GECKO: Gigapixel Vision-Concept Contrastive Pretraining in Histopathology

{Saarthak.Kapse; Prateek.Prasanna}@stonybrook.edu

1 Stony Brook University
2 UNIVERSITY OF NORTH CAROLINA CHARLOTTE
ImagineLab

Saarthak Kapse[1]* Pushpak Pati* Srikar Yellapragada[1] Srijan Das[2] Rajarsi R. Gupta[1]
Joel Saltz[1] Dimitris Samaras[1] Prateek Prasanna[1]

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Impact

We introduce a novel pretraining method for Multiple Instance learning (MIL). Applications: Histopathology, Video analysis.

- **Problem**: Need for acquiring paired modality (for e.g. text, gene expression) for effective MIL pretraining.
- **Solution**: Computationally derive a concept-based modality!

Background

- Giga-Whole Slide Image (WSI) cohorts yield millions of patches, allowing large-scale Patch Foundation Model pretraining.

$\mathbb{R}^{N \times D}$
f_1
f_2
f_N
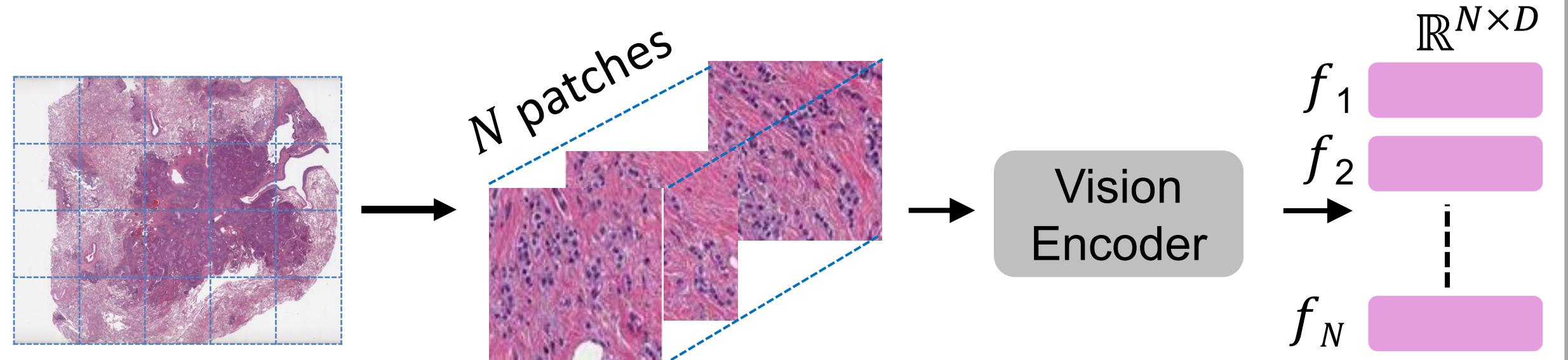
- Supervised MIL for slide-level tasks is prone to overfitting.

Multimodal Pretraining

TANGLE[2]

✓ Performance
✗ Cost-effective

Gigapixel Vision-Concept Pretraining

VLM[3]

