMIL employs a Transformer-based MIL architecture to model correlations among WSI patches (Shao et al. (2021b)), CLAM-SB combines single-branch attention pooling with instance-level clustering constraint (Lu et al. (2020)), and DPsurv uses dual prototypes and evidential fusion to obtain uncertainty-aware and interpretable survival predictions from WSIs(Xing et al. (2025)). Finally, several methods utilize both WSIs and genomics data. PORPOISE combines attention-aggregated WSI features with genomics representations through a weakly supervised multimodal fusion framework (Chen et al. (2022)), MCAT uses co-attention to model dense interactions between genomics features and histology patches before Transformer-based aggregation (Chen et al. (2021b)), MOTCat introduces optimal-transport-based co-attention to align pathology and genomics instances while preserving their global structural relationship (Xu & Chen (2023)), CMTA uses parallel encoder–decoder structures and cross-modal attention to perform cross-modal translation and alignment (Zhou & Chen (2023)), SurvPath organizes transcriptomics data into biological pathway tokens and models their interactions with histology patches using a memory-efficient multimodal Transformer (Jaume et al. (2024)), PIBD uses prototypical information bottlenecking and disentanglement to reduce intra-modal and inter-modal redundancy (Zhang et al. (2024)), MMP represents WSIs and transcriptomics data using morphological and biological pathway prototypes before multimodal fusion. (Song et al. (2024)), LD-CVAE learns pathology and function-specific genomics distributions through a conditional variational autoencoder and supports survival prediction when genomics data are missing (Zhou et al. (2025)), SurMoE combines pathology clustering, pathway-based genomics encoding, cross-modal attention, and a mixture-of-experts architecture for multimodal survival prediction (Zhang et al. (2026a)), and SlotSPE uses slot attention to represent sparse patient-specific prognostic events and model their intra-modal and inter-modal interactions (Zhang et al. (2026b)).

All methods are evaluated using the same five-fold data splits. Tables 4 and 5 report the complete C-index results as the mean and standard deviation over the five folds. With both modalities available, DualRE-Surv achieves the highest mean C-index on all eight cancer cohorts, reaching 0.6477, 0.7192, 0.7318, 0.8114, 0.8120, 0.6104, 0.6830, and 0.6570 on HNSC, BLCA, BRCA, UCEC, KIRC, LUSC, SKCM, and STAD. Compared with the second-best methods, DualRE-Surv yields relative C-index improvements of 0.92%, 1.86%, 1.47%, 1.30%, 0.52%, 1.09%, 0.41%, and 1.06%, respectively.

Table 4: Complete compare results on HNSC, BLCA, BRCA, and UCEC. G. and P. denote genomics and pathology. **Bold** and underlined indicate the best and second-best results.

