DPsurv: Dual-Prototype Evidential Fusion for Uncertainty-Aware And Interpretable Whole Slide Image Survival Prediction

Yucheng Xing[1,2] · Ling Huang[1,3] · Jingying Ma[1] · Ruping Hong[4] · Jiangdong Qiu[4]
Pei Liu[5] · Kai He[1] · Huazhu Fu[6] · Mengling Feng[1]

[1]National University of Singapore · [2]NUS Guangzhou · [3]Imperial College London
[4]Peking Union Medical College Hospital · [5]Hunan University · [6]A*STAR

ICML International Conference On Machine Learning

SCAN for code

Motivation

WSIs hold rich histopathological information for cancer survival prediction. Yet tissue heterogeneity and censoring make reliable prognosis difficult, and most methods still give a single risk score with little interpretability.

- Existing methods handle the gigapixel scale but rarely address heterogeneity. Tumor, stroma, and necrosis each carry their own prognostic value.
- Interpretability stops at the feature level, such as attention heatmaps or coarse clusters. No transparent path runs from feature to reasoning to decision.
- Point estimates leave out uncertainty, and most uncertainty research still focuses on classification rather than survival.

Q: Can we turn WSIs into tissue-level evidence rather than point scores, so survival prediction becomes both interpretable and uncertainty aware?

Core Idea

Decompose each WSI into distinct tissue components. Each component carries its own risk evidence, a Gaussian random fuzzy number (GRFN) with its own uncertainty. Fusing them makes the prediction interpretable and uncertainty aware by design.

Evidence form:

$$\tilde{Y} \sim \tilde{\mathcal{N}}(u, \sigma^2, h)$$

$u \in \mathbb{R}$	$\sigma \in \mathbb{R}$	$h \in \mathbb{R}_{\geq 0}$
plausible log survival time	aleatoric uncertainty	epistemic uncertainty

Takeaways

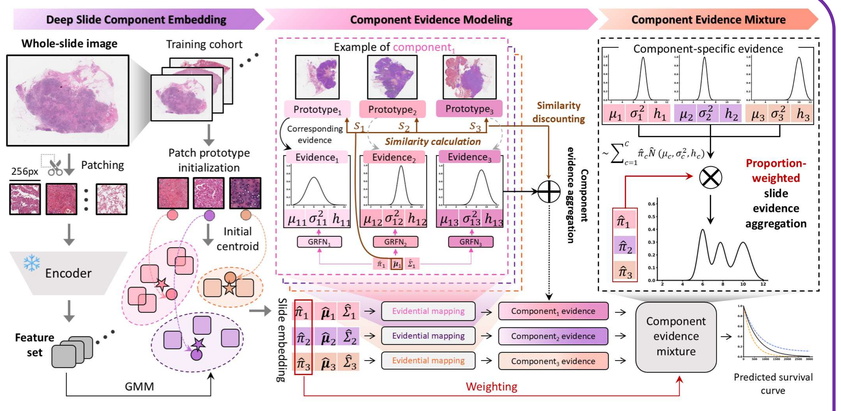
IDEA
WSIs are represented as mixtures of tissue-level risk evidence.

THEORY
Belief and plausibility provide calibrated uncertainty bounds.

METHOD
Prototype-based evidential modeling captures phenotype-specific survival signals.

PRACTICE
Best calibration and strong performance across 5 TCGA cohorts.

