Noise-Aware Importance–Uncertainty Disentangled Multimodal Learning for Robust Cancer Survival Prediction

Wenlong Ming[1], Wenbin Ye[2], Mingxin Liu[1], Depin Chen[1], Yiping Jiao[1], Jun Xu[1], and Xiangxue Wang[1⋆]

[1] Jiangsu Key Laboratory of Intelligent Medical Image Computing, School of Artificial Intelligence, Nanjing University of Information Science and Technology, 219 Ning Liu Road, Nanjing 210044, China
wming@nuist.edu.cn, mingxinliu@ieee.org, 202411620001@nuist.edu.cn, ping@nuist.edu.cn, jxu@nuist.edu.cn, xwang@nuist.edu.cn

[2] Jiangsu Key Laboratory of Intelligent Medical Image Computing, School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China
202412492765@nuist.edu.cn

Abstract. Accurate cancer prognosis prediction from multimodal data is critical for personalized treatment planning, yet remains challenging due to modality-specific technical noise and heterogeneous prognostic characteristics. In particular, whole slide images (WSIs) and mRNA expression profiles exhibit an overlooked asymmetry: WSIs provide stable but weak prognostic cues, whereas mRNA data offer strong prognostic relevance at the cost of high technical variability. This discrepancy between technical uncertainty and prognostic importance is not explicitly modeled by existing multimodal survival models, leading to fragile cross-modal fusion. We introduce the Noise-Aware Disentangled Multimodal Survival Network (NADMSurv), a framework that formalizes importance–uncertainty asymmetry as a guiding principle for multimodal representation learning. An importance–uncertainty dual-track module jointly performs feature pre-selection and patient-level noise quantification to encode heterogeneous prognostic patterns. A noise-aware disentanglement mechanism then separates multimodal representations into shared representations and modality-specific representations. To model high-order patient correlations, hypergraph neural networks are employed, and subjective logic is introduced to generate opinions with uncertainty quantification, ultimately achieving evidence-level fusion through cross-attention and consensus operators. Multi-cohort evaluation on TCGA datasets demonstrates consistent improvements in C-index over state-of-the-art methods. By analyzing modality-specific technical noise characteristics and disentangling the importance–uncertainty asymmetry, NADMSurv establishes a principled paradigm for robust multimodal survival prediction. Code and sufficient information are available at: https://github.com/imicjs/NADMSurv.

⋆ Corresponding author.

Keywords: Cancer prognostication · Technical Noise Quantification · Multimodal Fusion · Hypergraph neural network · Subjective Logic.

1 Introduction

Cancer poses a major global disease burden, with the International Agency for Research on Cancer (IARC) reporting nearly 20 million new cases in 2022 [1,2]. While traditional assessment based on tumor staging remains standard, it cannot fully reflect tumor biological heterogeneity [15,8], necessitating accurate prognostic models for effective clinical decision-making.

While transcriptome-based models identify molecular subtypes, their clinical use is limited by high costs [9,17]. WSIs provide rich morphological information yet exhibit limited prognostic power [6,12]. Consequently, multimodal fusion has gained attention, with methods such as HFBSurv [14] and MMsurv [26] employing advanced fusion strategies for survival prediction.

However, existing methods remain limited to direct feature concatenation, rarely disentangling shared and modality-specific representations [3,4]. More critically, they overlook the modality intrinsic importance–uncertainty asymmetry: WSIs exhibit high technical stability but limited prognostic power, whereas mRNA contains strong signals yet suffers from high technical noise [23,20].

To this end, we propose the Noise-Aware Disentangled Multimodal Survival Network (NADMSurv), explicitly modeling this asymmetry via an importance–uncertainty dual-track module for feature pre-selection and noise quantification, a noise-aware disentanglement mechanism separating shared and modality-specific representations, and hypergraph neural networks with subjective logic for evidence-level fusion. 5-fold cross-validation C-index results on five TCGA cohorts demonstrate state-of-the-art performance.