Method	Modality		C-index			
	G.	P.	HNSC	BLCA	BRCA	UCEC
MLP	✓		0.5674 ± 0.0512	0.6570 ± 0.0449	0.6443 ± 0.0837	0.7054 ± 0.0318
SNN	✓		0.5987 ± 0.0414	0.6080 ± 0.0328	0.6368 ± 0.0567	0.7570 ± 0.0411
SNNTrans	✓		0.5896 ± 0.0409	0.6254 ± 0.0558	0.6482 ± 0.0631	0.7432 ± 0.0409
DualRE-Surv	✓		0.6402 ± 0.0428	0.6704 ± 0.0582	0.6792 ± 0.0566	0.7822 ± 0.0548
ABMIL		✓	0.6213 ± 0.0511	0.6421 ± 0.0368	0.6490 ± 0.0326	0.7794 ± 0.0632
TransMIL		✓	0.6055 ± 0.0546	0.6677 ± 0.0687	0.6411 ± 0.0324	0.7800 ± 0.0377
CLAM-SB		✓	0.6005 ± 0.0397	0.6704 ± 0.0572	0.6681 ± 0.0501	0.7836 ± 0.0459
DPsurv		✓	0.5763 ± 0.1373	0.6227 ± 0.0373	0.6661 ± 0.0946	0.6358 ± 0.1156
DualRE-Surv		✓	0.6336 ± 0.0579	0.6667 ± 0.0344	0.6806 ± 0.0349	0.7917 ± 0.0390
PORPOISE	✓	✓	0.5998 ± 0.0470	0.6398 ± 0.0596	0.6541 ± 0.0503	0.7953 ± 0.0740
MCAT	✓	✓	0.6125 ± 0.0468	0.6679 ± 0.0785	0.6812 ± 0.0977	0.7986 ± 0.0575
MOTCat	✓	✓	0.6137 ± 0.0523	0.6896 ± 0.0437	0.7008 ± 0.0574	0.7809 ± 0.0604
CMTA	✓	✓	0.6249 ± 0.0445	0.6783 ± 0.0290	0.7196 ± 0.0688	0.8003 ± 0.0189
SurvPath	✓	✓	0.6036 ± 0.0368	0.6935 ± 0.0556	0.6754 ± 0.0601	0.7991 ± 0.0204
PIBD	✓	✓	0.6368 ± 0.0365	0.6821 ± 0.0377	0.6812 ± 0.0512	0.7812 ± 0.0355
MMP	✓	✓	0.6225 ± 0.0477	0.7033 ± 0.0534	$\underline{0.7212 \pm 0.0479}$	0.7654 ± 0.0753
LD-CVAE	✓	✓	0.6198 ± 0.0509	0.6850 ± 0.0479	0.7135 ± 0.0633	$\underline{0.8010 \pm 0.0246}$
SurMoE	✓	✓	0.6334 ± 0.0788	0.6689 ± 0.0308	0.7000 ± 0.0217	0.7582 ± 0.0273
SlotSPE	✓	✓	$\underline{0.6418 \pm 0.0345}$	$\underline{0.7061 \pm 0.0225}$	0.7083 ± 0.0607	0.7975 ± 0.0637
DualRE-Surv	✓	✓	$\mathbf{0.6477 \pm 0.0382}$	$\mathbf{0.7192 \pm 0.0302}$	$\mathbf{0.7318 \pm 0.0650}$	$\mathbf{0.8114 \pm 0.0604}$

H Ablation experiment

To evaluate the Evidence-Aware Path, we conduct four full-modality ablations by removing the entire path, DVE, CCED, and CEE, respectively. Table 6 present the complete ablation study results of DualRE-Surv across all eight cohorts. Removing the entire path or any of its components reduces performance across all eight cohorts, confirming their contributions to the model.

I Survival risk stratification analysis

Patients were stratified into high- and low-risk groups based on the median predicted risk. Kaplan–Meier curves (Kaplan & Meier (1958)) and stratified log-rank tests (Peto & Peto (1972)) were used to assess the separation between the two groups. DualRE-Surv achieved statistically significant separation across all eight TCGA cohorts and consistently yielded smaller stratified log-rank p-values than SlotSPE and LD-CVAE. The complete results are shown in Figures 7.

J Modality Conflict Analysis

For modality $m \in \{w, o\}$, the mean of its Dirichlet distribution is:

$$\mathbf{p}_m = \frac{\boldsymbol{\alpha}_m}{S_m} \tag{66}$$

Table 5: Complete compare results on KIRC, LUSC, SKCM, and STAD. G. and P. denote genomics and pathology. **Bold** and underlined indicate the best and second-best results.

