UT-MIL: Uncertainty-Rectified and TME-Decoupling Dual-Stream Aggregation for Robust WSI Survival Prediction

Taiyuan Hu[1,2,†], Haijing Luan[1,2,†], Jifang Hu[1,2], Han Li[3], Yue Pei[1,2], Kaixing Yang[2], Ruilin Li[1], Beifang Niu[1], Xuebin Chi[1], Rui Yan[4,5,⋆], and Jinrong Jiang[1,2,⋆]

[1] Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Institute of Pathology, Technical University of Munich, Munich, Germany
[4] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei, China
[5] Suzhou Institute for Advanced Research, USTC; Jiangsu Provincial Key Laboratory of Multimodal Digital Twin Technology, Suzhou, China
yanrui@ustc.edu.cn, jjr@sccas.cn

Abstract. Survival analysis using Whole Slide Images (WSIs) is often challenged by histological artifacts and ambiguous tissue morphology within the tumor microenvironment (TME). While Multiple Instance Learning (MIL) has become a standard paradigm for this task, existing methods frequently overlook feature-level reliability, leading to overconfident yet erroneous prognostic stratifications. For advanced prognostic modeling, explicitly managing uncertainty is imperative to reduce out-of-distribution noise and refine ambiguous signals. To this end, we propose UT-MIL, an uncertainty-aware TME-decoupling framework that integrates clinical prognostic logic through two synergistic phases. First, we introduce the Uncertainty-Guided Feature Rectifier (UGFR) to estimate feature-level uncertainty using perturbation sensitivity and a Histomorphological Memory Bank. Unlike passive estimation, UGFR performs active feature intervention via a tri-branch routing mechanism: removing highly uncertain features, preserving reliable features, and refining ambiguous representations via prototype retrieval. Second, to aggregate the resulting purified manifold, we design the TME-Decoupling Contextual Aggregator (TDCA). TDCA employs a dual-stream architecture comprising a *Critical Stream* to highlight high-risk malignant foci and a *Context Stream* that utilizes learnable latent queries to decouple multifaceted TME patterns. Comprehensive evaluations across five TCGA benchmarks demonstrate that UT-MIL achieves superior prognostic performance over representative MIL baselines. Beyond empirical gains, our framework yields high-fidelity, noise-resilient spatial interpretability that integrates localized risk determinants with global histological contexts. Our code is available at https://github.com/hutaiyuan/UTMIL.

[†] Equal Contribution, [⋆] Corresponding Authors.

Keywords: Survival Analysis · Whole Slide Images · Multiple Instance Learning

1 Introduction

Computational pathology has emerged as a transformative tool for cancer prognosis, utilizing Whole Slide Images (WSIs) to guide clinical treatment [1,2,3,4]. Distinct from conventional categorical tasks like subtyping, WSI survival analysis is a higher-order prognostic challenge that necessitates modeling continuous-time risk trajectories under weak supervision [5,6,7]. It requires aggregating gigapixel-level visual information characterized by immense spatial heterogeneity and complex tissue morphology [8,9].

Multiple Instance Learning (MIL) has become a widely adopted paradigm for WSI survival analysis [10]. Mainstream approaches [9,11,12,13] typically rely on attention-based mechanisms to compute a weighted aggregation of patch features. However, oncology prognosis presents a highly heterogeneous learning objective. A patient's survival trajectory may be primarily driven by the most aggressive, localized high-risk malignant foci, corresponding to critical instances in MIL, yet simultaneously modulated by the broader tumor microenvironment (TME), such as immune cell infiltration and stromal response [14,15]. Conventional attention-based models may dilute these sharp, critical malignant signals by averaging them with numerous less informative patches. Conversely, max-pooling-centric methods can emphasize critical instances but struggle to explicitly capture the rich, multifaceted semantic patterns of the surrounding TME [16].