2 Methodology

2.1 Importance–uncertainty dual-track module

WSI Feature Pre-screening WSIs are Macenko-normalized, Otsu-segmented, and cropped into 256×256 patches ($N \leq 3,000$), encoded by ResNet50 into $\mathbf{F} \in \mathbb{R}^{N \times 2048}$. Patches are clustered into $M = 3$ pseudo-bags with MIL attention-based selection [26].

Uncertainty is quantified via entropy $u_i = -\sum_{j=1}^{M} p_{ij} \log p_{ij}$, where i and j index patches and pseudo-bags, respectively, and p_{ij} is the soft-assignment probability of patch i to bag j. Within each bag, an MIL attention scorer ranks patches by importance; the top-20% are retained as candidates and split by the median entropy into high- and low-uncertainty subsets, from which 4 patches each are selected, yielding 24 key patches per patient.

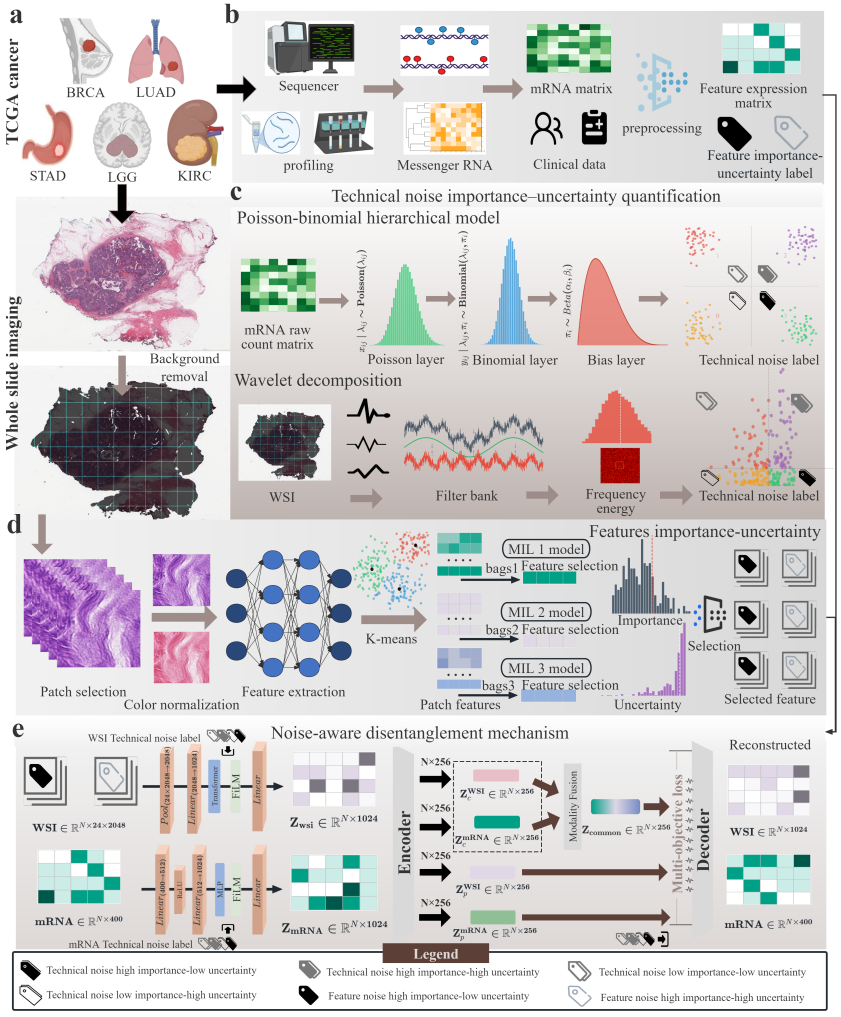

Fig. 1. Overview of the NADMSurv. (a) Collection of multi-modal data (WSI and mRNA) from five TCGA cancer cohorts; (b) mRNA feature pre-screening; (c) Patient-level technical noise importance–uncertainty quantification (Poisson-binomial hierarchical model for mRNA; wavelet decomposition for WSI); (d) WSI feature pre-screening; (e) Noise-aware disentanglement mechanism.