Method	Modality		C-index			
	G.	P.	KIRC	LUSC	SKCM	STAD
MLP	✓		0.7643 ± 0.0412	0.5307 ± 0.0652	0.6320 ± 0.0714	0.5485 ± 0.0772
SNN	✓		0.7322 ± 0.0380	0.5681 ± 0.0468	0.6159 ± 0.0329	0.5278 ± 0.0523
SNNTrans	✓		0.7082 ± 0.0293	0.5612 ± 0.0454	0.6288 ± 0.0547	0.5857 ± 0.0453
DualRE-Surv	✓		0.7906 ± 0.0238	0.5935 ± 0.0372	0.6624 ± 0.0208	0.6298 ± 0.0320
ABMIL		✓	0.7367 ± 0.0436	0.5275 ± 0.0133	0.6405 ± 0.0367	0.6036 ± 0.0327
TransMIL		✓	0.7751 ± 0.0297	0.5563 ± 0.0360	0.6098 ± 0.0104	0.5983 ± 0.0295
CLAM-SB		✓	0.6908 ± 0.0616	0.5173 ± 0.0417	0.6431 ± 0.0245	0.6145 ± 0.0467
DPsurv		✓	0.6942 ± 0.1034	0.5322 ± 0.0954	0.5978 ± 0.0809	0.6043 ± 0.0981
DualRE-Surv		✓	0.7847 ± 0.0084	0.6064 ± 0.0482	0.6593 ± 0.0528	0.6471 ± 0.0185
PORPOISE	✓	✓	0.7149 ± 0.0189	0.5988 ± 0.0322	0.6659 ± 0.0496	0.6239 ± 0.0770
MCAT	✓	✓	0.8017 ± 0.0214	0.5764 ± 0.0193	0.6580 ± 0.0301	0.6348 ± 0.0221
MOTCat	✓	✓	0.7868 ± 0.0103	0.5868 ± 0.0457	0.6577 ± 0.0252	0.6308 ± 0.0374
CMTA	✓	✓	0.7915 ± 0.0246	0.5760 ± 0.0326	$\underline{0.6802 \pm 0.0364}$	0.6417 ± 0.0286
SurvPath	✓	✓	0.7595 ± 0.0019	0.5901 ± 0.0705	0.6594 ± 0.0270	0.6296 ± 0.0584
PIBD	✓	✓	0.7787 ± 0.0457	0.5895 ± 0.0611	0.6531 ± 0.0311	$\underline{0.6501 \pm 0.0442}$
MMP	✓	✓	0.8046 ± 0.0370	0.5793 ± 0.0514	0.6653 ± 0.0632	0.6377 ± 0.0573
LD-CVAE	✓	✓	0.7954 ± 0.0276	$\underline{0.6038 \pm 0.0357}$	0.6688 ± 0.0436	0.6469 ± 0.0859
SurMoE	✓	✓	0.7660 ± 0.0509	0.5964 ± 0.0437	0.6572 ± 0.0490	0.6295 ± 0.0117
SlotSPE	✓	✓	$\underline{0.8078 \pm 0.0179}$	0.5775 ± 0.0604	0.6694 ± 0.0472	0.6337 ± 0.0093
DualRE-Surv	✓	✓	$\mathbf{0.8120 \pm 0.0156}$	$\mathbf{0.6104 \pm 0.0357}$	$\mathbf{0.6830 \pm 0.0485}$	$\mathbf{0.6570 \pm 0.0183}$

Table 6: Complete ablation study results of DualRE-Surv across all eight TCGA cohorts.

Variant	C-index			
	HNSC	BLCA	BRCA	UCEC
w/o Evidence-Aware Path	0.6379 ± 0.0461	0.7005 ± 0.0377	0.7029 ± 0.0740	0.8016 ± 0.0918
w/o DVE	0.6397 ± 0.0585	0.7159 ± 0.0245	0.7250 ± 0.0629	0.8040 ± 0.0645
w/o CCED	0.6413 ± 0.0552	0.7137 ± 0.0261	0.7001 ± 0.0403	0.8023 ± 0.0472
w/o CEE	0.6467 ± 0.0431	0.7087 ± 0.0148	0.7307 ± 0.0296	0.8082 ± 0.0623
DualRE-Surv	$\mathbf{0.6477 \pm 0.0382}$	$\mathbf{0.7192 \pm 0.0302}$	$\mathbf{0.7318 \pm 0.0650}$	$\mathbf{0.8114 \pm 0.0604}$

