Interpretable Vision-Language Survival Analysis with Ordinal Inductive Bias for Computational Pathology

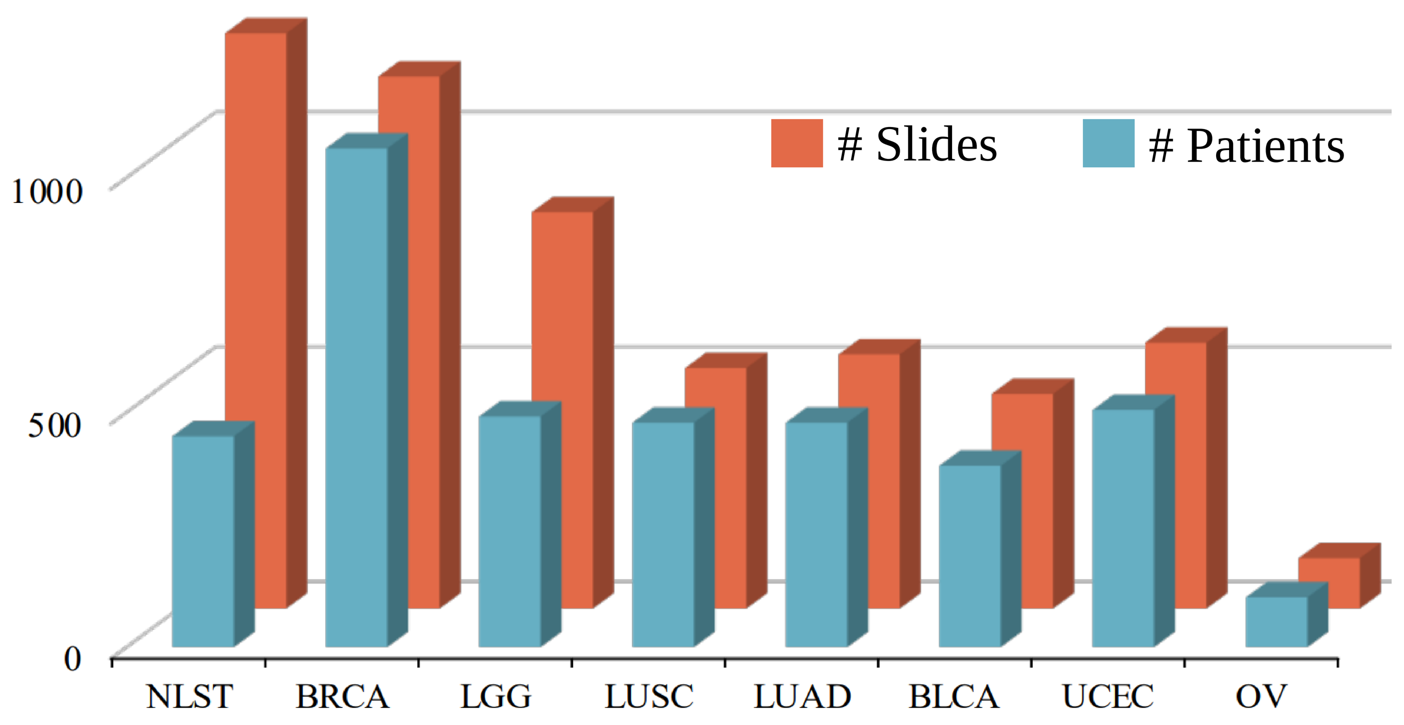
Pei Liu Luping Ji ✉ Jiaxiang Gou Bo Fu Mao Ye | School of Computer Science and Engineering, UESTC