Concept Prompts → Concept Prior
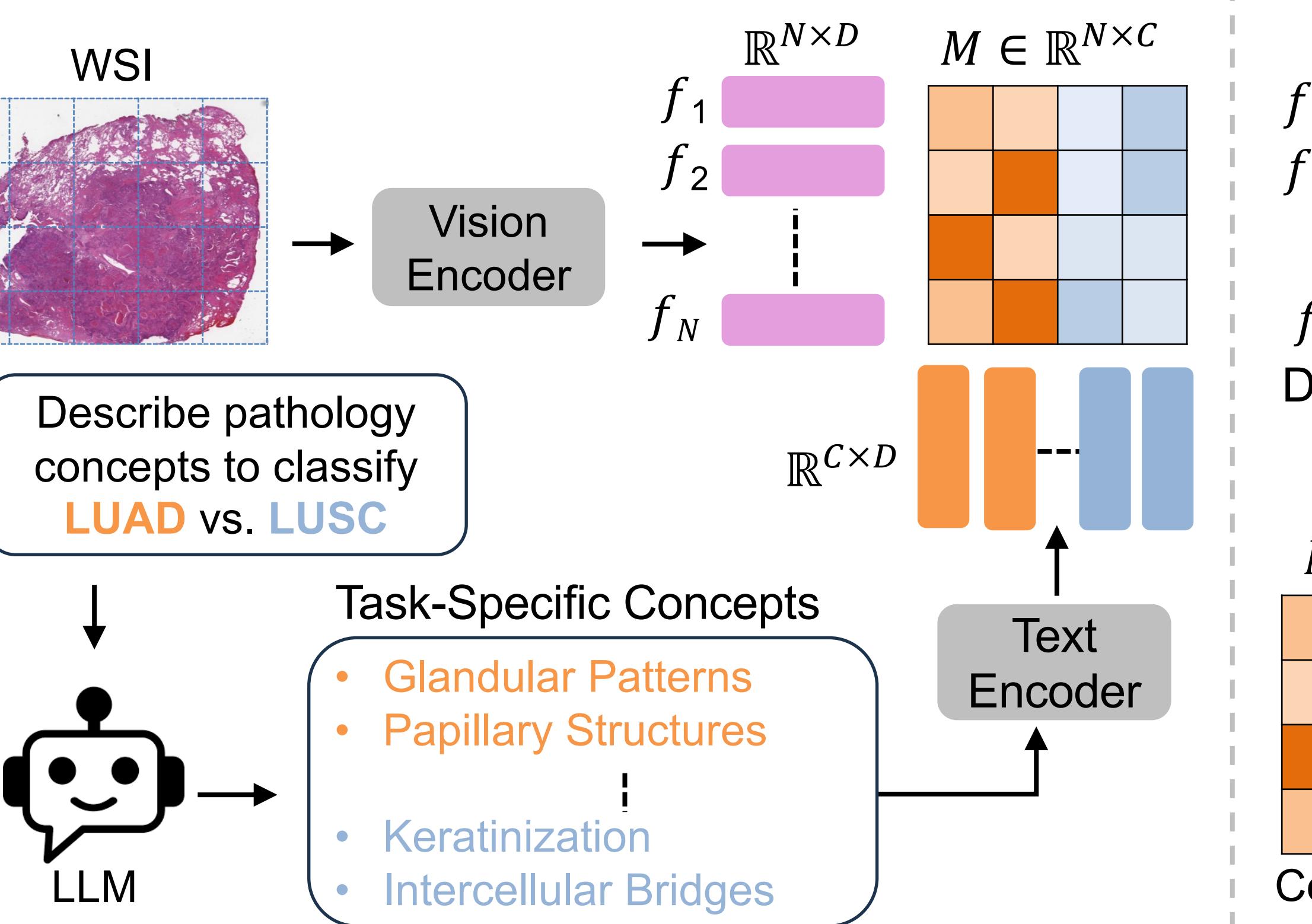
✓ Performance
✓ Cost-effective
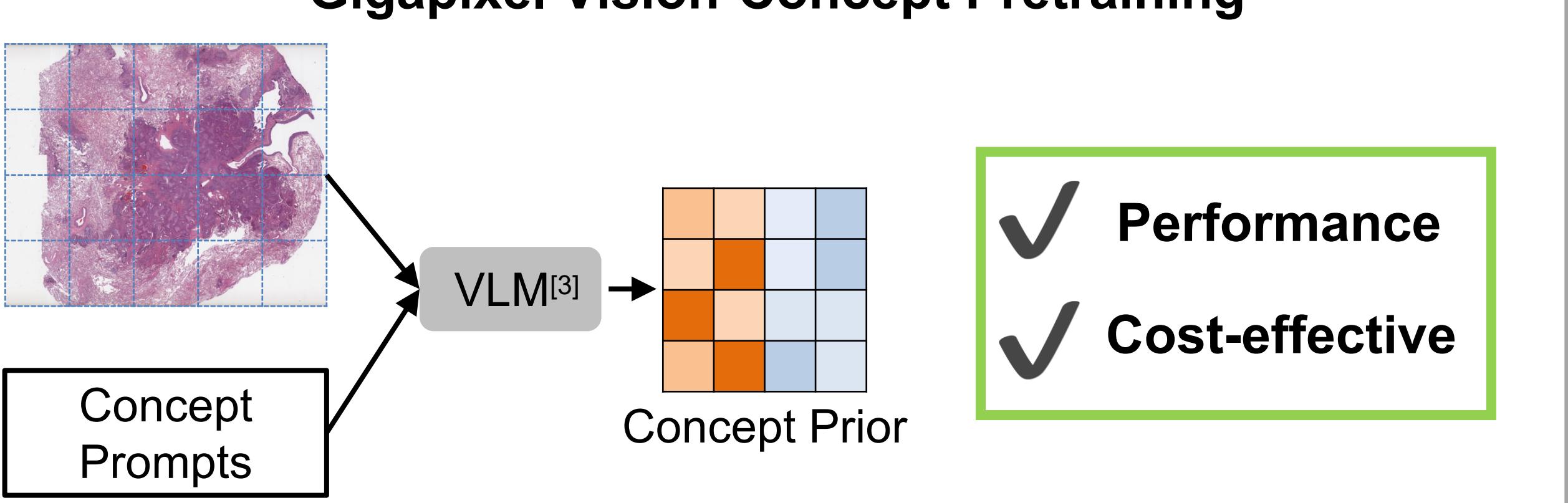
References:
[1] Kapse, Saarthak, et al. "Si-mil: Taming deep mil for self-interpretability in gigapixel histopathology." CVPR 2024