Variant	C-index			
	KIRC	LUSC	SKCM	STAD
w/o Evidence-Aware Path	0.7902 ± 0.0353	0.5829 ± 0.0510	0.6703 ± 0.0574	0.6418 ± 0.0133
w/o DVE	0.7997 ± 0.0191	0.6038 ± 0.0370	0.6732 ± 0.0482	0.6504 ± 0.0204
w/o CCED	0.8099 ± 0.0173	0.6077 ± 0.0647	0.6799 ± 0.0636	0.6366 ± 0.0182
w/o CEE	0.8105 ± 0.0182	0.5988 ± 0.0467	0.6765 ± 0.0368	0.6547 ± 0.0166
DualRE-Surv	$\mathbf{0.8120 \pm 0.0156}$	$\mathbf{0.6104 \pm 0.0357}$	$\mathbf{0.6830 \pm 0.0485}$	$\mathbf{0.6570 \pm 0.0183}$

where $\boldsymbol{\alpha}_m$ denotes the dirichlet parameters and $S_m = \sum_{k=1}^{K} \alpha_{m,k}$ is the corresponding dirichlet strength. The vector $\mathbf{p}_m$ represents the event-time distribution predicted by modality m. The patient-level modality conflict score is defined as:

$$C = \frac{\mathrm{JS}(\mathbf{p}_w, \mathbf{p}_o)}{\log 2} \sqrt{(1 - V_w)(1 - V_o)} \tag{67}$$

where JS$(\cdot)$ denotes the Jensen–Shannon divergence, which measures the discrepancy between the event-time distributions predicted by the two modalities. A larger value indicates greater disagree-

