TRIAGE-MIL: Multi-Axis Instance Selection and Semantic Hypergraph Modeling for Survival Prediction from Whole-Slide Images

Barathi Subramanian[1], Rathinaraja Jeyaraj[1], Songmi Noh[2], George A. Fisher[1], and Jeanne Shen[1,*]

[1] Stanford University, Palo Alto, CA 94305, USA
[2] CHA University College of Medicine, Seoul 13488, Republic of Korea

Abstract. Cancer survival prognostication using whole-slide images (WSI) remains challenging, as WSI contain rich yet complex prognostic signals embedded within a broader tissue architectural and tumor microenvironmental context. The gigapixel resolution of WSI also imposes substantial computational constraints on deep learning models. Multiple instance learning (MIL) mitigates this by aggregating tile-level features, but often treats tiles as an unordered set, discarding architectural context critical for modeling tumor biology and disease progression. Graph-based MIL restores spatial structure but is typically limited to pairwise edges and fails to capture higher-order tissue interactions. To address these limitations, we propose TRIAGE-MIL, a two-stage framework that couples unsupervised instance selection with semantic hypergraph modeling. First, we introduce multi-axis stratified sampling (MASS), which selects a fixed-size subset of tiles using statistical proxies along four phenotype-inspired axes: morphologic heterogeneity, tissue regularity, microenvironmental interface, and tissue diversity. This yields a proxy-enriched slide representation without requiring region-level ground-truth labels. Second, we introduce a semantic hierarchical hypergraph network that models intra- and inter-tissue relationships via constrained hyperedges induced by MASS groupings, enabling higher-order relational reasoning beyond pairwise neighborhood. Across six different cancer cohorts, TRIAGE-MIL achieved the highest mean C-index among 13 state-of-the-art baselines and showed significant Kaplan−Meier risk-group separation in all cohorts (log-rank $p < 0.05$), indicating consistent patient risk stratification. `https://github.com/barathi-1993/TRIAGE-MIL`

Keywords: Hypergraph Modeling · MIL · Survival Prediction.

1 Introduction

Survival prognostication is a central task in clinical oncology because it guides patient risk stratification and treatment planning [1,2]. In routine clinical practice, prognostic assessment relies on pathologists' microscopic review of tissue

*jeannes@stanford.edu

slides, but this process is constrained by time, interobserver variability, and the difficulty of consistently extracting subtle signals from heterogeneous specimens. Computational pathology provides a complementary approach by quantifying patterns linked to tissue composition, architecture, and the tumor microenvironment (TME) [1,3,4,5]. These cues are often sparse, spatially localized, and morphologically heterogeneous [6,7], making consistent identification difficult when relevant regions are small or distributed across tissue compartments.

Recent advances in digital pathology and artificial intelligence (AI) have enabled prognostication from gigapixel whole-slide images (WSI), in which deep learning models estimate time-to-event outcomes such as recurrence [8] and overall survival [9,10]. However, the ultra-high resolution of WSI creates a computational bottleneck. At typical magnifications (e.g., 20×), a single WSI yields tens of thousands of tiles, making dense supervision expensive and region-level survival annotation impractical [11,12]. Multiple instance learning (MIL) has therefore become dominant, representing each WSI as a bag of tile instances paired with a patient-level outcome. While effective, standard MIL pipelines face persistent challenges in survival modeling [4,13,14,15]. First, prognostic heterogeneity is long-tailed. Only a small subset of tiles may drive risk, so random sampling can miss critical cues. Second, many MIL aggregators, including max-pooling and attention pooling, treat tiles as an unordered set of independent instances [14,15]. This can discard the spatial and structural context represented in invasive fronts, glandular organization, stromal boundaries, and immune infiltrates, which are often prognostically relevant [16,17,18].