mRNA Feature Pre-screening Low-variance filtering with threshold 0.5 and univariate Cox regression at $P < 0.01$ select top 1,000 genes by absolute hazard ratio. Uncertainty is quantified by the standard error of the Cox coefficient, $SE_{\beta_i} = \sqrt{[I(\hat{\beta}_i)^{-1}]_{ii}}$, where i indexes the gene and regression coefficient β_i, and $[I(\hat{\beta}_i)^{-1}]_{ii}$ denotes the i-th diagonal element of the inverse Fisher information matrix, representing $\mathrm{Var}(\hat{\beta}_i)$.

Low-SE genes provide stable cross-patient consensus anchors, while high-SE genes capture genuine tumor subtype heterogeneity and individualized signals. The 200 lowest-SE and 200 highest-SE genes are retained, yielding 400 genes total.

WSI Patient-Level Technical Noise Quantification For each patient, we evaluate technical noise from $K \leq 3{,}000$ patches. Histological spatial continuity manifests as energy in the low-frequency approximation subband cA_2, whereas technical artifacts introduce abrupt transitions captured by high-frequency detail subbands cH_2, cV_2, and cD_2 under two-level wavelet decomposition [20].

Per-patch noise intensity is defined as $n_i = E_{\text{HF}}/E_{\text{tot}}$, where $E_{\text{HF}} = \|cH_2\|_F^2 + \|cV_2\|_F^2 + \|cD_2\|_F^2$ and $E_{\text{tot}} = E_{\text{HF}} + \|cA_2\|_F^2$. Here $i \in \{1, \ldots, K\}$ indexes the 256×256 patch; cA_2, cH_2, cV_2, cD_2 are the approximation, horizontal, vertical, and diagonal subband coefficients. The patient-level stability score is $s_{\text{WSI}}^{(j)} = 1 - \frac{1}{K}\sum_{i=1}^{K} n_i \in [0,1]$, where $j \in \{1, \ldots, N\}$ indexes the patient; higher values indicate lower overall noise and more reliable morphological features. Uncertainty is quantified by the coefficient of variation (CV) of per-patch noise levels. Median-based stratification of both importance and uncertainty yields four pseudo-labels for noise-aware supervision.

mRNA Patient-Level Technical Noise Quantification Technical noise in mRNA sequencing arises from capture efficiency variation and batch effects [21,22]. We employ a Poisson-binomial hierarchical model [24]: (i) Poisson layer $x_{ij} \mid \lambda_{ij} \sim \text{Poisson}(\lambda_{ij})$ for true transcriptional abundance; (ii) Binomial layer $y_{ij} \mid \lambda_{ij}, \pi_i \sim \text{Binomial}(\lambda_{ij}, \pi_i)$ with capture efficiency $\pi_i \in [0,1]$, where y_{ij} denotes the read count after sequencing capture dropout; (iii) Beta layer $\pi_i \sim \text{Beta}(\alpha_i, \beta_i)$.

Parameters are estimated via iterative EM. The E-step yields posterior mean $\hat{\pi}_i = (\alpha_i + \sum_{j=1}^{N} \lambda_{ij})/(\alpha_i + \beta_i + \sum_{j=1}^{N} \lambda_{ij})$ and variance $\text{Var}(\pi_i) = \alpha_i\beta_i/((\alpha_i + \beta_i)^2(\alpha_i+\beta_i+1))$. The M-step updates Beta priors via moment estimation as $\hat{\alpha}_i = \hat{\pi}_i(\hat{\pi}_i(1-\hat{\pi}_i)/\text{Var}(\pi_i)-1)$ and $\hat{\beta}_i = (1-\hat{\pi}_i)(\hat{\pi}_i(1-\hat{\pi}_i)/\text{Var}(\pi_i)-1)$, iterating until the change in π_i falls below 10^{-6}. Gene importance is $w_i = 1 - |\hat{\pi}_i - 0.5|$ (closer to 1 indicates weaker noise), with uncertainty measured by posterior variance. The patient-level noise score is $\text{noise}_{\text{mRNA}}^{(j)} = \frac{1}{G}\sum_{i=1}^{G} \sqrt{\frac{x_{ij}}{\max_k x_{kj}}}\,(1 - w_i)$, where x_{ij} is the raw count of gene i in patient j.