Furthermore, existing MIL frameworks operate under the implicit assumption that attention modules can autonomously filter out artifacts (e.g., marker ink, tissue folds) and handle morphologically ambiguous regions (e.g., poorly differentiated tissues) [17]. This absence of explicit reliability modeling compels models to assign overconfident risk scores to noisy or borderline inputs [18]. Because inherent biological variability and slide preparation artifacts can limit prognostic prediction [19], explicitly quantifying and managing uncertainty is essential to prevent misleading risk stratification in clinical workflows.

To bridge these gaps, we propose **UT-MIL**, an **U**ncertainty-Rectified **T**ME-Decoupling Contextual Aggregation framework designed for robust and interpretable survival prediction. Our method operates through two synergistic phases. First, in the Uncertainty Rectification Phase, we estimate feature-level uncertainty using perturbation sensitivity and a learnable Histomorphological Memory Bank. Guided by this quantified uncertainty, a dynamic tri-branch routing mechanism removes highly uncertain features, retains reliable semantic features, and actively rectifies ambiguous instances by interpolating them with canonical pathological prototypes. Second, in the Aggregation Phase, the purified feature manifold is processed by a dual-stream aggregation module. The *Critical Stream* selectively isolates the most prognostically salient instances representing peak malignancy. Concurrently, the *Context Stream* utilizes a set of learnable latent

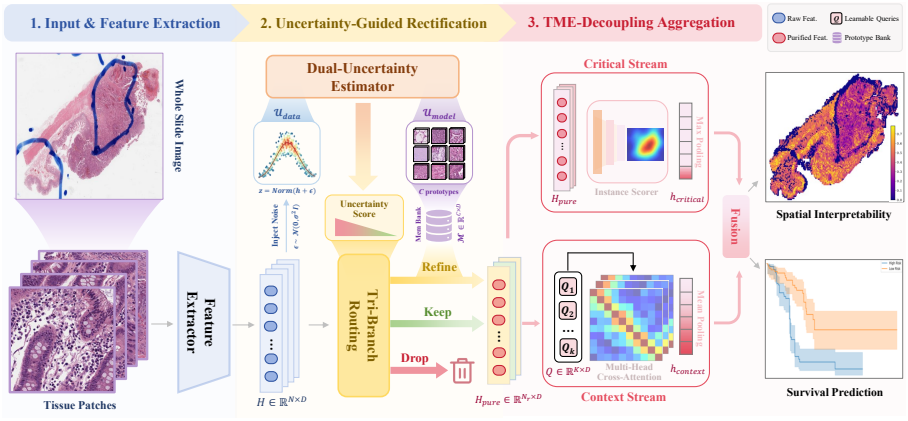

Fig. 1. Overview of the proposed UT-MIL framework.

queries, which function as diverse histological pattern extractors, to condense the encompassing tissue regions into a representation of the global TME.

Our main contributions are summarized as follows:

- To mitigate the impact of histological artifacts and morphological ambiguities in advanced prognostic modeling, we introduce a Dual-Uncertainty Estimator that disentangles aleatoric and epistemic uncertainties. Coupled with a Histomorphological Memory Bank, this module transforms uncertainty from a passive metric into an active feature-level rectifier through a dynamic tri-branch routing mechanism.
- To align MIL architectures with the multi-scale nature of clinical risk assessment, we propose the TME-Decoupling Contextual Aggregator. By synergizing a Critical Stream for localized malignant signals with a Context Stream for global TME patterns, UT-MIL decouples dual prognostic determinants to establish a robust representation that captures the complex landscape of tissue-based risk.
- Beyond predictive performance, UT-MIL demonstrates exceptional robustness against histological artifacts. It provides high-fidelity spatial risk mappings that offer intuitive, pathologist-verifiable evidence, improving the transparency of MIL-based WSI survival prediction.