To incorporate spatial context, recent works have extended MIL with graph-based formulations, where tiles serve as nodes and edges connect nearby tiles based on spatial proximity [17,19]. While graph-based MIL captures local topology better than unordered pooling, it is typically restricted to pairwise edges [20]. Consequently, higher-level interactions between multiple regions are only captured implicitly through multi-hop message passing, rather than being modeled explicitly as collective tissue relationships [21,22]. In WSI, prognostic evidence can depend on coordinated patterns across multiple compartments, which are not naturally represented by pairwise edges. This limitation is often compounded by low-quality and site-specific signatures that contaminate tile embeddings and bias downstream risk estimation [23,24,25,26]. In practice, morphology-based prognostication often depends on higher-order architectural context (e.g., co-occurrence of tumor nests and immune-infiltrated stroma), where the prognostic signal arises from sets of regions acting together, not isolated tiles or simple pairwise neighborhoods. Additionally, many graph-based pipelines still rely on random or uniform tile sampling, which can discard sparse but prognostically important regions (e.g., invasive fronts) before modeling is performed [3,15].

Motivated by these research gaps, we propose TRIAGE-MIL, a two-stage framework that couples phenotype-motivated tile selection with semantic hypergraph modeling. Our key contributions are: (i) multi-axis stratified sampling (MASS), an unsupervised tile selector that stratifies tiles using embedding-derived proxies along four phenotype-inspired axes: morphologic heterogeneity,

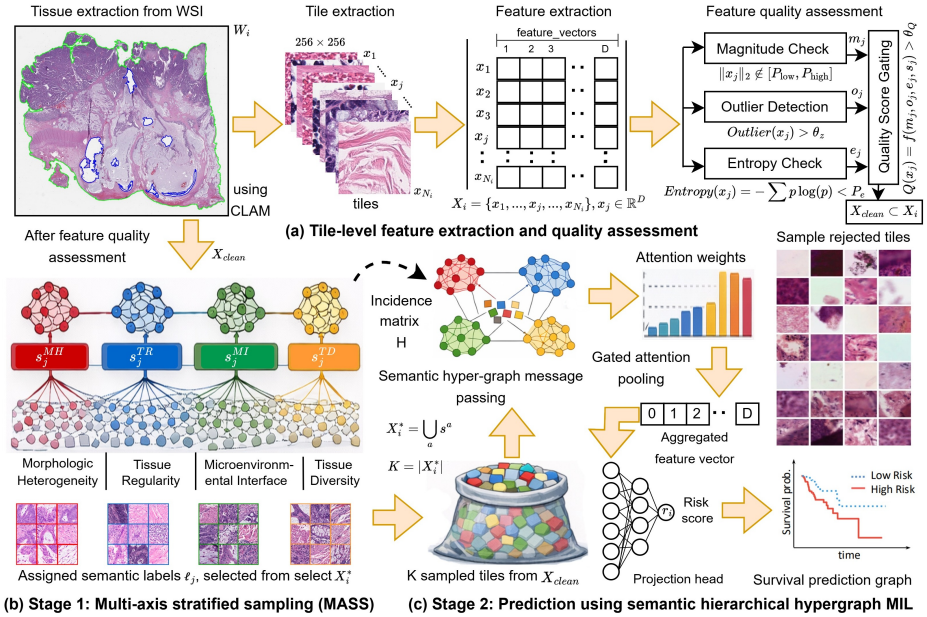

Fig. 1. Overview of TRIAGE-MIL framework. (a) CLAM masking [3], tile and feature extraction, and feature quality filtering (rejects low-quality tiles) to obtain X_{clean}. (b) MASS selects a fixed-budget (K) subset X_i^* using four phenotype-inspired axes: morphologic heterogeneity (MH), tissue regularity (TR), microenvironmental interface (MI), and tissue diversity (TD). (c) A semantic hierarchical hypergraph MIL model with gated attention pooling predicts patient risk and survival.

tissue regularity, microenvironmental interface, and tissue diversity; (ii) a semantic hierarchical hypergraph with constrained hyperedges and a "super-node" structure that links tiles by spatial proximity and semantic similarity, enabling message passing over higher-order groups beyond pairwise neighborhoods [21,22]; and (iii) an evaluation against 13 baselines across six cohorts [17,27,28] using standard survival ranking and risk stratification analyses [4,5], supporting WSI-based computational risk stratification in clinical workflows [29].

2 Method

As illustrated in Fig. 1, TRIAGE-MIL is a two-stage framework for WSI-based survival prognostication that combines unsupervised instance selection with higher-order relational modeling. Given a cohort of N patients with WSI W_i and outcome (t_i,δ_i), where t_i denotes the observed time-to-event and $\delta_i \in \{0,1\}$ indicates censoring status, the objective is to learn a mapping, f, that predicts a patient risk score, r_i. Each W_i is divided into N_i non-overlapping tiles with feature set $X_i = \{x_j\}_{j=1}^{N_i}$, where $x_j \in \mathbb{R}^D$, and coordinates $\mathcal{C}_i = \left\{c_j \in \mathbb{R}^2\right\}_{j=1}^{N_i}$.