2.2 Noise-Aware Disentangled Mechanism

The WSI of size 24×2048 and mRNA of dimension 400 encoders output 1024-dim features. For each patient, four discrete noise pseudo-labels are generated by cohort-level median binarization of the WSI stability score $s_{\text{WSI}}^{(j)}$, the mRNA noise score $\text{noise}_{\text{mRNA}}^{(j)}$, and their corresponding uncertainty scores. These labels are mapped to continuous vectors $e^{(j)} \in \mathbb{R}^{32}$ via a learnable embedding matrix

$\mathbf{E} \in \mathbb{R}^{4\times 32}$. FiLM modulation [16] is inserted at an intermediate layer of each encoder, after the second Transformer block for WSI and after the second MLP block for mRNA. Concretely, applying MLP_γ and MLP_β to $e^{(j)}$ yields $\gamma^{(j)}$ and $\beta^{(j)}$, and the FiLM transform is defined as $\gamma^{(j)} \odot h + \beta^{(j)}$. Let $z_{\mathrm{enc}}^{(j)}$ denote the final noise-conditioned encoder output. This noise-conditioned scaling modulates feature responses according to the learned embedding, with sample-specific gains determined by the pseudo-label-driven FiLM parameters, thereby bridging noise quantification and representation disentanglement. Let $z_{\mathrm{enc}}^{(j)} = \mathrm{FiLM}(h^{(j)}, e^{(j)})$ denote the noise-conditioned encoder feature. Features are then projected into 256-dim common (z_c) and private (z_p) subspaces via $z_c = \mathrm{MLP}_{\mathrm{common}}(z_{\mathrm{enc}})$ and $z_p = \mathrm{MLP}_{\mathrm{private}}(z_{\mathrm{enc}})$. Let $\mathcal{B}$ denote a mini-batch of patients. Four weighted losses optimize disentanglement:

Alignment Loss enforces bimodal z_c convergence:

$$\mathcal{L}_{\mathrm{align}} = \frac{1}{|\mathcal{B}|} \sum_j w^{\mathrm{WSI},(j)} w^{\mathrm{mRNA},(j)} \left\| z_c^{\mathrm{WSI},(j)} - z_c^{\mathrm{mRNA},(j)} \right\|_2^2 \tag{1}$$

where $w^{\mathrm{WSI}} = [0.0, 1.0, 0.5, 0.0]$ and $w^{\mathrm{mRNA}} = [0.0, 0.0, 1.0, 1.0]$ for labels 0 (high-noise stable), 1 (high-noise unstable), 2 (low-noise stable), and 3 (low-noise unstable).

Adversarial Loss [7] ensures z_c is uninformative for noise prediction:

$$\mathcal{L}_{\mathrm{adv}} = \frac{1}{|\mathcal{B}|} \sum_j \frac{w^{\mathrm{WSI},(j)} + w^{\mathrm{mRNA},(j)}}{2} \ell_{\mathrm{CE}} \left(p\left(\hat{y}_j \mid \mathrm{GRL}(z_c^{(j)}) \right), y_j \right) \tag{2}$$

where j indexes the patient; y_j is the discrete noise pseudo-label of patient.

Separation Loss maximizes cosine distance between bimodal z_p:

$$\mathcal{L}_{\mathrm{sep}} = \frac{1}{|\mathcal{B}|} \sum_j w_{\mathrm{sep}}^{\mathrm{WSI},(j)} w_{\mathrm{sep}}^{\mathrm{mRNA},(j)} \cos\left(z_p^{\mathrm{WSI},(j)}, z_p^{\mathrm{mRNA},(j)} \right) \tag{3}$$

where $w_{\mathrm{sep}}^{\mathrm{WSI}} = [1.0, 0.0, 0.5, 1.0]$ and $w_{\mathrm{sep}}^{\mathrm{mRNA}} = [1.0, 1.0, 0.0, 0.0]$.