[2] Jaume, Guillaume, et al. "Transcriptomics-guided slide representation learning in computational pathology." CVPR 2024.
[3] Lu, Ming Y., et al. "A visual-language foundation model for computational pathology." Nature medicine 30.3 (2024): 863-874.
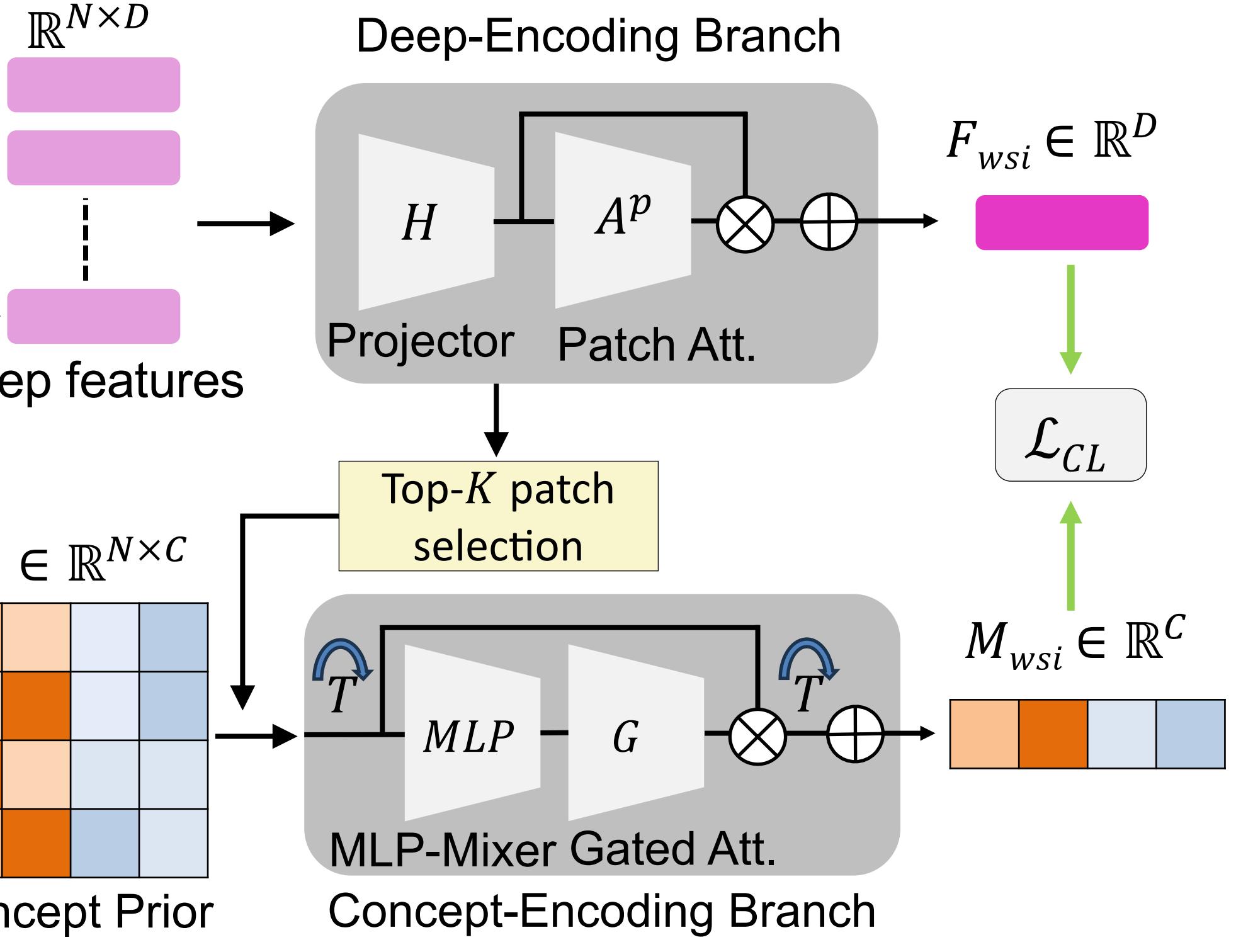
Method