2 Methods

Let a WSI be represented as a bag of instances $X = \{x_1, \ldots, x_N\}$, with corresponding extracted features $H = \{h_1, \ldots, h_N\} \in \mathbb{R}^{N \times D}$, where N denotes the number of patches and D is the feature dimension. Our objective is to accurately predict the continuous survival risk score for each patient. The proposed UT-MIL framework is built upon two sequential modules: the Uncertainty-Guided Feature Rectifier (UGFR) and the TME-Decoupling Contextual Aggregator (TDCA).

2.1 Phase I: Uncertainty-Guided Feature Rectifier (UGFR)

To guarantee that the downstream aggregator processes exclusively high-fidelity representations, UGFR dynamically evaluates and rectifies the reliability of each patch feature via a Dual-Uncertainty Estimator.

Dual-Uncertainty Estimation. We estimate feature-level uncertainty from two complementary perspectives.

Aleatoric Uncertainty ($\mathcal{U}_{data}$) models data-inherent degradation. We simulate feature perturbation by injecting scaled Gaussian noise $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma^2 I)$ to generate two augmented views: $z_1 = \text{Norm}(h + \epsilon_1)$ and $z_2 = \text{Norm}(h + \epsilon_2)$. The aleatoric uncertainty is defined as the cosine distance between these views:

$$\mathcal{U}_{data} = 1 - \frac{z_1 \cdot z_2}{\|z_1\|_2 \|z_2\|_2}.$$

Epistemic Uncertainty ($\mathcal{U}_{model}$) quantifies out-of-distribution morphologic patterns. We maintain a learnable Histomorphological Memory Bank $\mathcal{M} = \{p_1, \ldots, p_C\} \in \mathbb{R}^{C \times D}$, representing C canonical tissue prototypes initialized via K-Means on the training corpus [20]. Epistemic uncertainty is measured by the Euclidean distance to the nearest prototype:

$$\mathcal{U}_{model} = \min_{c \in \{1,\ldots,C\}} \|h - p_c\|_2.$$

The total uncertainty for a given patch h_i is computed as the normalized convex combination:

$$u_i = \lambda \mathcal{U}_{data}(h_i) + (1 - \lambda)\mathcal{U}_{model}(h_i),$$

where $\lambda \in [0, 1]$ is a balancing hyperparameter.

Tri-Branch Dynamic Routing. Rather than relying on conventional soft-attention masking, which often struggles to entirely suppress severe artifacts, we introduce a hard-routing mechanism guided by batch-wise quantiles (q_{low}, q_{high}) of the total uncertainty distribution. The routing function $\mathcal{R}(h_i, u_i)$ is mathematically formulated as:

$$\mathcal{R}(h_i, u_i) = \begin{cases} h_i, & \text{if } u_i < q_{low} \text{ (Keep)} \\ (1 - \alpha)h_i + \alpha p_{c^*}, & \text{if } q_{low} \leq u_i < q_{high} \text{ (Refine)} \\ \emptyset, & \text{if } u_i \geq q_{high} \text{ (Drop)} \end{cases}$$

where $p_{c^*} \in \mathcal{M}$ denotes the closest prototype to h_i, and α acts as an interpolation weight. Specifically, the *Keep* branch strictly preserves highly reliable patches. The *Refine* branch actively pulls ambiguous boundary features toward distinct semantic clusters. The *Drop* branch identifies irreparable noise and physically eradicates them from the sequence. This triage process yields a purified, length-reduced feature set $H_{pure} \in \mathbb{R}^{N_r \times D}$, where $N_r \leq N$.

2.2 Phase II: TME-Decoupling Contextual Aggregator (TDCA)

To mathematically formalize the clinical diagnostic workflow, TDCA processes H_{pure} through two parallel streams, capturing localized high-risk malignant signals and broader TME context.