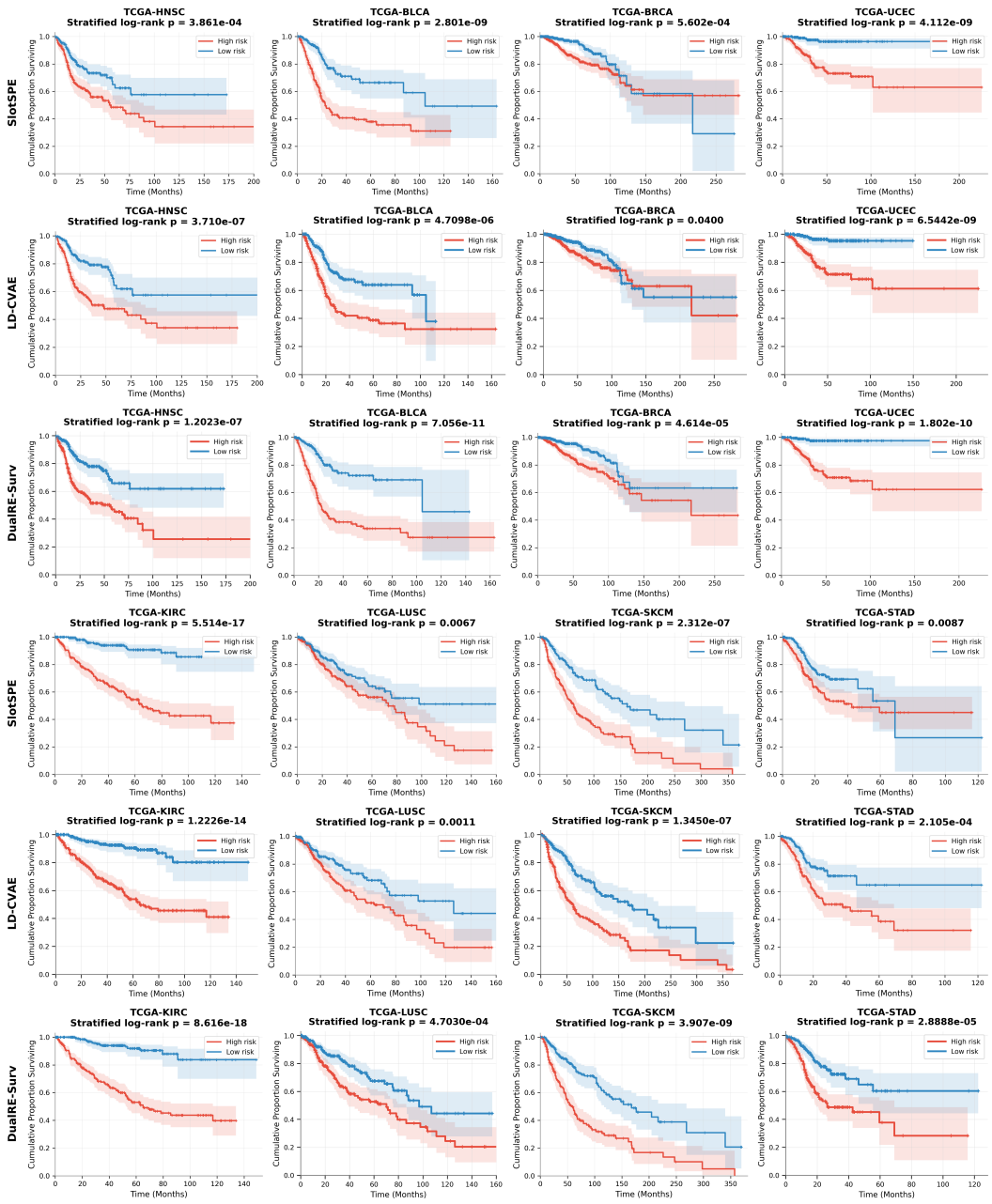

Figure 7: Risk stratification results of SlotSPE, LD-CVAE, and DualRE-Surv on eight TCGA cohorts. The red and blue curves represent the high-risk and low-risk groups, respectively.

ment between the two modalities. Patients are ranked according to C and divided into low-conflict, medium-conflict, and high-conflict groups. To evaluate how closely the three prediction paths match the observed survival outcomes, we use the patient-level discrete-time survival NLL. This metric directly evaluates the predicted survival distribution while accounting for both observed events and censored samples. For prediction path $q \in \{\mathrm{rep}, \mathrm{evi}, \mathrm{final}\}$, its mean NLL in conflict group g is defined as:

$$\overline{\mathcal{L}}_{q,g} = \frac{1}{N_g} \sum_{i \in g} \mathcal{L}_{\mathrm{NLL},q}^{(i)} \tag{68}$$

where N_g is the number of patients in group g, and $\mathcal{L}^{(i)}_{\mathrm{NLL},q}$ is the survival NLL of patient i produced by path q. The complete results are shown in Figure 8. Across the eight cohorts, the final prediction obtains a lower mean NLL than the Representation Path. The same pattern is retained in the high-conflict groups, indicating that the contribution of the Evidence-Aware Path remains useful when pathology and genomics produce more discrepant predictions.