a) Data curation: Concept Prior is computationally derived from WSI (no additional modality).

WSI → Vision Encoder → $f_1, f_2, \dots f_N$ → $\mathbb{R}^{N \times D}$; $M \in \mathbb{R}^{N \times C}$

Describe pathology concepts to classify **LUAD** vs. **LUSC** → LLM → Task-Specific Concepts:
- Glandular Patterns
- Papillary Structures
- Keratinization
- Intercellular Bridges
→ Text Encoder → $\mathbb{R}^{C \times D}$

b) Pretraining: Contrastively aligns the branches of dual-branch MIL (SI-MIL[1]).

Deep-Encoding Branch
$\mathbb{R}^{N \times D}$: $f_1, f_2, \dots f_N$ Deep features → H Projector, A^p Patch Att. $\otimes \oplus$ → $F_{wsi} \in \mathbb{R}^D$

Top-K patch selection

$M \in \mathbb{R}^{N \times C}$ Concept Prior → T MLP G $\otimes$ $T \oplus$ MLP-Mixer Gated Att. Concept-Encoding Branch → $M_{wsi} \in \mathbb{R}^C$

$\mathcal{L}_{CL}$

c) Supervised: Linear Probing

Deep features $f_1, f_2, \dots f_N$ → GECKO's Dual-branch MIL 🔒 → F_{wsi} GECKO_deep → Probing

Concept Prior → M_{wsi} GECKO_concept → Probing

GECKO_ensemble
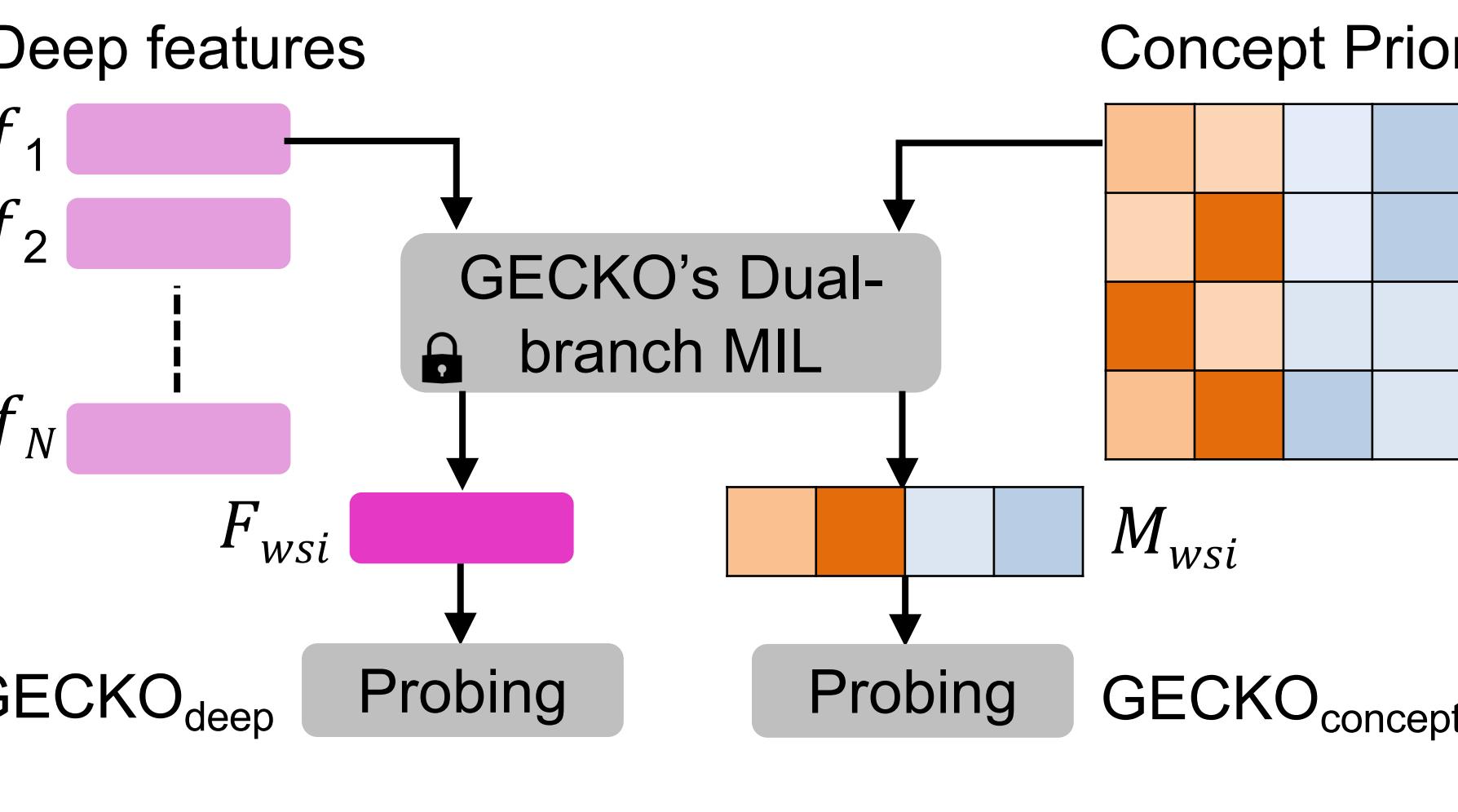
d) Unsupervised: Pathologist driven heuristic