Critical Instance Stream. Cancer prognosis is disproportionately driven by the most aggressive tumor sub-clone[14,15]. To model this, an Instance Scorer projects H_{pure} into instance-level risk scores $S \in \mathbb{R}^{N_r}$ via a learnable mapping. We actively isolate the most critical patch feature via an arg max operation:

$$h_{critical} = H_{pure}[\arg\max(S)].$$

Latent Query Context Stream. Because the TME encompasses diverse prognostic indicators beyond the tumor core, we define a set of K learnable latent queries $Q \in \mathbb{R}^{K \times D}$. Utilizing a Multi-Head Cross-Attention (MHA) operation, these queries dynamically scan the purified slide:

$$Z_{context} = \mathrm{MHA}(Q, H_{pure}, H_{pure}) = \mathrm{Softmax}\left(\frac{QH_{pure}^{\top}}{\sqrt{D}}\right) H_{pure}.$$

During optimization, each query autonomously specializes in capturing a specific histological pattern (e.g., stromal desmoplasia, lymphocytic infiltration). These K distinct contextual responses are subsequently aggregated via mean pooling to construct a holistic global representation:

$$h_{context} = \frac{1}{K}\sum_{k=1}^{K} Z_{context}^{(k)}.$$

Prognostic Fusion and Optimization. The final WSI-level representation fuses the critical instance representation with the global contextual representation via concatenation:

$$h_{fusion} = [h_{critical} \parallel h_{context}].$$

A prediction head maps h_{fusion} to a continuous prognostic risk score. The entire network is trained end-to-end optimizing the Cox Partial Likelihood loss.

3 Experiments

3.1 Datasets and Experimental Setup

We evaluate our method on five public TCGA benchmarks[21,22]: TCGA-BRCA ($n = 1127$), TCGA-LUSC ($n = 505$), TCGA-BLCA ($n = 450$), TCGA-STAD ($n = 349$), and TCGA-CRC ($n = 448$), following a rigorous 5-fold cross-validation protocol. WSIs are segmented into non-overlapping 256×256 patches at $20\times$ magnification. We utilize a pretrained UNI2-h[23] model to extract $D = 1536$ dimensional features, which are subsequently projected to 512 dimensions. The

Table 1. Performance comparison (C-Index) across five TCGA benchmarks. Results are reported as mean ± std over 5-fold cross-validation. The best results are highlighted in **bold** and the second-best are underlined.

Method	BRCA	LUSC	BLCA	STAD	CRC	Mean
ABMIL [9]	$0.610_{\pm 0.035}$	$0.674_{\pm 0.033}$	$0.736_{\pm 0.018}$	$0.574_{\pm 0.046}$	$0.631_{\pm 0.055}$	0.645
CLAM [25]	$0.637_{\pm 0.075}$	$0.689_{\pm 0.041}$	$0.746_{\pm 0.041}$	$0.558_{\pm 0.068}$	$0.632_{\pm 0.064}$	0.652
DSMIL [16]	$0.627_{\pm 0.053}$	$0.683_{\pm 0.031}$	$\underline{0.753_{\pm 0.035}}$	$\underline{0.599_{\pm 0.041}}$	$0.642_{\pm 0.051}$	$\underline{0.661}$
TransMIL [11]	$0.524_{\pm 0.036}$	$0.571_{\pm 0.039}$	$0.641_{\pm 0.043}$	$0.541_{\pm 0.028}$	$0.625_{\pm 0.024}$	0.580
DTFD-MIL [26]	$0.653_{\pm 0.058}$	$\underline{0.690_{\pm 0.046}}$	$0.733_{\pm 0.040}$	$0.556_{\pm 0.045}$	$0.634_{\pm 0.072}$	0.653
HIPT [27]	$\underline{0.657_{\pm 0.018}}$	$0.675_{\pm 0.049}$	$0.747_{\pm 0.044}$	$0.582_{\pm 0.038}$	$0.624_{\pm 0.060}$	0.657
MambaMIL [28]	$0.647_{\pm 0.061}$	$0.665_{\pm 0.013}$	$0.740_{\pm 0.027}$	$0.597_{\pm 0.037}$	$\underline{0.656_{\pm 0.032}}$	$\underline{0.661}$
OTSurv [12]	$0.606_{\pm 0.079}$	$0.661_{\pm 0.027}$	$0.655_{\pm 0.058}$	$0.593_{\pm 0.060}$	$0.644_{\pm 0.055}$	0.632
Panther [13]	$0.654_{\pm 0.032}$	$0.687_{\pm 0.028}$	$0.712_{\pm 0.031}$	$0.592_{\pm 0.023}$	$0.641_{\pm 0.050}$	0.657
UT-MIL	$\mathbf{0.712}_{\pm \mathbf{0.063}}$	$\mathbf{0.693}_{\pm \mathbf{0.048}}$	$\mathbf{0.783}_{\pm \mathbf{0.019}}$	$\mathbf{0.605}_{\pm \mathbf{0.036}}$	$\mathbf{0.661}_{\pm \mathbf{0.068}}$	**0.691**