Reconstruction Loss preserves modality-specific information:

$$\mathcal{L}_{\mathrm{recon}} = \left\| \mathrm{Dec}_{\mathrm{WSI}}(z_p^{\mathrm{WSI}}) - h_{\mathrm{high}} \right\|_2^2 + \left\| \mathrm{Dec}_{\mathrm{mRNA}}(z_p^{\mathrm{mRNA}}) - x^{\mathrm{mRNA}} \right\|_2^2 \tag{4}$$

The total loss is: $\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{align}}\mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{adv}}\mathcal{L}_{\mathrm{adv}} + \lambda_{\mathrm{sep}}\mathcal{L}_{\mathrm{sep}} + \lambda_{\mathrm{recon}}\mathcal{L}_{\mathrm{recon}}$.

2.3 Evidential Hypergraph Fusion

The disentangled representations are processed by Hypergraph Neural Networks (HGNN) [5] with Subjective Logic fusion, combining cross-attention and evidential consensus.

Hypergraphs are constructed via K-nearest neighbors, with convolution performing node-hyperedge-node message passing [10,18]. Three parallel HGNNs

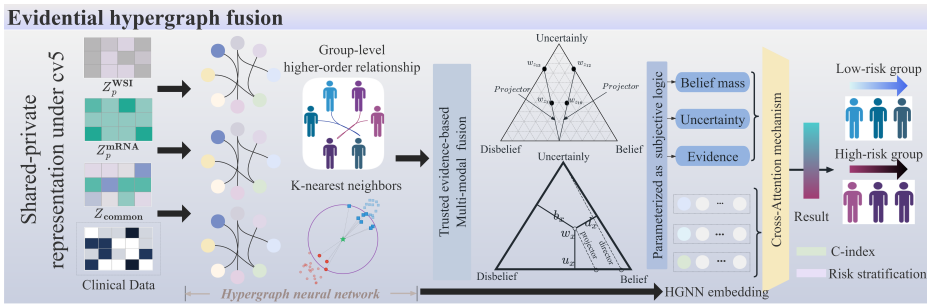

Fig. 2. Architecture of evidential hypergraph fusion: The disentangled representations (shared Z_{common}, WSI-private Z_p^{WSI}, and mRNA-private Z_p^{mRNA}) are processed by parallel HGNN via K-nearest neighbor construction. Subjective Logic converts HGNN outputs into evidential opinions.

process the common (z_c), WSI-private (z_p^{WSI}), and mRNA-private (z_p^{mRNA}) representations via K-NN hypergraphs.

Subjective Logic [11] maps risk scores to opinions $\omega = (b, u, a)$ comprising belief b, uncertainty u, and base rate a. Evidence is derived via sigmoid $e_{\text{risk}} = \sigma(r)$, $e_{\text{safe}} = 1 - \sigma(r)$ with total evidence $E = e_{\text{risk}} + e_{\text{safe}} + \epsilon$ ($\epsilon = 0.1$), yielding belief masses $b_{\text{risk}} = e_{\text{risk}}/E$ and $b_{\text{safe}} = e_{\text{safe}}/E$. The uncertainty u is estimated by each HGNN from the variance of its risk scores with base rate, and clamped to $[0.1, 0.5]$. The expected risk integrates both belief and uncertainty: $\mathbb{E}[r] = b_{\text{risk}} + u \cdot a$.

Private representations are fused via cross-attention, z_c as Query, private features as Key/Value, yielding $\mathbf{z}_{\text{fused}}$. Final risk prediction integrates cross-attention weights with evidential beliefs, where private contributions are modulated by attention α_m and certainty $(1 - u_m)$: $r_{\text{final}} = 0.5 \cdot \text{MLP}(\mathbf{z}_{\text{fused}}) + 0.3 \cdot b_c(1 - u_c) + \sum_m 0.1 \cdot b_m \alpha_m (1 - u_m)$. The model is trained via Cox partial log-likelihood.