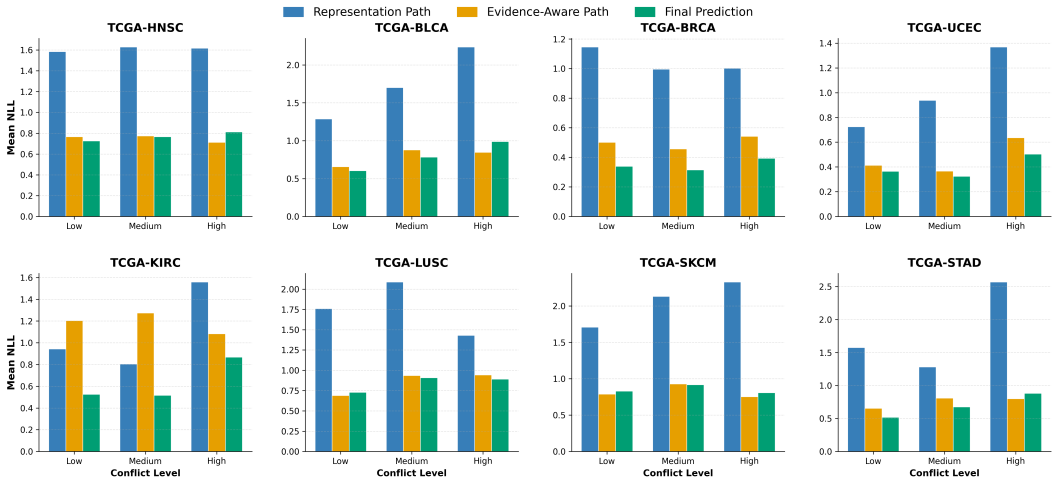

Figure 8: Comparison of the mean NLLs of the Representation Path, Evidence-Aware Path, and final prediction on eight TCGA cohorts.

K Conflict Perturbation Analysis

To examine the response of the Evidence-Aware Path to controlled modality conflict, we construct perturbations by partially replacing the genomics evidence while keeping the pathology evidence unchanged. For each patient i, we compare their pathology evidence with the genomics evidence of other patients from the same cancer cohort. The conflict score between the pathology evidence of patient i and the genomics evidence of patient j is defined as:

$$C_{i,j} = \frac{\mathrm{JS}\left(\mathbf{p}_{w,i}, \mathbf{p}_{o,j}\right)}{\log 2} \sqrt{(1 - V_{w,i})(1 - V_{o,j})} \tag{69}$$

where $\mathbf{p}_{w,i}$ and $\mathbf{p}_{o,j}$ denote the corresponding event-time distributions; $V_{w,i}$ and $V_{o,j}$ denote their evidential vacuities; $\mathrm{JS}(\cdot)$ denotes the Jensen–Shannon divergence measures the discrepancy between the two event-time distributions. For each patient i, the other patients are ranked according to $C_{i,j}$. The patients with conflict scores in the top 20% form the candidate set, from which one patient is randomly selected. Let $d(i)$ denote the index of the selected patient. The perturbed genomics evidence is then constructed by linear interpolation:

$$\widetilde{\mathbf{e}}^{(\lambda)}_{o,i} = (1 - \lambda)\mathbf{e}_{o,i} + \lambda \mathbf{e}_{o,d(i)}, \qquad \lambda \in \{0, 0.25, 0.5, 0.75, 1\} \tag{70}$$

When $\lambda = 0$, the original genomics evidence is retained. As λ increases, a larger proportion of the selected genomics evidence is introduced. When $\lambda = 1$, the original genomics evidence is completely replaced. The pathology evidence remains unchanged throughout the experiment. At each perturbation level, the random selection is repeated 200 times, and the resulting measurements are averaged. The remaining results are shown in Figure 9. As shown in the left panels, the mean modality conflict increases consistently with λ, confirming that the interpolation progressively enlarges the discrepancy between pathology and genomics evidence. The right panels show that the survival-curve drift increases with the perturbation strength under all three evidence combination strategies. At every nonzero perturbation level, the full evidential mechanism produces smaller deviations than direct evidence summation and the variant without vacuity weighting, with the differences becoming more evident at moderate and high perturbation levels.