Paper & Code

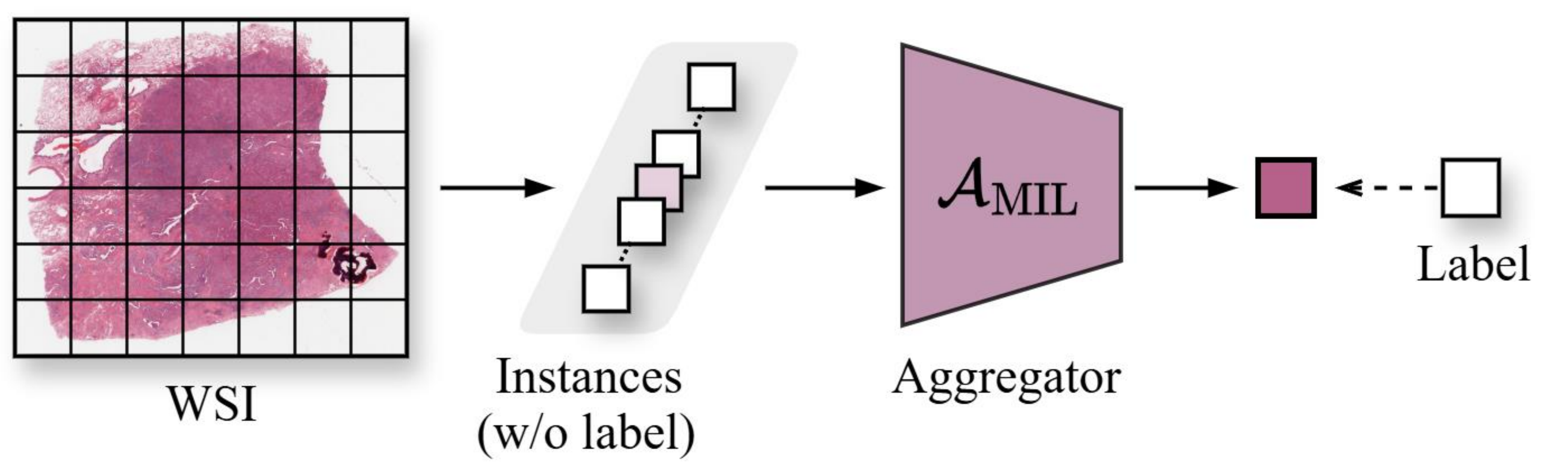
Motivation

■ The survival analysis of WSI (Whole-Slide Image) data has always faced two critical challenges:

- **Scarce training data** (often limited to 1,000 patients)

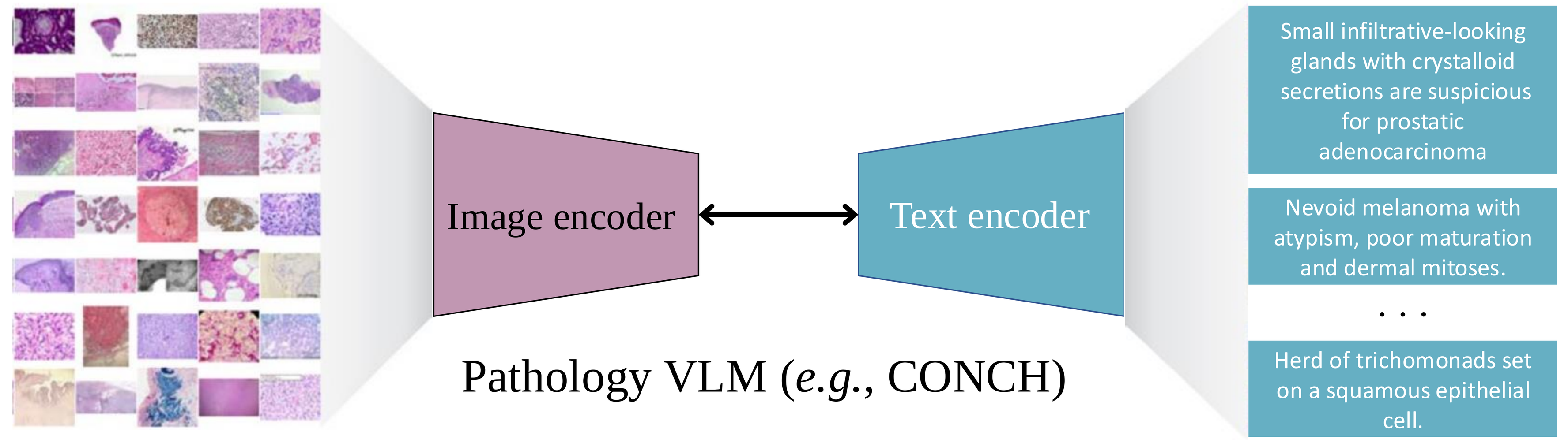


- Learning from gigapixel images under **weak supervision**



■ **VLMs** (Vision-Language Models) offer promising means:

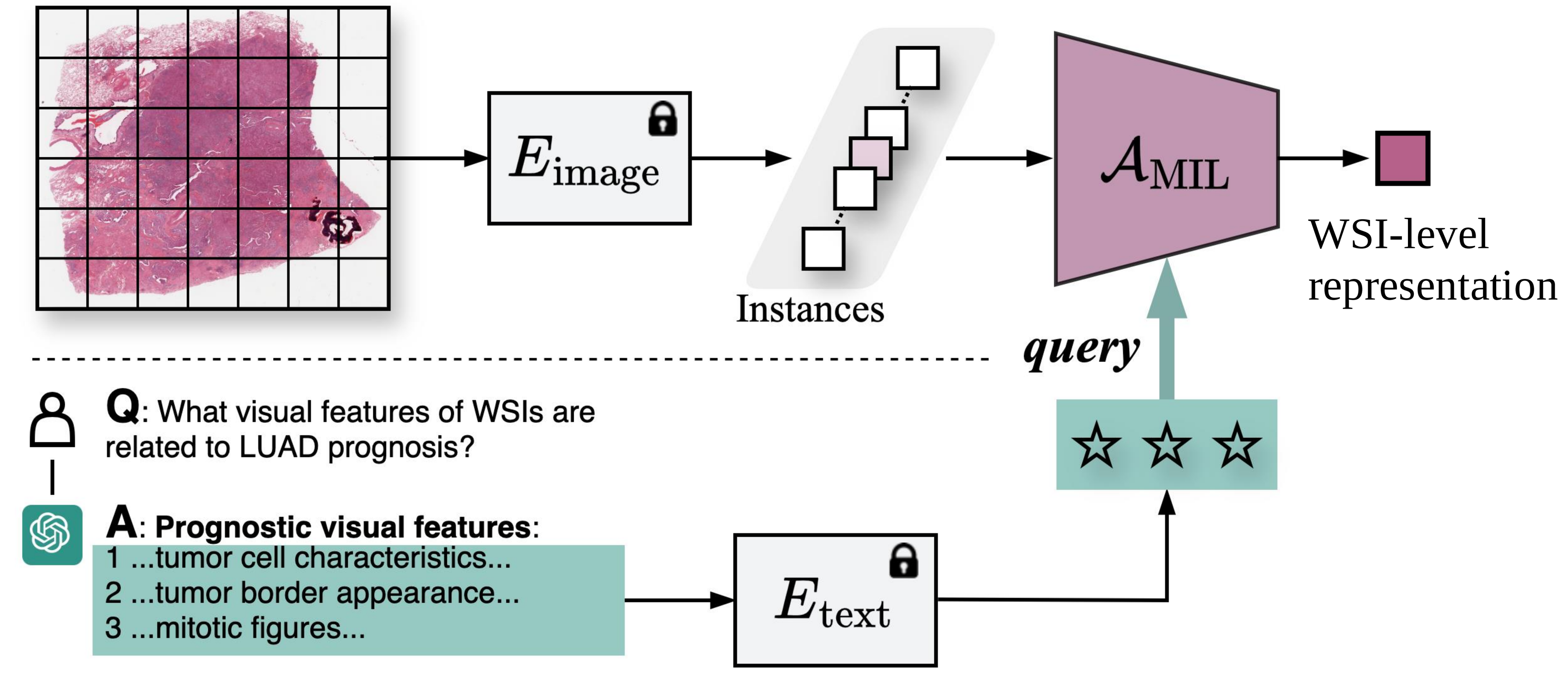
- VLMs show clear advantages in **data-efficiency**
- Prior knowledge as text prompt to provide **auxiliary signals**