M_{wsi} → Split Per-Class → $\arg\max_i \left(\sum \square, \sum \square \right)$
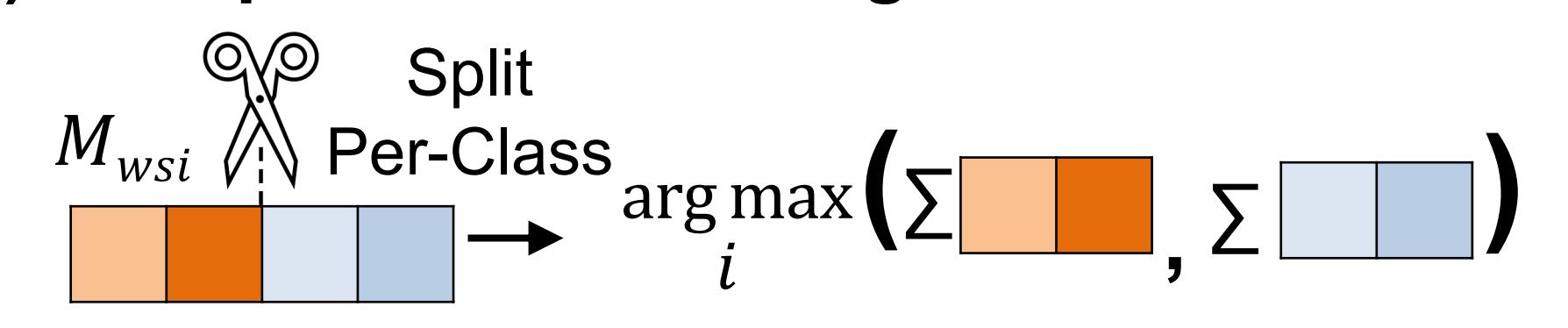
Results and Interpretability Analysis

Takeaway: The Concept Prior, a stronger FM, and gene expression, each significantly improves performance.