Table 2. Ablation study of UT-MIL. We evaluate the incremental contribution of the TME-Decoupling Contextual Aggregator (TDCA), Aleatoric-based Hard Drop (HD), and Epistemic-based Prototype Refinement (PR).

TDCA	HD	PR	BRCA	LUSC	BLCA	STAD	CRC	Mean
–	–	–	$0.610_{\pm 0.035}$	$0.674_{\pm 0.033}$	$0.736_{\pm 0.018}$	$0.574_{\pm 0.046}$	$0.631_{\pm 0.055}$	0.645
✓	–	–	$0.680_{\pm 0.061}$	$0.675_{\pm 0.034}$	$0.739_{\pm 0.011}$	$0.571_{\pm 0.023}$	$0.637_{\pm 0.090}$	0.660
✓	✓	–	$0.706_{\pm 0.021}$	$0.663_{\pm 0.031}$	$0.768_{\pm 0.011}$	$0.601_{\pm 0.044}$	$0.651_{\pm 0.080}$	0.678
✓	✓	✓	$\mathbf{0.712}_{\pm \mathbf{0.063}}$	$\mathbf{0.693}_{\pm \mathbf{0.048}}$	$\mathbf{0.783}_{\pm \mathbf{0.019}}$	$\mathbf{0.605}_{\pm \mathbf{0.036}}$	$\mathbf{0.661}_{\pm \mathbf{0.068}}$	**0.691**

Histomorphological Memory Bank stores $C = 16$ prototypes. The uncertainty routing thresholds are set to $q_{low} = 0.3$ and $q_{high} = 0.8$, with a noise scale of $\sigma = 0.1$. The TDCA employs $K = 8$ latent queries. The network is trained for 20 epochs with a batch size of 32 using the Adam optimizer[24] (learning rate 2×10^{-4}) on a single NVIDIA A800 GPU. The Concordance Index (C-Index) serves as the primary evaluation metric.

3.2 Results and Discussion

Comparison with State-of-the-Art. Utilizing the same frozen feature representations, Table 1 demonstrates that UT-MIL consistently establishes a superior prognostic baseline across all five TCGA benchmarks. Compared with the strongest competing baselines, UT-MIL improves the mean C-Index from 0.661 to 0.691, corresponding to a 4.5% relative gain, with a fold-level paired bootstrap 95% CI of $\Delta C = [0.0029, 0.0582]$. Our framework outperforms standard attention-pooling methods, including ABMIL [9] and CLAM [25], while surpassing advanced architectures such as DSMIL [16], TransMIL [11], DTFD-MIL