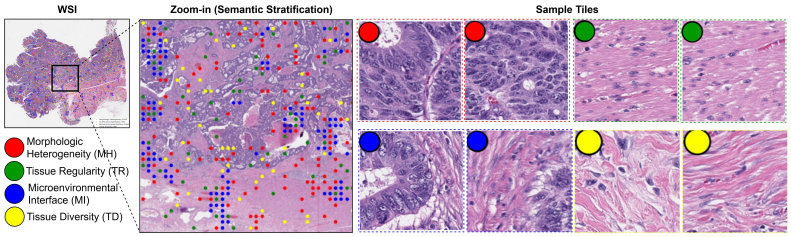

Fig. 2. MASS stratification overview. Left: WSI thumbnail with the selected ROI. Center: zoomed ROI with MASS-selected tiles colored by axis label. Right: representative tile examples for each axis: (MH, red), (TR, green), (MI, blue), and (TD, yellow).

As N_i is typically large, MASS selects a fixed-size (K) subset, $X_i^* \subset X_i$, with $|X_i^*| = K \ll N_i$, which is used to predict $r_i = f(X_i^*) \in \mathbb{R}$.

2.1 Multi-Axis Stratified Sampling (MASS)

MASS is an unsupervised tile selection module that reduces a large tile bag, X_i, to a fixed-size subset, X_i^*, by (i) feature-space quality filtering and (ii) stratified sampling across complementary phenotype-inspired axes. This design is consistent with unsupervised morphology-driven slide representation learning that leverages visual prototypes without task labels [30].

Quality-Aware Filtering. Given $X_i = \{x_j\}_{j=1}^{N_i}$, each tile embedding, $x_j \in \mathbb{R}^D$, is evaluated using robust statistical criteria to filter out low-quality or uninformative tiles (shown in Fig. 1(a)) not removed by tissue masking. We compute (i) feature magnitude $m_j = \|x_j\|_2$ and discard extreme values using slide-adaptive percentiles ($m_j < P_2$ or $m_j > P_{98}$); (ii) outlierness $o_j = \max_d |(x_{j,d} - \mu_d)/(\sigma_d + \epsilon)|$ with a retention threshold of $o_j > 3.5$; and (iii) feature entropy $e_j = -\sum_{d=1}^{D} p_{j,d} \log(p_{j,d} + \epsilon)$, removing tiles with $e_j < P_3$. Finally MASS computes a composite quality score $Q(x_j) \in [0, 1]$, which integrates m_j, o_j, e_j. Tiles with $Q(x_j) \geq 0.30$ comprise $X_{\text{clean}} \subset X_i$.

Proxy-based Stratification. The filtered tiles, X_{clean}, are stratified along four axes computed from embedding and coordinate-derived proxies. These axes are designed to capture distinct modes of morphologic variation relevant to survival modeling, without using survival labels. Fig. 2 provides qualitative examples of the tile phenotypes retrieved by each axis. Concretely, morphologic heterogeneity (MH) emphasizes visually heterogeneous tiles; tissue regularity (TR) emphasizes more regular/structured patterns; microenvironmental interface (MI) emphasizes spatial neighborhoods with high local variation; and tissue diversity (TD) promotes inclusion of rare morphologies via feature-space diversity. Each axis is operationalized by a simple proxy score and a fixed budget. We define the four axis scores as follows:

MH: Tiles exhibiting high morphologic variability are identified using feature variance, $s_j^{MH} = Var(x_j)$.

TR: Tiles with more regular/structured appearances are scored using the ratio of feature magnitude to variance $s_j^{TR} = \|x_j\|_2 / (Var(x_j) + \epsilon)$, where ϵ is a small constant to ensure numerical stability.

MI: Tiles located at tissue interfaces are identified by measuring local spatial variation in MH $s_j^{MI} = Std\left(\left\{s_k^{MH} | c_k \in \mathcal{N}_\tau(c_j)\right\}\right)$, where $\mathcal{N}_\tau(c_j)$ denotes the set of τ nearest spatial neighbors of tiles j and k inside $\mathcal{N}_\tau(c_j)$.