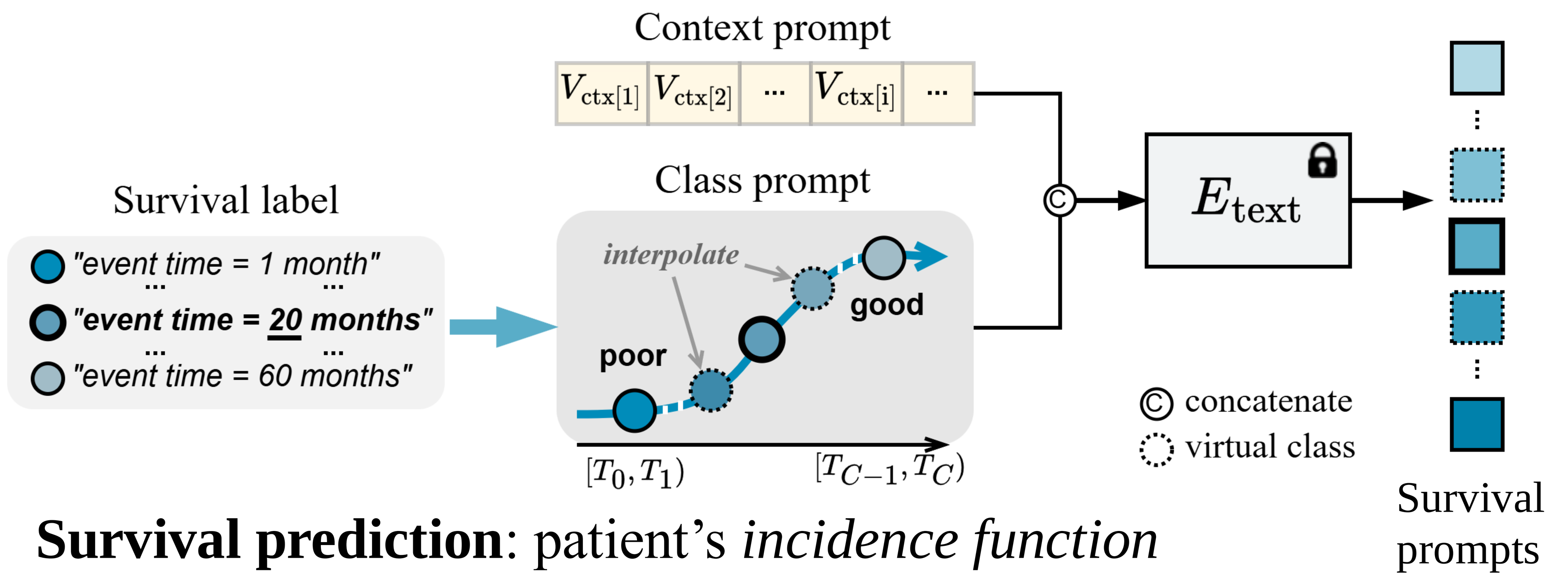
Method

■ **Vision-Language Survival Analysis** (VLSA):

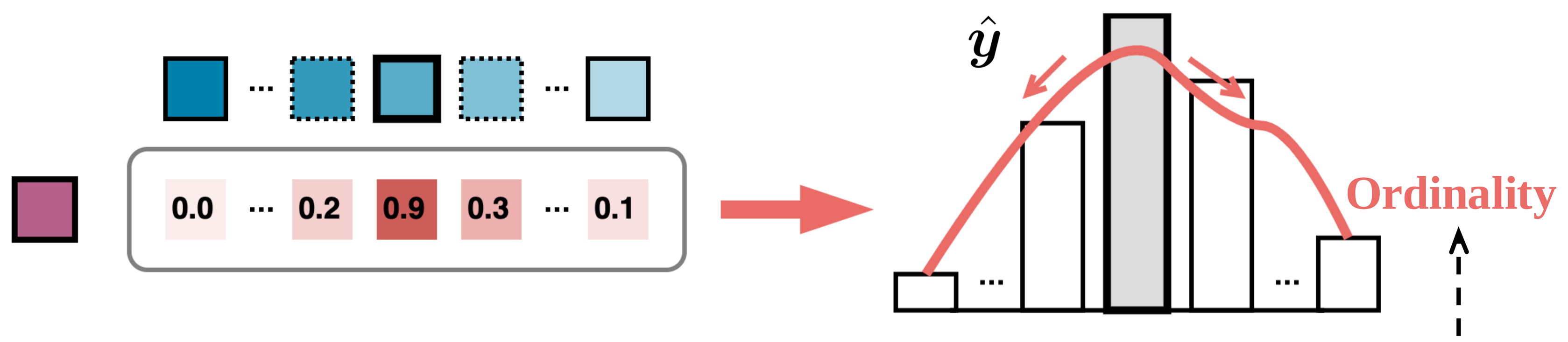
- **WSI representation learning** with *prognostic priors* (from LLMs)