3 Experiments

3.1 Dataset

Following, we conducted extensive experiments on five cancer types from The Cancer Genome Atlas (TCGA) including Breast Invasive Carcinoma (BRCA), Lung Adenocarcinoma (LUAD), Stomach Adenocarcinoma (STAD), Brain Lower Grade Glioma (LGG), and Kidney Renal Clear Cell Carcinoma (KIRC). We utilized hematoxylin and eosin (H&E) stained formalin-fixed paraffin-embedded (FFPE) whole slide images (WSIs) and mRNA expression profiles as the multi-modal inputs of our model.

Table 1. Performance comparison of different methods across five cancer types.

Methods	Modality P.	Modality G.	BRCA (n=925)	LUAD (n=425)	STAD (n=311)	LGG (n=430)	KIRC (n=493)
MLP		✓	0.6415±0.0669	0.6363±0.0878	0.6181±0.0689	0.8007±0.0683	0.6566±0.0631
SNN		✓	0.6768±0.0688	0.6196±0.0429	0.6262±0.0463	0.7869±0.0567	0.6662±0.0334
TransMIL	✓		0.6211±0.0346	0.5758±0.0383	0.5634±0.0571	0.6659±0.0679	0.6241±0.0355
AttentionMIL	✓		0.5626±0.0750	0.5644±0.0525	0.5383±0.0483	0.6434±0.0722	0.6012±0.0486
DeepSet	✓		0.5462±0.0663	0.5441±0.0457	0.5276±0.0406	0.6194±0.0786	0.6222±0.0373
CMTA	✓	✓	0.7268±0.0618	0.6436±0.0613	0.6178±0.0550	0.7486±0.0640	0.6156±0.0557
GPDBN	✓	✓	0.6783±0.0200	0.6110±0.0574	0.6367±0.0185	0.8024±0.0760	0.6707±0.0569
HFBSurv	✓	✓	0.7062±0.0327	0.6503±0.0636	0.6347±0.0293	0.6971±0.0241	0.5823±0.0305
MMSurv	✓	✓	0.7360±0.0739	0.6443±0.0442	0.6182±0.0677	0.8137±0.0380	0.6536±0.0395
Ours(NADMSurv)	✓	✓	**0.7470±0.0910**	**0.6788±0.0588**	**0.6420±0.0547**	**0.8322±0.0430**	**0.6959±0.0674**

Bold indicates the best, underlined indicates the second best. "P." indicates pathological images, "G." indicates genomic profiles.

3.2 Comparison with State-of-the-art Methods

We compared NADMSurv against nine methods (TABLE 1), including MLP and SNN [13] for mRNA-based prediction, TransMIL [19], AttentionMIL and DeepSet [27] for WSI-based prediction, and CMTA [28], GPDBN [25], HFBSurv [14] and MMSurv [26] for multimodal fusion. Evaluated via 5-fold cross-validation using C-index, NADMSurv achieves superior performance across all cohorts. Multimodal methods consistently outperform unimodal baselines, underscoring the value of integrating both modalities.

Table 2. Ablation study results of NADMSurv across five cancer types.

Method	BRCA (n=925)	LUAD (n=425)	STAD (n=311)	LGG (n=430)	KIRC (n=493)
Baseline(mRNA)	0.6040±0.0288	0.5524±0.0601	0.5709±0.0477	0.7884±0.0435	0.6172±0.0324
Baseline(WSI)	0.5788±0.0993	0.5335±0.0410	0.5673±0.0541	0.5658±0.0367	0.5337±0.0232
Baseline(mRNA+WSI)	0.6214±0.0442	0.5747±0.0446	0.5801±0.0268	0.6451±0.0624	0.6177±0.0490
With Module1	0.7006±0.0868	0.6196±0.0471	0.6203±0.0582	0.7973±0.0426	0.6282±0.0980
With Module2	0.6705±0.0637	0.6507±0.0569	0.5934±0.0794	0.8187±0.0455	0.6427±0.0496
Ours(NADMSurv)	**0.7470±0.0910**	**0.6788±0.0588**	**0.6420±0.0547**	**0.8322±0.0430**	**0.6959±0.0674**

Bold indicates the best, underlined indicates the second best.