Method



- **Feature level:** a GMM with patch prototypes decomposes each WSI into a few distinct tissue components.

$$\mathbf{z}^{\mathbf{i}}_{\mathbf{WSI}} = \left[\sum_{n=1}^{N_i} \phi^i(\mathbf{z}^{\mathbf{i}}_{\mathbf{n}}, \mathbf{v}_1), \ldots, \sum_{n=1}^{N_i} \phi^i(\mathbf{z}^{\mathbf{i}}_{\mathbf{n}}, \mathbf{v}_C)\right] = \left[\widehat{\pi_1^i}, \widehat{\eta_1^i}, \widehat{\Sigma_1^i}, \ldots, \widehat{\pi_C^i}, \widehat{\eta_C^i}, \widehat{\Sigma_C^i}\right].$$

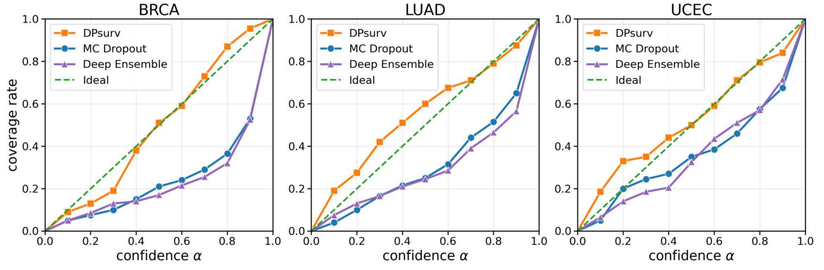
- **Reasoning level:** component prototypes turn each tissue component into one piece of risk evidence as a GRFN.

$$\left[\widehat{\pi_c}, \widehat{\eta_c}, \widehat{\Sigma_c}\right] \xrightarrow{\text{Evidence Modelling}} \tilde{\mathcal{N}}(\mu_c, \sigma_c^2, h_c)$$

- **Decision level:** the component evidence is fused into a single interval valued survival prediction.

$$\tilde{Y} \sim \sum_{c=1}^{C} \widehat{\pi_c}\, \tilde{\mathcal{N}}(\mu_c, \sigma_c^2, h_c)$$