	Methods	Embedding	Interpretable (patch level-feature level)	EBV+MSI vs. Others (70 vs. 199)	MSI vs. Others (44 vs. 225)
WSI only	Intra	CONCH	✓- ×	83.5 ± 8.3	83.9 ± 7.0
	GECKO	deep	✓- ×	85.3 ± 8.8	83.8 ± 4.8
		concept	✓- ✓	82.8 ± 6.1	84.4 ± 10.9
		ensemble	✓- ×	**86.4 ± 8.4**	**86.5 ± 7.8**
WSI + Gene	TANGLE	CONCH	✓- ×	85.4 ± 8.0	86.6 ± 5.0
	GECKO	deep	✓- ×	86.1 ± 7.2	88.1 ± 4.3
	(CONCH)	concept	✓- ✓	85.6 ± 6.3	86.6 ± 6.5
		ensemble	✓- ×	**87.1 ± 7.0**	**89.4 ± 5.3**

	Methods	Embedding	EBV+MSI vs. Others $k = 10$	$k = 25$
WSI only	Intra	CONCH	75.6 ± 5.3	79.4 ± 7.3
	GECKO	CONCH	79.8 ± 4.8	82.5 ± 7.4
		CONCHv1.5	**82.1 ± 4.7**	**84.6 ± 5.5**
MM	TITAN	CONCHv1.5	78.7 ± 3.6	84.7 ± 4.6
	GECKO	CONCH	81.9 ± 4.8	84.5 ± 6.2
		CONCHv1.5	**84.4 ± 5.4**	**86.0 ± 5.8**

MM – Multi-Modal. Ensemble for eval.

Full supervision setting

- GECKO outperforms unimodal pretraining.
- GECKO's 3-way loss (WSI-Gene-Concept) outperforms TANGLE's 2-way loss (WSI-Gene).
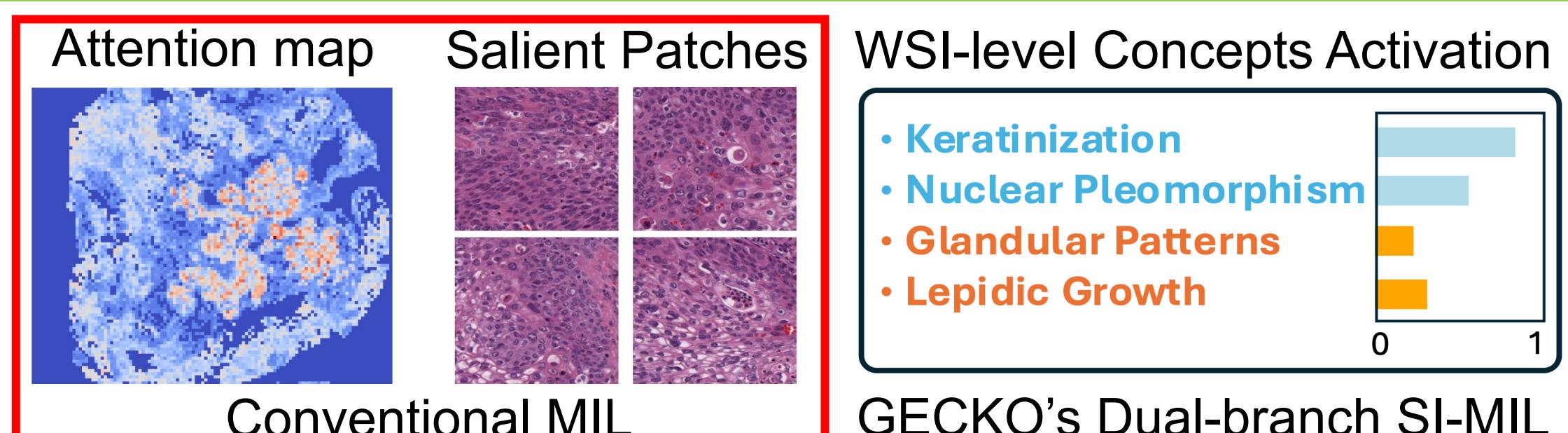
Ablations (few-labels setting)

- GECKO is agnostic to choice of FM.
- GECKO is on-par or outperforms TITAN, despite TITAN's extensive pretraining on 100K+ WSI-Report pairs.

Takeaway: Beyond identifying key regions in a WSI, GECKO also provides salient concepts driving the prediction.

Slide-Level Interpretability

Attention map | Salient Patches | WSI-level Concepts Activation
Conventional MIL

- Keratinization
- Nuclear Pleomorphism
- Glandular Patterns
- Lepidic Growth

0 ... 1

GECKO's Dual-branch SI-MIL

Feature-Level Interpretability

Handcrafted features
➤ Skewness of cells' intensity std
➤ Graph modularity with cell types as community

GECKO's Concepts
➤ Alveolar growth pattern
➤ Keratinization and Intercellular Bridges

✗ **Pathologist-friendly?** ✓