3.3 Ablation Study

We established three baselines: unimodal mRNA/WSI and a simple concatenation fusion (TABLE 2). Results show that *Baseline(mRNA)* consistently outperforms *Baseline(WSI)*. Separately introducing the *Noise-aware Disentanglement module* (Module 1) and the *Evidential Hypergraph Fusion module* (Module 2) both achieved performance gains compared to the baseline models. The complete NADMSurv achieved optimal performance across all cohorts, verifying

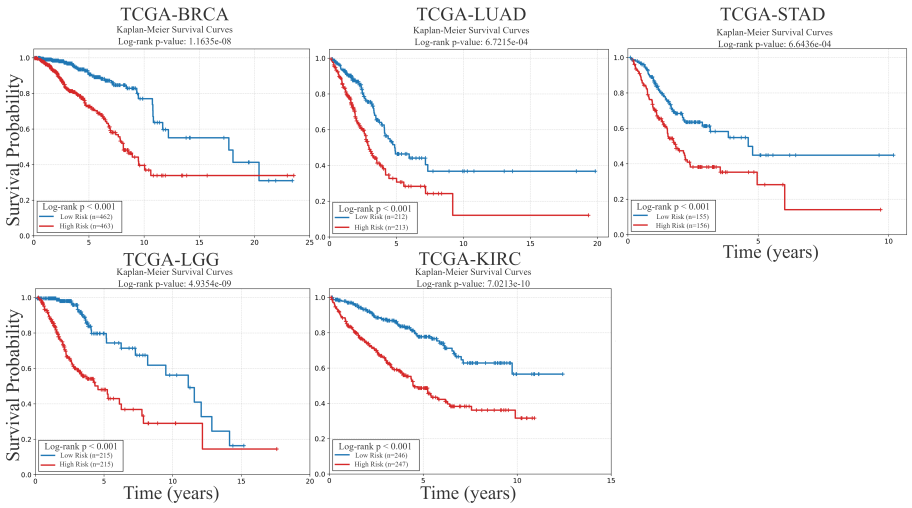

Fig. 3. Kaplan-Meier survival curves for risk stratification based on NADMSurv across five cancer types.

that noise-aware disentanglement combined with evidential fusion effectively excavates complementary prognostic information while resolving inter-modal conflicts.

3.4 Risk Stratification

To further validate the clinical efficacy of the model, we stratified patients into high-risk and low-risk groups based on the median of predicted risk scores and generated Kaplan-Meier survival curves to evaluate inter-group survival differences (Fig. 3). Across five cancer cohorts including BRCA ($p = 1.16 \times 10^{-8}$), LUAD ($p = 6.72 \times 10^{-4}$), STAD ($p = 6.64 \times 10^{-4}$), LGG ($p = 4.93 \times 10^{-9}$), and KIRC ($p = 7.02 \times 10^{-10}$), the high-risk group exhibited significantly lower survival probabilities compared to the low-risk group, with all Log-rank test p-values less than 0.001. These results demonstrate that NADMSurv enables clinically actionable risk stratification for personalized treatment.

4 Conclusion

This study proposes NADMSurv,which models the importance–uncertainty asymmetry between WSIs and mRNA. The importance–uncertainty dual-track module jointly performs feature pre-selection and patient-level noise quantification, while noise-aware disentanglement separates multimodal representations into shared and modality-specific components, and Subjective Logic enables evidential fusion. Experiments on five TCGA cohorts validate robust prognostic prediction.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 62401272 and 82441029), the Talent Start-up Funding Project of Nanjing University of Information Science and Technology (No. 2024r090), and the Program from Anhui Key Laboratory of Computational Medicine and Intelligent Health (No. AHCM2024Z001).

Disclosure of Interests

The authors have no competing interests to disclose.