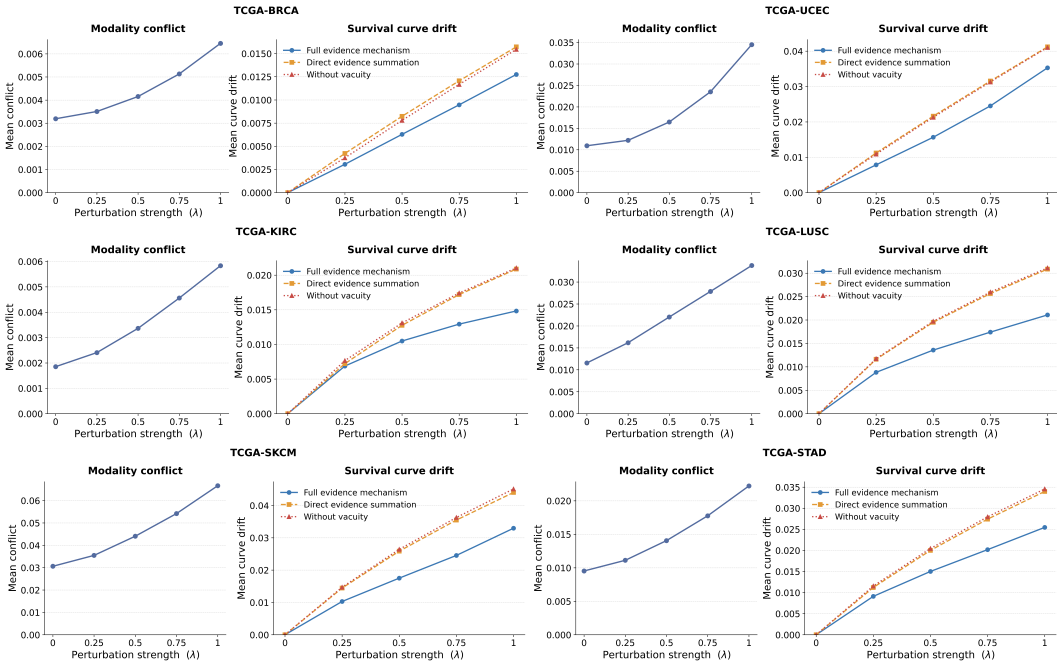

Figure 9: Conflict perturbation results on BRCA, UCEC, KIRC, LUSC, SKCM, STAD. The left panel shows the mean modality conflict at different perturbation strengths, while the right panel compares the mean survival-curve drift of the three settings.

L Robustness Analysis under Input Perturbations

We evaluate the robustness of DualRE-Surv by perturbing the pathology and genomics inputs during evaluation. Two types of perturbations are considered: Gaussian noise and dropout noise. For both types, the noise level is selected from $\lambda \in \{0, 0.1, 0.2, 0.3, 0.4\}$, where $\lambda = 0$ corresponds to the clean input. For Gaussian perturbations, noise of the corresponding magnitude is added to the pathology patch features or genomics pathway features. This setting evaluates model stability under continuous changes in the input representations. For dropout perturbations, a proportion of the pathology patches or genomics pathways is randomly removed according to λ, simulating partial information loss in the corresponding modality. For patient i, let $R_i^{(0)}$ denote the predicted risk obtained from the clean input and $R_i^{(\lambda)}$ the predicted risk at noise level λ. The absolute risk change is defined as:

$$\Delta R_i^{(\lambda)} = \left| R_i^{(\lambda)} - R_i^{(0)} \right| \tag{71}$$

The mean absolute risk change over all N patients is:

$$\overline{\Delta R}^{(\lambda)} = \frac{1}{N} \sum_{i=1}^{N} \Delta R_i^{(\lambda)} \tag{72}$$

A smaller $\overline{\Delta R}^{(\lambda)}$ indicates that the predicted risk changes less after the input is perturbed. To ensure comparability, the same definition of patient-level risk is used for DualRE-Surv and SlotSPE. The remaining results are shown in Figure 10. For both DualRE-Surv and SlotSPE, the mean absolute risk change generally increases with the noise level, indicating that stronger perturbations produce larger changes in the predicted risk. Across the eight datasets, DualRE-Surv produces smaller mean absolute risk changes than SlotSPE at the nonzero noise levels. These results show that DualRE-Surv exhibits stronger robustness.