- **Survival prompt learning**: context & *ordinal class prompts*



- **Survival prediction**: patient's *incidence function*



- **Objective function**: $\mathcal{L} = \mathcal{L}_{\text{MLE}} + \beta \cdot \mathcal{L}_{\text{EMD}}$

$$\mathcal{L}_{\text{MLE}} = -\left[\delta \cdot \log(\hat{y}_c) + (1-\delta) \cdot \log\left(1 - \sum_{i=1}^{c-1} \hat{y}_i\right)\right]$$

Experiments & Results

■ **Comparison with baselines & Ablation study on MIL methods**

	Method	TCGA BLCA	TCGA BRCA	TCGA GBMLGG	TCGA LUAD	TCGA UCEC	Average CI	Average MAE	D-cal Count
V	ABMIL	0.5581 (± 0.031)	0.5825 (± 0.035)	0.7935 (± 0.032)	0.6121 (± 0.050)	0.6667 (± 0.033)	0.6426	29.83	4
V	TransMIL	0.5885 (± 0.055)	0.6140 (± 0.060)	0.7956 (± 0.015)	0.5708 (± 0.050)	0.6380 (± 0.067)	0.6414	30.43	5
V	ILRA	0.5549 (± 0.053)	0.5705 (± 0.067)	0.7742 (± 0.014)	0.5179 (± 0.081)	0.6503 (± 0.064)	0.6136	32.59	4
V	R^2T-MIL	0.5775 (± 0.024)	0.5473 (± 0.095)	0.7757 (± 0.024)	0.5711 (± 0.076)	0.6510 (± 0.087)	0.6245	32.54	4
V	DeepAttnMISL	0.5646 (± 0.035)	0.5346 (± 0.036)	0.6750 (± 0.048)	0.4678 (± 0.039)	0.6259 (± 0.085)	0.5736	52.10	5
V	Patch-GCN	0.6124 (± 0.031)	0.6375 (± 0.033)	0.7999 (± 0.021)	0.5922 (± 0.053)	0.7212 (± 0.025)	0.6726	26.70	2
VL	$\text{MI-Zero}_{\text{Surv}}$ †	0.5541 (± 0.034)	0.5788 (± 0.028)	0.3842 (± 0.063)	0.5209 (± 0.049)	0.6623 (± 0.059)	0.5400	25.63	0
VL	$\text{ABMIL}_{\text{Prompt}}$	0.5717 (± 0.035)	0.6215 (± 0.084)	0.7825 (± 0.020)	0.5984 (± 0.052)	0.6762 (± 0.063)	0.6500	25.68	4
VL	CoOp	0.5971 (± 0.033)	0.5994 (± 0.086)	0.7853 (± 0.015)	0.5750 (± 0.064)	0.6840 (± 0.070)	0.6482	28.70	5
VL	OrdinalCLIP	0.6037 (± 0.043)	0.6202 (± 0.046)	0.7893 (± 0.018)	0.6053 (± 0.065)	0.6836 (± 0.036)	0.6604	28.01	5
VL	**VLSA (ours)**	**0.6176** (± 0.025)	**0.6652** (± 0.057)	**0.8002** (± 0.010)	**0.6370** (± 0.027)	**0.7571** (± 0.045)	**0.6954**	**25.15**	5

Method	BLCA	BRCA	GBMLGG	LUAD	UCEC	CI	MAE	D-cal Count
Attention	0.6083 (± 0.047)	0.6180 (± 0.046)	0.7908 (± 0.017)	0.6048 (± 0.063)	0.6908 (± 0.035)	0.6625	26.78	5
Learnable prototypes	0.5872 (± 0.048)	0.6201 (± 0.050)	0.7853 (± 0.013)	0.6061 (± 0.053)	0.6845 (± 0.052)	0.6566	26.76	4
Prognostic texts	0.6159 (± 0.025)	0.6614 (± 0.047)	0.7985 (± 0.009)	0.6314 (± 0.028)	0.7491 (± 0.049)	0.6912	**25.05**	4
Prognostic texts + FT	**0.6176** (± 0.025)	**0.6652** (± 0.057)	**0.8002** (± 0.010)	**0.6370** (± 0.027)	**0.7571** (± 0.045)	**0.6954**	25.15	5

■ **Visualization results**: ordinality & prediction interpretation