References

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)
2. Bray, F., Laversanne, M., Weiderpass, E., Soerjomataram, I.: The ever-increasing importance of cancer as a leading cause of premature death worldwide. Cancer **127**(16), 3029–3030 (2021)
3. Cheerla, A., Gevaert, O.: Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics **35**(14), i446–i454 (2019)
4. 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)
5. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y.: Hypergraph neural networks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 3558–3565 (2019)
6. Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., Gerstung, M.: Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nature cancer **1**(8), 800–810 (2020)
7. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
8. Hausser, J., Alon, U.: Tumour heterogeneity and the evolutionary trade-offs of cancer. Nature Reviews Cancer **20**(4), 247–257 (2020)
9. He, X., Liu, X., Zuo, F., Shi, H., Jing, J.: Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. In: Seminars in Cancer Biology. vol. 88, pp. 187–200. Elsevier (2023)
10. Hu, Y., Li, X., Yi, Y., Huang, Y., Wang, G., Wang, D.: Deep learning-driven survival prediction in pan-cancer studies by integrating multimodal histology-genomic data. Briefings in Bioinformatics **26**(2), bbaf121 (2025)
11. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **9**(03), 279–311 (2001)

12. Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K.A., Bankhead, P., et al.: Pan-cancer image-based detection of clinically actionable genetic alterations. Nature cancer **1**(8), 789–799 (2020)
13. Li, J., Xu, Q., Wu, M., Huang, T., Wang, Y.: Pan-cancer classification based on self-normalizing neural networks and feature selection. Frontiers in Bioengineering and Biotechnology **8**, 766 (2020)
14. Li, R., Wu, X., Li, A., Wang, M.: Hfbsurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. Bioinformatics **38**(9), 2587–2594 (2022)
15. Marusyk, A., Almendro, V., Polyak, K.: Intra-tumour heterogeneity: a looking glass for cancer? Nature reviews cancer **12**(5), 323–334 (2012)
16. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
17. Pizurica, M., Zheng, Y., Carrillo-Perez, F., Noor, H., Yao, W., Wohlfart, C., Vladimirova, A., Marchal, K., Gevaert, O.: Digital profiling of gene expression from histology images with linearized attention. Nature Communications **15**(1), 9886 (2024)
18. Qu, M., Yang, G., Di, D., Su, T., Gao, Y., Song, Y., Fan, L.: Multimodal cancer survival analysis via hypergraph learning with cross-modality rebalance. arXiv preprint arXiv:2505.11997 (2025)
19. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
20. Singh, S., Kumar, R.: A blind noise estimation and removal in histopathological images (2022)
21. Tang, W., Jørgensen, A.C.S., Marguerat, S., Thomas, P., Shahrezaei, V.: Modelling capture efficiency of single-cell rna-sequencing data improves inference of transcriptome-wide burst kinetics. Bioinformatics **39**(7), btad395 (2023)
22. Vallejos, C.A., Marioni, J.C., Richardson, S.: Basics: Bayesian analysis of single-cell sequencing data. PLoS computational biology **11**(6), e1004333 (2015)
23. Varabyou, A., Salzberg, S.L., Pertea, M.: Effects of transcriptional noise on estimates of gene and transcript expression in rna sequencing experiments. Genome research **31**(2), 301–308 (2021)
24. Wang, Y., Shu, Z., Cao, Z., Grima, R.: From noise to models to numbers: Evaluating negative binomial models and parameter estimations in single-cell rna-seq. PLOS Computational Biology **22**(3), e1014014 (2026)
25. Wang, Z., Li, R., Wang, M., Li, A.: Gpdbn: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. Bioinformatics **37**(18), 2963–2970 (2021)
26. Yang, H., Wang, J., Wang, W., Shi, S., Liu, L., Yao, Y., Tian, G., Wang, P., Yang, J.: Mmsurv: a multimodal multi-instance multi-cancer survival prediction model integrating pathological images, clinical information, and sequencing data. Briefings in Bioinformatics **26**(3), bbaf209 (2025)
27. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. Advances in neural information processing systems **30** (2017)
28. Zhou, F., Chen, H.: Cross-modal translation and alignment for survival analysis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21485–21494 (2023)