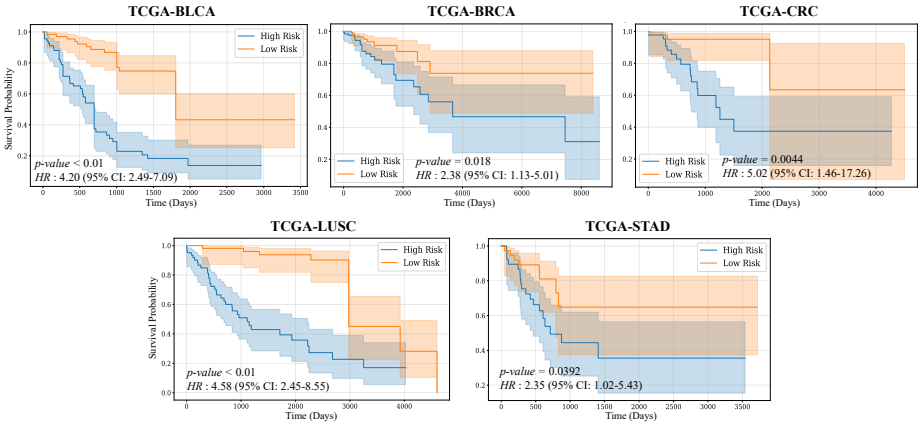

Fig. 2. Kaplan-Meier survival curves across diverse TCGA cohorts.

[26], and the multi-scale HIPT [27]. Notably, UT-MIL also outperforms recent competitive models like MambaMIL [28], OTSurv [12], and Panther [13]. The efficacy of our approach is further corroborated by the Kaplan-Meier survival curves in Fig. 2, which demonstrate statistically significant risk stratification across diverse patient cohorts [29].

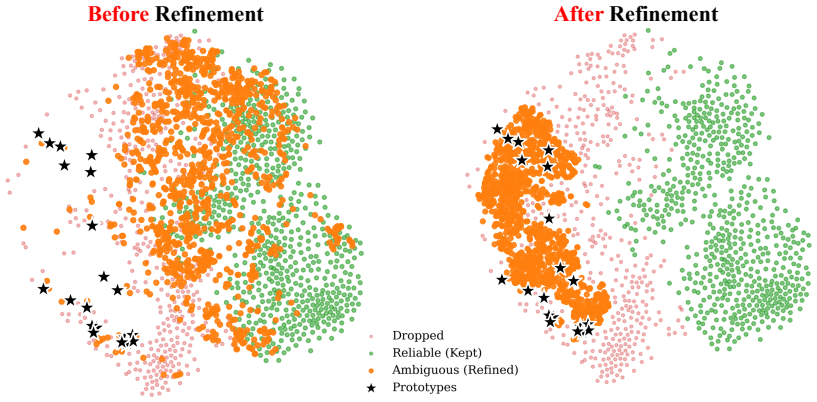

Fig. 3. t-SNE visualization before and after Prototype Refinement. Ambiguous instances (orange) move toward established pathological prototypes (black stars).

Ablation Study. Table 2 highlights the hierarchical performance gains of the proposed modules. Introducing TDCA improves the mean C-Index over conventional pooling, suggesting the benefit of jointly modeling critical instances and

global TME patterns. Adding Aleatoric-guided Hard Drop further improves performance by reducing the influence of highly uncertain features, such as potential histological artifacts. The full model achieves the best overall performance after incorporating Prototype Refinement, indicating that refining ambiguous representations further benefits WSI survival prediction.

The efficacy of Prototype Refinement is visually illustrated in Fig. 3. While ambiguous features are initially scattered and entangled with noise, the refinement process drives these boundary patches toward semantic convergence with established histological prototypes. This effectively purifies the feature manifold and clarifies the representation space.