TD: Rare or underrepresented morphologies are selected using farthest-point sampling in the feature space to promote tissue diversity among retained tiles. We select a total of K tiles across 4 axes (K_{MH},K_{TR},K_{MI},K_{TD}), satisfying $\sum_a K_a = K$. Axis fractions (f_{MH},f_{TR},f_{MI},f_{TD}) are tuned with Optuna hyperparameter search [31] without survival labels by maximizing a proxy objective based on embedding statistics (mean selected-tile variance), f_{TD} = 1 - (f_{MH}, f_{TR}, f_{MI}) and $f_{TD} \geq 0.05$. To ensure mutually exclusive semantic assignments, selection is performed sequentially (MH->TR->MI->TD), removing selected tiles from the candidate pool at each step. The final selected subset, X_i^*, is obtained from X_{clean} by $X_i^* = \bigcup_{a \in \{MH,TR,MI,TD\}} s_j^a$, $|X_i^*| = K$, where s_j^a denotes the subset of tiles selected along axis a and assigned a semantic label $l_j \in \{0, 1, 2, 3\}$ corresponding to its axis, which is subsequently used for semantic hypergraph construction (Fig. 1b-c).

2.2 Semantic Hierarchical Hypergraph Network

To model higher-order tissue organization, we represent the MASS-selected tile set X_i^* as a semantic hierarchical hypergraph. Following prior motivation for higher-order relational prognosis modeling [32,33], the model uses tile nodes and semantic super-node groups to define higher-order hyperedges over selected tiles. For each MASS semantic label $l \in \{0, 1, 2, 3\}$, tiles within the label group are clustered in embedding space using k-means to form semantic super-nodes. The number of super-nodes is determined by the fixed granularity of tiles/super-node (reported in Section 3), and each super-node is represented by the mean embedding of its assigned member tiles. We construct the tile-level incidence matrix H using three hyperedge families: (i) containment hyperedges define one hyperedge per super-node, containing all tiles assigned to that super-node; (ii) intra-semantic hyperedges group each anchor super-node with its k_{intra} nearest super-nodes from the same MASS semantic label, capturing within-axis coherence; (iii) inter-semantic hyperedges group each anchor super-node with its k_{inter} nearest super-nodes from different allowed MASS labels, enabling cross-compartment interactions such as tumor–stroma or tumor–immune interfaces. Each nearest-neighbor hyperedge therefore contains the tiles belonging to the anchor super-node and its selected neighboring super-nodes (fixed topology parameters are reported in Section 3.2). Feature propagation is performed with degree-normalized hypergraph message passing, $X^{(m)} = X^{(m-1)} + D_v^{-1} H D_e^{-1} H^T X^{(m-1)} \Theta^{(m)}$, where $X^{(0)} \in \mathbb{R}^{K \times D}$ are tile embeddings in X_i^*, H is the incidence matrix, D_v and D_e are node and hyperedge degree matrices, and $\Theta^{(m)} \in \mathbb{R}^{D \times D}$ is a learnable weight matrix at the message-passing layer m.

3 Experiments

Dataset. TRIAGE-MIL was evaluated on six H&E-stained, formalin fixed, paraffin-embedded (FFPE) WSI patient cohorts, using one representative tumor slide per patient: an internal single-center colorectal adenocarcinoma cohort (790 patients who underwent surgical resection at Stanford University Medical Center between the years 2009-2022 (1 representative tumor WSI/patient; endpoint: overall-survival; 228 events; 71.1% censoring) and five public TCGA cohorts with survival outcomes obtained from the TCGA Pan-Cancer Clinical Data Resource [34]: BRCA (1008 patients; 141 events; 86.0% censoring), LUAD (382; 128; 66.5%), COAD-READ (565; 120; 78.8%), STAD (365; 145; 60.3%), and BLCA (385; 175; 54.5%).We used patient-level stratified 5-fold cross-validation with splits stratified jointly by event status and survival-time quartile. In each fold, data were partitioned into three patient-disjoint sets: 60% for training, 20% for validation and 20% as a held-out test set. Fold-level C-indices were computed exclusively on the held-out test set, which was not used during training or checkpoint selection, and cohort-level performance was reported as mean±s.d. (standard deviation) across folds. The same protocol was applied to all 13 baselines.