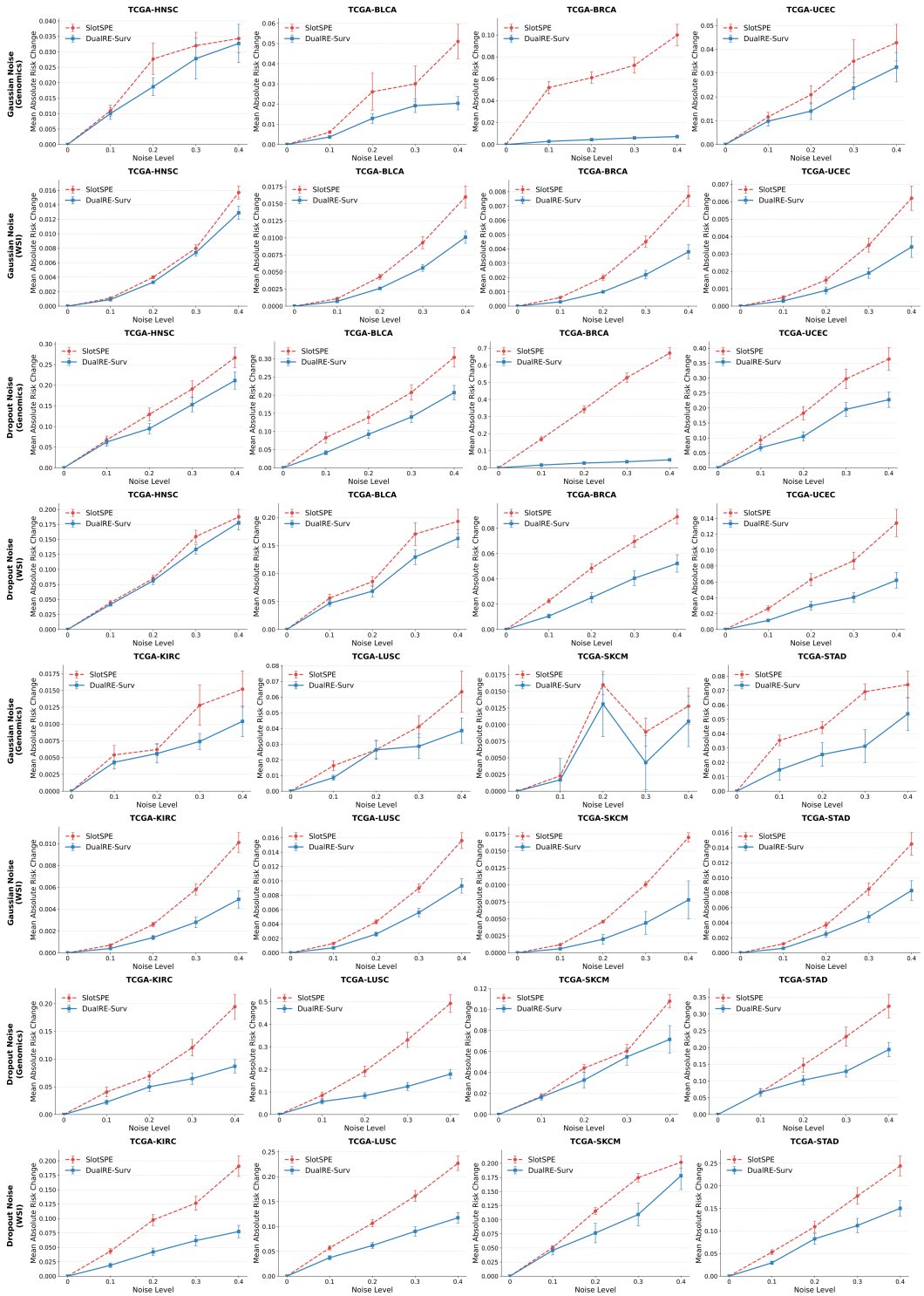

Figure 10: Robustness comparison between SlotSPE and DualRE-Surv under input perturbations on eight TCGA cohorts. From top to bottom, the rows show Gaussian noise on genomics and pathology features, followed by dropout noise on the same features.