RESEARCH PRESENTATION | NeurIPS 2026

Concept-Aware Wasserstein Routing for Few-Shot WSI Classification

Ankit Kumar

C-MInDS, IIT Bombay | MeDAL Lab

Centre for Machine Intelligence & Data Science

NEURAL INFORMATION PROCESSING SYSTEMS

MeDAL

Introduction

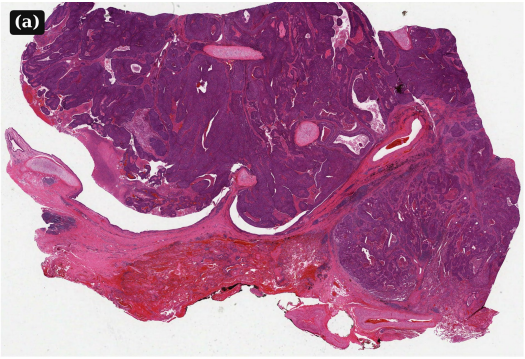
Research question

How can pathology concepts and optimal transport improve classification from only a few labelled whole-slide images?

Task: predict the cancer subtype of a whole-slide image (WSI) from only a few labelled slides per class.

- A WSI is cut into thousands of patches; only the slide has a label
- Pathology vision–language models (PLIP, CONCH) link patches to text

Why WSI classification is hard



TCGA lung squamous carcinoma slide (from the paper).

- **Huge image, one label**
- **Few examples:** 4, 8 or 16 slides per class
- **Scattered evidence:** small patterns in different regions

Existing methods

- **MIL methods** (ABMIL, TransMIL, . . .) treat patches as independent.
- **Vision–language MIL** (TOP, ViLa-MIL, MAPLE, . . .) use text only for the final decision.
- **One prototype per class** mixes different tissue patterns into one point.

Our idea

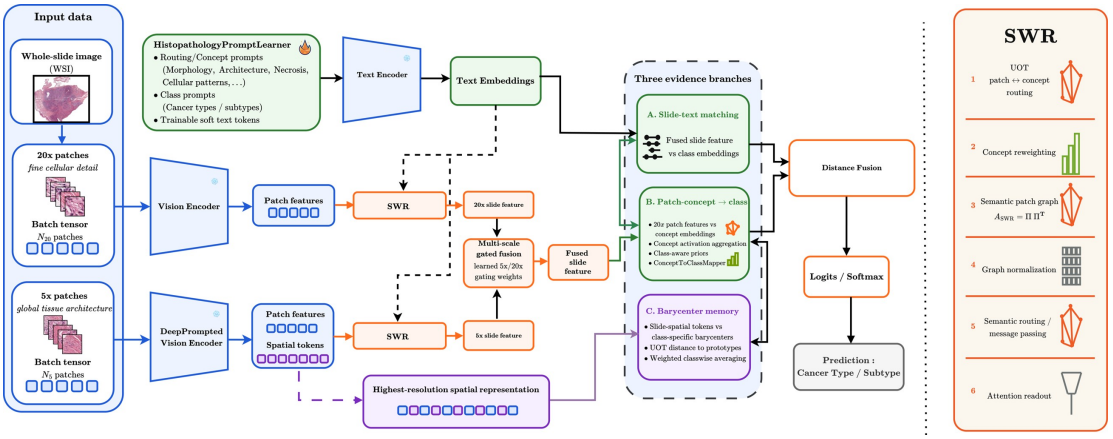
Let concepts decide which patches interact, and use several prototypes per class.

Contributions

1. **Semantic Wasserstein Routing (SWR).** Optimal transport assigns patches to pathology concepts; patches with the same concept share information.
2. **Barycentric prototype memory.** Each class is described by several typical tissue patterns instead of one average.
3. **Concept-guided reasoning.** Fuse fine-resolution patch evidence, pathology concepts, and class-level distance signals.

Only about 1.5M parameters are trained; the encoders stay frozen.

Architecture



Architecture figure from the paper.

How the architecture works

1. **Encode** patches at 5× (tissue) and 20× (cells)
2. **Prompt** concepts and class names into text embeddings
3. **Route** patches to concepts; patches sharing a concept exchange information
4. **Compare** with class text, class prototypes and concept scores
5. **Fuse** the three into one score per class

Experimental setup

- **Datasets:** TCGA breast (BRCA), kidney (RCC), lung (NSCLC)
- **Shots:** 4, 8, 16 slides per class; five folds
- **Metrics:** AUC, F1, accuracy
- **Encoders:** CONCH (main results) and PLIP
- **Baselines:** ABMIL, TransMIL, GTMIL, WiKG, TOP, ViLa-MIL, MSCPT, FOCUS, MAPLE

Results table

Dataset	Shots	MAPLE (strongest baseline) AUC	F1	ACC	**Ours** AUC	F1	ACC
RCC	4	.947	.791	.805	**.955**	**.802**	**.817**
	8	.971	.888	**.899**	**.975**	**.895**	.897
	16	.984	.910	.919	**.985**	**.915**	**.929**
NSCLC	4	.889	.774	.786	**.902**	**.802**	**.812**
	8	.964	.894	.894	**.972**	**.915**	**.908**
	16	.981	.914	.914	**.985**	**.917**	**.918**
BRCA	4	.844	.653	.695	**.862**	**.675**	**.720**
	8	.900	.748	.797	**.918**	**.770**	**.820**
	16	.916	.790	.825	**.930**	**.808**	**.845**

CONCH encoder, mean over five folds (from the paper). Bold = better of the two.

Ablation tables

Component (TCGA-RCC)	ACC (%)
Full model	**91.7**
without prototype memory	86.5
without SWR	88.2
only 5× patches	85.4
only 20× patches	87.2

Transport (TCGA-RCC)	ACC (%)	AUC
Unbalanced OT	**91.7**	**0.986**
Balanced OT	89.0	0.956
Cosine only (no OT)	85.4	0.918

From the paper. Left: single run. Right: mean over 3–5 seeds.

Conclusion and limitations

Conclusion

Concept-guided routing and several prototypes per class give consistent gains in few-shot WSI classification.

Limitations

- About 2× slower than MAPLE
- Concept weights had no measurable effect
- Three TCGA tasks only

Thank you!

Questions and discussion

Concept-Aware Wasserstein Routing for Few-Shot WSI Classification





Centre for
Machine
Intelligence &
Data Science



NEURAL INFORMATION
PROCESSING SYSTEMS



MeDAL