Implementation Details. WSI were processed at 20× magnification using CLAM tissue masking [3], tiled into non-overlapping 256 × 256 patches, and encoded with the UNI foundation model [11] to obtain 1024-d embeddings. All models used a unified training protocol with AdamW [35], learning rate 2×10^{-4}, batch size 4, and gradient accumulation 8, training up to 200 epochs with early stopping on validation performance. TRIAGE-MIL applies feature-space quality filtering within MASS to remove low-quality embeddings before tile selection and uses a fixed per-slide budget of K=4096 tiles for hypergraph training. For baselines requiring fixed-size bags, the same K-tile input budget was used under their native sampling strategies, while MASS-based selection was used only for TRIAGE-MIL. The value of K was set to 4096 as a compute-performance trade-off: smaller budgets under-covered prognostic morphology, whereas larger budgets provided only marginal performance gains at substantially higher memory cost. Hypergraph topology was fixed across cohorts (k_{intra}=10, k_{inter}=5, tiles/super-node =12) to standardize comparisons.

3.1 Survival Prognostication Performance

We compared TRIAGE-MIL against 13 baselines, including pooling (Mean, Max), attention-based MIL (ABMIL, CLAM, DSMIL, DTFD-MIL), sequence-based MIL (TransMIL, MambaMIL), graph-based MIL (DeepAttnMISL, IBMIL), and survival-specific methods (OTSurv, ILRA, PANTHER). Table 1 reports the cohort-wise C-indices. TRIAGE-MIL achieved the best mean C-index (0.685), improving over the strongest baseline, OTSurv (0.652), by 3.3 percentage points. Kaplan–Meier analysis showed significant risk-group separation (log-rank $p < 0.05$) across all six cohorts (Fig. 3a). For KM analysis, risk scores were generated only for held-out test patients that were not used for training or

Table 1. Comparison of C-index performance (mean±s.d.) across five TCGA cohorts and an internal colorectal cancer cohort for different survival prediction methods. Best and second-best results are bolded and underlined, respectively.

Model	In-House (790 cases)	CRC (565 cases)	LUAD (382 cases)	STAD (365 cases)	BLCA (385 cases)	BRCA (1008 cases)	Mean±s.d. (6 cohorts)
Max-Pooling [13]	0.672±0.047	0.648±0.031	0.616±0.077	0.584±0.051	0.636±0.025	0.625±0.040	0.630±0.029
Mean-Pooling [13]	0.673±0.030	0.662±0.068	0.622±0.028	0.600±0.033	0.633±0.025	0.649±0.051	0.640±0.025
ABMIL [14]	0.667±0.029	0.642±0.054	0.607±0.031	0.596±0.027	0.621±0.023	0.639±0.054	0.629±0.024
CLAM [3]	0.690±0.027	0.665±0.076	0.620±0.024	0.605±0.033	0.632±0.036	0.666±0.056	0.646±0.030
TransMIL [36]	0.687±0.045	0.657±0.061	0.624±0.034	0.610±0.025	0.642±0.038	0.671±0.069	0.649±0.027
DSMIL [37]	0.675±0.027	0.662±0.071	0.617±0.020	0.607±0.029	0.637±0.030	0.663±0.052	0.644±0.025
DTFD-MIL [38]	0.654±0.023	0.658±0.066	0.624±0.035	0.605±0.030	0.627±0.033	0.649±0.049	0.636±0.019
MambaMIL [27]	0.629±0.018	0.646±0.055	0.601±0.030	0.601±0.026	0.632±0.026	0.631±0.041	0.623±0.029
DeepAttnMISL [15]	0.615±0.017	0.644±0.055	0.581±0.048	0.608±0.029	0.628±0.032	0.626±0.031	0.617±0.020
OTSurv [28]	0.695±0.028	0.663±0.074	0.630±0.027	0.618±0.026	0.643±0.026	0.665±0.050	0.652±0.028
IB-MIL [39]	0.687±0.029	0.662±0.071	0.611±0.018	0.604±0.033	0.635±0.035	0.663±0.058	0.644±0.029
ILRA [40]	0.684±0.027	0.657±0.071	0.606±0.017	0.600±0.033	0.639±0.029	0.660±0.057	0.641±0.029
PANTHER [30]	0.681±0.030	0.660±0.067	0.617±0.034	0.609±0.034	0.639±0.029	0.660±0.056	0.644±0.025
TRIAGE-MIL	**0.712**±0.035	**0.687**±0.061	**0.676**±0.013	**0.651**±0.038	**0.668**±0.021	**0.714**±0.062	**0.685**±0.025

Table 2. Performance impact of model design components in TRIAGE-MIL.