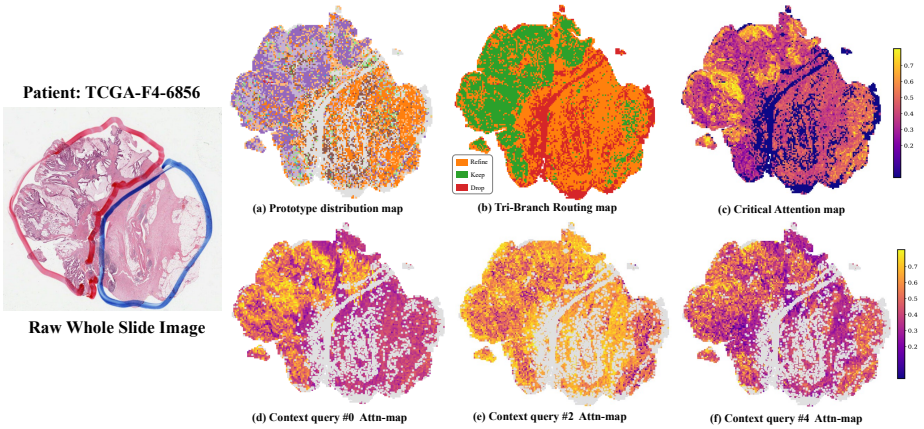

Fig. 4. Clinical interpretability of UT-MIL. The *Critical Stream* localizes the most malignant tumor cores, while the *Context Stream* queries autonomously decouple the microenvironment into distinct, semantically meaningful spatial masks.

Clinical Interpretability. As illustrated in Fig. 4, UT-MIL shows robustness to common histological artifacts, such as the marker pen annotations in patient TCGA-F4-6856. The UGFR module assigns high uncertainty to the red and blue ink traces and routes them to the Drop branch (Fig. 4b), reducing their influence on downstream aggregation.

Based on this reliability-aware representation, the TDCA module provides interpretable spatial evidence. The *Critical Stream* highlights the most aggressive tumor regions (Fig. 4c), while the *Context Stream* leverages latent queries to autonomously decouple the TME into semantically distinct masks (Fig. 4d-f). These queries specialize in diverse patterns—from broad architectures to punctate immune infiltrates—providing artifact-free prognostic evidence.

4 Conclusion

This paper presents UT-MIL, an uncertainty-rectified and TME-decoupling framework for WSI survival analysis. Given the imperative of managing uncertainty for advanced prognostic modeling, UGFR is introduced to filter out-of-distribution noise and rectify ambiguous signals, thereby ensuring the robust aggregation of reliable prognostic features. TDCA further processes the purified manifold by synergistically fusing localized critical malignant signals with query-extracted global TME contexts. Extensive evaluations on TCGA benchmarks demonstrate that UT-MIL achieves state-of-the-art prognostic performance among representative MIL baselines and provides high-fidelity, noise-resilient spatial risk mappings, improving the interpretability of MIL-based WSI survival prediction.

Acknowledgments This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2024ZD0533500), National Natural Science Foundation of China (No. 62402473), and Basic Research Program of Jiangsu (No.BK20255001).

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

References

1. Song, A.H., Jaume, G., Williamson, D.F., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F.: Artificial intelligence for digital and computational pathology. Nature Reviews Bioengineering **1**(12), 930–949 (2023)
2. Hu, T., Luan, H., Yan, R., Hu, J., Yang, K., Han, X., Liu, W., He, J., Duan, X., Li, R., et al.: Transformer-based multi-scale fusion for robust predicting microsatellite instability from pathological images. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2046–2053. IEEE (2024)
3. Luan, H., Hu, T., Hu, J., Liu, W., Yang, K., Pei, Y., Li, R., He, J., Gao, Y., Sun, D., et al.: Breast cancer homologous recombination deficiency prediction from pathological images with a sufficient and representative transformer. NPJ Precision Oncology **9**(1), 160 (2025)
4. Wang, B., Zhang, K., Wang, Y., Gu, Y., Luan, H., Zhou, Y., Hu, T., Wang, R., Yang, Z., Jiang, Z., et al.: Wsisum: Wsi summarization via dual-level semantic reconstruction. Medical Image Analysis p. 103970 (2026)
5. El Nahhas, O.S., van Treeck, M., Wölflein, G., Unger, M., Ligero, M., Lenz, T., Wagner, S.J., Hewitt, K.J., Khader, F., Foersch, S., et al.: From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology. Nature protocols **20**(1), 293–316 (2025)
6. Yu, J., Chen, H., Hu, L., Wu, B., Zhou, S., Zhu, J., Jiang, Y., Han, S., Wang, Z.: Exploring multi-instance learning in whole slide imaging: Current and future perspectives. Pathology-Research and Practice **271**, 156006 (2025)
7. Yang, Z., Shi, X., Ba, W., Song, Z., Luan, H., Hu, T., Lin, S., Wang, J., Zhou, S.K., Yan, R.: Fusion of multi-scale heterogeneous pathology foundation models for whole slide image analysis. arXiv preprint arXiv:2510.27237 (2025)