Uncertainty-Awareness

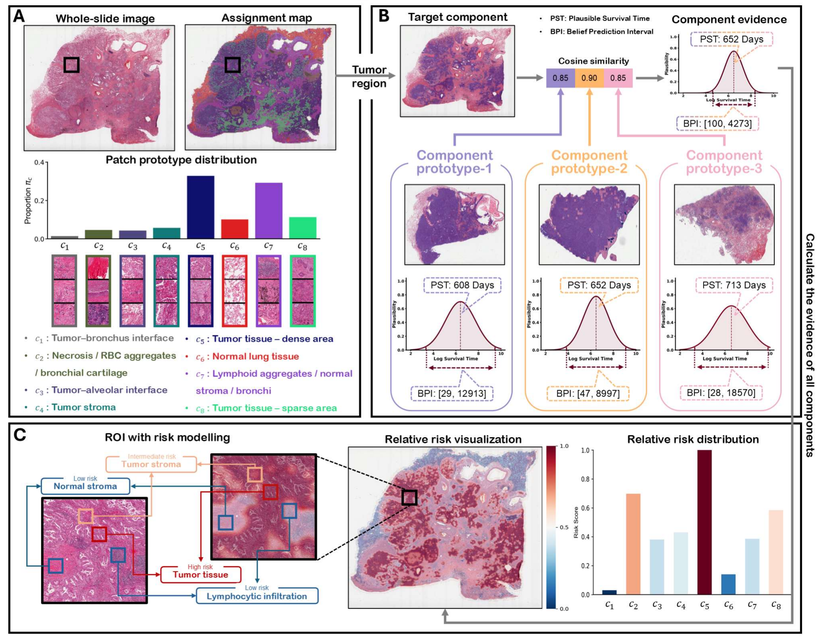


Results

	Methods	BRCA C-index↑	BRCA IBS↓	BRCA IBLL↓	BLCA C-index↑	BLCA IBS↓	BLCA IBLL↓	LUAD C-index↑	LUAD IBS↓	LUAD IBLL↓
✗ UA	ABMIL	0.687±0.06	0.752±0.14	3.019±1.16	0.557±0.04	0.535±0.09	1.540±0.34	0.611±0.11	0.627±0.19	1.938±0.70
	TransMIL	0.607±0.11	0.934±0.09	5.231±2.01	0.578±0.08	0.823±0.12	3.239±1.10	0.618±0.08	0.792±0.16	3.351±1.21
	DSMIL	0.652±0.04	0.694±0.16	2.208±0.54	0.583±0.03	0.417±0.10	1.080±0.26	0.619±0.10	0.473±0.11	1.227±0.28
	AttnMISL	0.654±0.06	0.801±0.15	3.563±1.63	0.545±0.05	0.499±0.12	1.537±0.44	0.593±0.12	0.465±0.17	1.314±0.54
	ILRA	0.665±0.05	0.932±0.04	4.491±1.09	0.552±0.07	0.865±0.07	3.999±1.02	0.578±0.07	0.846±0.17	4.654±0.74
	PANTHER	0.696±0.05	0.837±0.08	2.683±0.59	0.601±0.06	0.530±0.11	1.331±0.27	0.588±0.04	0.667±0.16	1.819±0.54
✓ UA	EVREG	0.585±0.03	0.665±0.01	8.467±0.51	0.587±0.01	0.461±0.00	5.469±0.06	0.565±0.02	0.483±0.00	5.964±0.18
	UMSA	0.640±0.08	0.730±0.11	2.336±0.51	0.573±0.07	0.501±0.10	1.381±0.35	0.634±0.10	0.569±0.17	1.663±0.59
	BayesMIL	0.707±0.05	0.721±0.06	2.024±0.31	0.602±0.06	0.414±0.08	1.058±0.19	0.622±0.11	0.461±0.08	1.180±0.18
	DPsurv	**0.720**±0.03	**0.199**±0.03	**0.566**±0.08	**0.625**±0.06	**0.410**±0.12	**0.855**±0.14	**0.667**±0.08	**0.381**±0.06	**1.130**±0.24

	Methods	UCEC C-index↑	UCEC IBS↓	UCEC IBLL↓	KIRC C-index↑	KIRC IBS↓	KIRC IBLL↓	Average C-index↑	Average IBS↓	Average IBLL↓
✗ UA	ABMIL	0.636±0.11	0.788±0.07	2.802±0.17	0.703±0.10	0.517±0.20	1.670±0.94	0.639	0.644	2.194
	TransMIL	0.698±0.07	0.936±0.05	4.590±0.93	0.725±0.08	0.755±0.13	3.313±1.09	0.645	0.848	3.945
	DSMIL	0.690±0.09	0.652±0.13	1.871±0.45	0.693±0.05	0.455±0.11	1.216±0.28	0.647	0.538	1.520
	AttnMISL	0.673±0.10	0.703±0.14	2.190±0.60	0.683±0.03	0.501±0.15	1.479±0.54	0.630	0.594	2.017
	ILRA	0.716±0.07	0.920±0.04	4.074±0.92	0.659±0.09	0.858±0.07	4.325±1.13	0.634	0.884	4.309
	PANTHER	0.709±0.06	0.866±0.10	3.045±0.86	0.683±0.09	0.701±0.15	1.926±0.48	0.655	0.720	2.161
✓ UA	EVREG	0.657±0.02	0.612±0.01	7.071±0.29	0.594±0.01	0.541±0.00	5.544±0.17	0.598	0.552	6.503
	UMSA	0.641±0.11	0.758±0.14	2.717±1.08	0.693±0.08	0.506±0.17	1.524±0.57	0.636	0.613	1.924
	BayesMIL	0.736±0.10	0.679±0.12	1.941±0.47	0.703±0.05	0.434±0.15	1.171±0.41	0.674	0.542	1.475
	DPsurv	**0.766**±0.05	**0.251**±0.08	**0.692**±0.19	**0.741**±0.08	**0.310**±0.11	**0.879**±0.28	**0.704**	**0.310**	**0.824**

Interpretability



- **Patch prototype:** a global morphological pattern shared across slides.
- **Component prototype:** a local risk expert for a representative phenotype.