Method/ Parameter	In-House	CRC	LUAD	STAD	BLCA	BRCA	Mean±s.d.
TRIAGE-MIL **(Final)**	0.712±0.035	0.687±0.061	0.676±0.013	0.651±0.038	0.668±0.021	0.714±0.062	**0.685**±0.025
Ablation on model design choices							
w/o Quality-Aware Filtering ($Q(x_j)$)	0.701±0.038	0.679±0.064	0.665±0.015	0.640±0.040	0.660±0.023	0.708±0.065	0.676±0.026
w/o MASS (Naive Top-K sampling)	0.688±0.041	0.670±0.065	0.650±0.018	0.628±0.042	0.652±0.025	0.699±0.068	0.665±0.026
w/o Semantic Labels (ℓ_j)	0.695±0.039	0.678±0.063	0.661±0.016	0.635±0.039	0.658±0.024	0.705±0.064	0.672±0.026
w/o Relational modeling	0.685±0.042	0.665±0.066	0.645±0.019	0.622±0.043	0.648±0.026	0.692±0.070	0.660±0.027
w/o Hierarchical Super-nodes	0.698±0.037	0.675±0.062	0.659±0.015	0.638±0.040	0.655±0.024	0.702±0.063	0.671±0.025

checkpoint selection. In each fold, high- and low-risk groups were assigned using the median risk threshold estimated from the corresponding training set, and the assigned held-out test groups were pooled across folds for log-rank testing (one out-of-fold risk prediction contributed per patient).

3.2 Ablation Studies

Model choice: Table 2 quantifies the contribution of each component under controlled ablations. Replacing MASS with naïve top-K sampling using only the MH proxy $s_j^{MH} = Var(x_j)$ over X_{clean} with K=4096 reduced the mean C-index from 0.685 to 0.665, indicating that multi-axis stratification provides complementary coverage beyond a single proxy. Removing quality-aware filtering $X_{\text{clean}} \equiv X_i$ decreased the performance (0.685→0.676), suggesting that filtering low-quality embeddings prior to selection improves stability. Eliminating semantic labels, l_j, in hypergraph construction also reduced performance (0.685→0.672). Disabling relational modeling (i.e., removing hypergraph message passing and aggregating MASS-selected tiles directly) yielded the largest drop (0.685→0.660), supporting the importance of modeling cross-compartment interactions. Removing hierarchical super-nodes reduced the performance (0.685 → 0.671).