8. Luan, H., Hu, T., Hu, J., Li, R., Ji, D., He, J., Duan, X., Yang, C., Gao, Y., Chen, F., et al.: Multi-class cancer classification of whole slide images through transformer and multiple instance learning. In: International symposium on bioinformatics research and applications. pp. 150–164. Springer (2023)
9. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
10. Luan, H., Yang, K., Hu, T., Hu, J., Liu, S., Li, R., He, J., Yan, R., Guo, X., Qian, N., et al.: Review of deep learning-based pathological image classification: From task-specific models to foundation models. Future Generation Computer Systems **164**, 107578 (2025)
11. 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)
12. Ren, Q., Wang, Y., Fang, R., Ling, H., You, C.: Otsurv: A novel multiple instance learning framework for survival prediction with heterogeneity-aware optimal transport. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 439–449. Springer (2025)
13. Song, A.H., Chen, R.J., Ding, T., Williamson, D.F., Jaume, G., Mahmood, F.: Morphological prototyping for unsupervised slide representation learning in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11566–11578 (2024)
14. Ge, R., Wang, Z., Cheng, L.: Tumor microenvironment heterogeneity an important mediator of prostate cancer progression and therapeutic resistance. NPJ precision oncology **6**(1), 31 (2022)
15. Chen, H., Deng, C., Gao, J., Wang, J., Fu, F., Wang, Y., Wang, Q., Zhang, M., Zhang, S., Fan, F., et al.: Integrative spatial analysis reveals tumor heterogeneity and immune colony niche related to clinical outcomes in small cell lung cancer. Cancer Cell **43**(3), 519–536 (2025)
16. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
17. Xiong, C., Chen, H., Sung, J.J., King, I.: Diagnose like a pathologist: Transformer-enabled hierarchical attention-guided multiple instance learning for whole slide image classification. arXiv preprint arXiv:2301.08125 (2023)
18. Zhang, H., Burrows, L., Meng, Y., Sculthorpe, D., Mukherjee, A., Coupland, S.E., Chen, K., Zheng, Y.: Weakly supervised segmentation with point annotations for histopathology images via contrast-based variational model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15630–15640 (2023)
19. Luo, X., Xiang, J., Ji, Y., Li, R.: nnmil: A generalizable multiple instance learning framework for computational pathology. arXiv preprint arXiv:2511.14907 (2025)
20. Likas, A., Vlassis, N., Verbeek, J.J.: The global k-means clustering algorithm. Pattern recognition **36**(2), 451–461 (2003)
21. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia **2015**(1), 68–77 (2015)
22. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al.: An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell **173**(2), 400–416 (2018)

23. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A.H., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine (2024)
24. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
26. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18802–18812 (2022)
27. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16144–16155 (2022)
28. Yang, S., Wang, Y., Chen, H.: Mambamil: Enhancing long sequence modeling with sequence reordering in computational pathology. In: International conference on medical image computing and computer-assisted intervention. pp. 296–306. Springer (2024)
29. Efron, B.: Logistic regression, survival analysis, and the kaplan-meier curve. Journal of the American statistical Association **83**(402), 414–425 (1988)