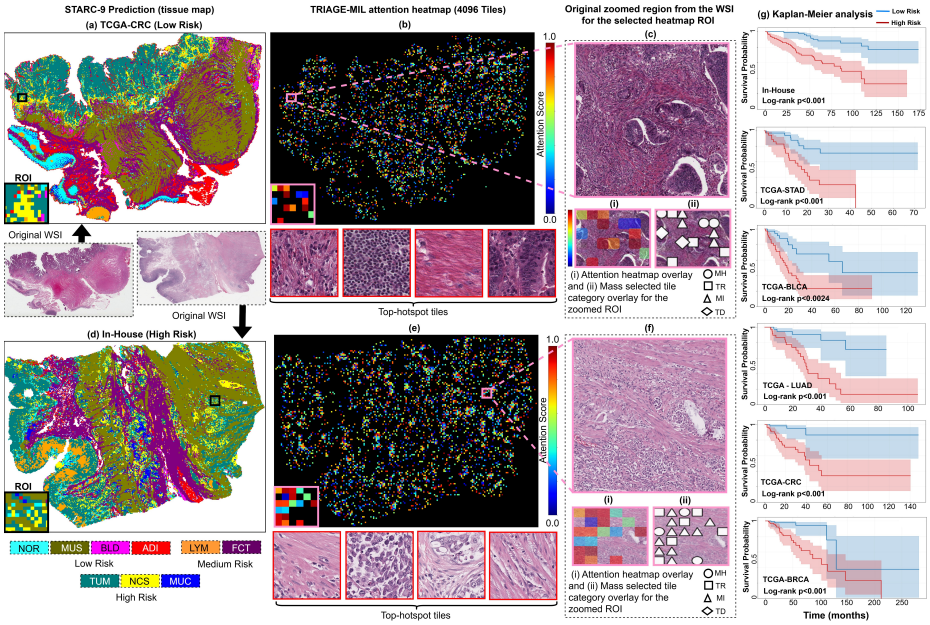

Fig. 3. TRIAGE-MIL slide and cohort level interpretability. (a,d) STARC-9 tissue-type predictions for representative low-risk (TCGA-CRC) and high-risk (internal) slides, with the ROI indicated. (b,e) TRIAGE-MIL tile-level attention over the MASS-selected fixed budget (K=4096); attention is visualized only on the selected subset, yielding a sparse/discrete map. (c,f) Corresponding ROI from the original WSI, with (i) attention overlay and (ii) MASS semantic categories (MH/TR/MI/TD) for the tiles within the ROI. Representative top-attention tiles are shown beneath (b,e). (g) Kaplan–Meier curves show risk-group separation across six cohorts (log-rank p).

3.3 Qualitative Analysis and Interpretability

Fig. 3 summarizes TRIAGE-MIL interpretability at both cohort and slide levels. For representative low-risk (top) and high-risk (bottom) cases, we first show tile-level tissue-type maps produced by a classifier trained on STARC-9 [41] (Fig. 3a,d) to provide morphological context. We then visualize TRIAGE-MIL attention over the MASS-selected tiles (K=4096) (Fig. 3b, e); because inference operates only on the selected subset, attention appears as a sparse tile set rather than a dense pixel-level map. Zoomed ROIs show the corresponding MASS semantic categories (MH/TR/MI/TD) for the selected tiles (Fig. 3c, f). In the high-risk case, attention concentrates on tumor-associated and interface-like regions within the ROI, and the top hotspots highlight morphologies consistent with architectural disruption. In contrast, the low-risk case shows a more distributed attention pattern, with hotspots more often aligned with organized tissue compartments. Overall, these visualizations suggest that TRIAGE-MIL bases risk estimates on MASS-selected tile phenotypes and their contextual relationships captured by the hypergraph.

4 Conclusion

We propose TRIAGE-MIL, a two-stage framework that uses MASS to select a compact, informative tile subset by filtering artifacts and stratifying tiles along four phenotype-inspired axes, applies a semantic hierarchical hypergraph MIL model to capture higher-order tissue interactions beyond pairwise graphs. Experiments show that TRIAGE-MIL achieves consistently strong performance across six cohorts and captures biologically meaningful features. A key limitation is that, due to compute-budget constraints, we did not perform an exhaustive hyperparameter sensitivity analysis or a head-to-head runtime and memory comparison against all baselines. Future work will systematically assess robustness to graph connectivity and filtering thresholds, evaluate computational complexity and scalability, and extend TRIAGE-MIL to multimodal risk modeling to improve discrimination and calibration in heterogeneous patient cohorts.

Acknowledgments. Funding for this study was provided by the United States National Cancer Institute (NCI), National Institutes of Health (NIH) (R01 CA270437).

Disclosure of Interests. The authors have no relevant competing interests to disclose.

References

1. Echle, A. et al.: Deep learning in cancer pathology: a new generation of clinical biomarkers (2 2021)
2. Bera, K. et al.: Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nature Reviews Clinical Oncology **16**, 703–715 (11 2019)
3. Lu, M.Y. et al.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**, 555–570 (6 2021)
4. Mobadersany, P. et al.: Predicting cancer outcomes from histology and genomics using convolutional networks. Proceedings of the National Academy of Sciences of the United States of America **115**, E2970–E2979 (10 2017)
5. Wulczyn, E. et al.: Deep learning-based survival prediction for multiple cancer types using histopathology images. PloS one **15**, e0233678 (2020)
6. Parvaiz, A. et al.: From pixels to prognosis: A survey on ai-driven cancer patient survival prediction using digital histology images. Journal of Imaging Informatics in Medicine **37**, 1728–1751 (8 2024)
7. Cen, M. et al.: C2mil: Synchronizing semantic and topological causalities in multiple instance learning for robust and interpretable survival analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 24392–24401 (2025)
8. Volinsky-Fremond, S. et al.: Prediction of recurrence risk in endometrial cancer with multimodal deep learning. Nature Medicine **30**, 1962–1973 (7 2024)
9. Li, C. et al.: A novel computational approach for survival prediction via fusion of whole-slide images and clinical data. J Clin Oncol **43**, e15119 (2025)
10. Chen, Y. et al.: Survmamba: State space model with multi-grained multi-modal interaction for survival prediction. arXiv (12 2024)

11. Chen, R.J. et al.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: CVPR. pp. 16144–16155 (2022)
12. Hou, L. et al.: Patch-based convolutional neural network for whole slide tissue image classification. In: CVPR. pp. 2424–2433 (3 2016)
13. Dietterich, T.G. et al.: Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence **89**, 31–71 (1997)
14. Ilse, M. et al.: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning. pp. 2127–2136 (6 2018)
15. Yao, J. et al.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Medical Image Analysis **65** (10 2020)
16. Saltz, J. et al.: Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell Reports **23**, 181–193.e7 (4 2018)
17. Chen, R.J. et al.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. arXiv (7 2021)
18. Maley, C.C. et al.: Classifying the evolutionary and ecological features of neoplasms (10 2017)
19. Jaume, G. et al.: Histocartography: A toolkit for graph analytics in digital pathology. In: Proceedings of the MICCAI Workshop on Computational Pathology. pp. 117–128 (7 2021)
20. Brussee, S. et al.: Graph neural networks in histopathology: Emerging trends and future directions. Medical Image Analysis **101** (4 2025)
21. Feng, Y. et al.: Hypergraph neural networks. In: AAAI. pp. 3558–3565 (2 2019)
22. Bai, S. et al.: Hypergraph convolution and hypergraph attention. Pattern Recognition **110**, 107637 (10 2020)
23. Kothari, S. et al.: Removing batch effects from histopathological images for enhanced cancer diagnosis. IEEE journal of biomedical and health informatics **18**, 765–772 (2013)
24. Tellez, D. et al.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical image analysis **58**, 101544 (4 2020)
25. Stacke, K. et al.: Measuring domain shift for deep learning in histopathology. IEEE Journal of Biomedical and Health Informatics **25**, 325–336 (2 2021)
26. Howard, F.M. et al.: The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nature Communications **12** (12 2021)
27. Yang, S. et al.: Mambamil: Enhancing long sequence modeling with sequence reordering in computational pathology. In: proceedings of Medical Image Computing and Computer Assisted Intervention. pp. 296 –306 (3 2024)
28. Ren, Q. et al.: 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 (6 2025)
29. Campanella, G. et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine **25**, 1301–1309 (8 2019)
30. Song, A.H. et al.: Morphological prototyping for unsupervised slide representation learning in computational pathology. In: CVPR. pp. 11569–11579 (5 2024)
31. Akiba, T. et al.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 2623–2631 (7 2019)

32. Ding, S. et al.: Hgmsurvnet: A two-stage hypergraph learning network for multi-modal cancer survival prediction. Medical Image Analysis **104** (8 2025)
33. Yang, H. et al.: Mmsurv: a multimodal multi-instance multi-cancer survival prediction model integrating pathological images, clinical information, and sequencing data. Briefings in Bioinformatics **26** (5 2025)
34. Liu, J. et al.: An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell **173**, 400–416.e11 (4 2018)
35. Loshchilov, I. et al.: Decoupled weight decay regularization. ICLR (1 2019)
36. Shao, Z. et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In: NeurIPS. pp. 2136–2147 (10 2021)
37. Li, B. et al.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: CVPR. pp. 14318–14328 (4 2021)
38. Zhang, H. et al.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. arXiv (3 2022)
39. Lin, T. et al.: Interventional bag multi-instance learning on whole-slide pathological images. In: CVPR. pp. 15802–15812 (3 2023)
40. Xiang, J. et al.: Exploring low-rank property in multiple in-stance learning for whole slide image classification. In: ICLR (2024)
41. Subramanian, B. et al.: Starc-9: A large-scale dataset for multi-class tissue classification for crc histopathology. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2025)