ALL-IN: A Local GLobal Graph-based DIstillatioN Model for Representation Learning of Gigapixel Histopathology Images With Application In Cancer Risk Assessment

Puria Azadi[1], Jonathan Suderman[3], Ramin Ebrahim Nakhli[2], Katherine Rich[2], Maryam Asadi[2], Sonia Kung[3], Htoo Zarni Oo[3], Mira Keyes[2], Hossein Farahani[2], Calum MacAulay[4], Larry Goldenberg[3], Peter Black[3], Ali Bashashati[2]

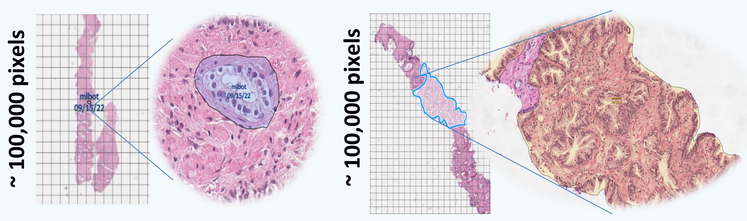
[1]Department of Electrical and Computer Engineering - UBC, [2]School of Biomedical Engineering - UBC, [3]*Vancouver Prostate Centre,* [4]BC Cancer Research Center

Risk Assessment:

- The holy grail application of machine learning in histopathology is providing **predictive** and **prognostic** information.
- Risk stratification is **under-explored** compared to grading and subtype classification.

Challenges:

- Aggregating **global** and **local** features to link histopathology images and outcome.
- Modeling **interactions at different scales** as well as **tissue heterogeneity**
- **Graph Neural Networks** can model patch relations; however, most GNN-based risk prediction have **limited nodes' receptive fields**.

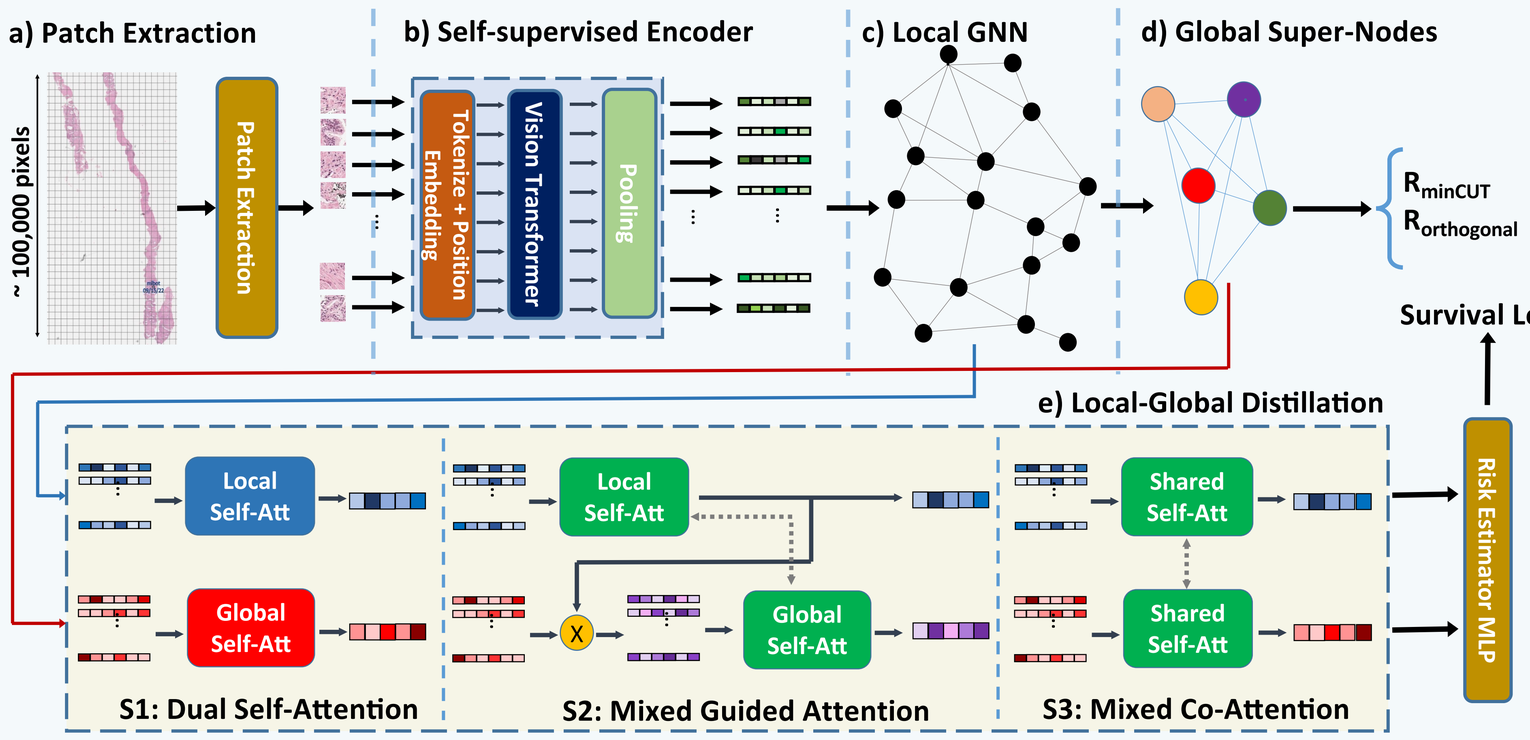


Sample of fine (left) and coarse (right) morphological feature

Contributions:

1. Proposing **ALL-IN: A Local GLobal Graph-based DIstillatioN**
2. ALL-IN is a novel graph-based model that extracts both local and global properties by identifying **morphological super-nodes.**
3. Introducing **a fine-coarse feature distillation module** with 3 various strategies to aggregate interactions at different scales.
4. Conducting an extensive study on the performance of our proposed model using two datasets.
5. Evaluating in two scenarios: **risk prediction** and **patient stratification**
6. Our results demonstrates **statistically significant stratification into risk groups** with clinical utility.

Methodology:



The overview of our proposed method. Steps are as follows:

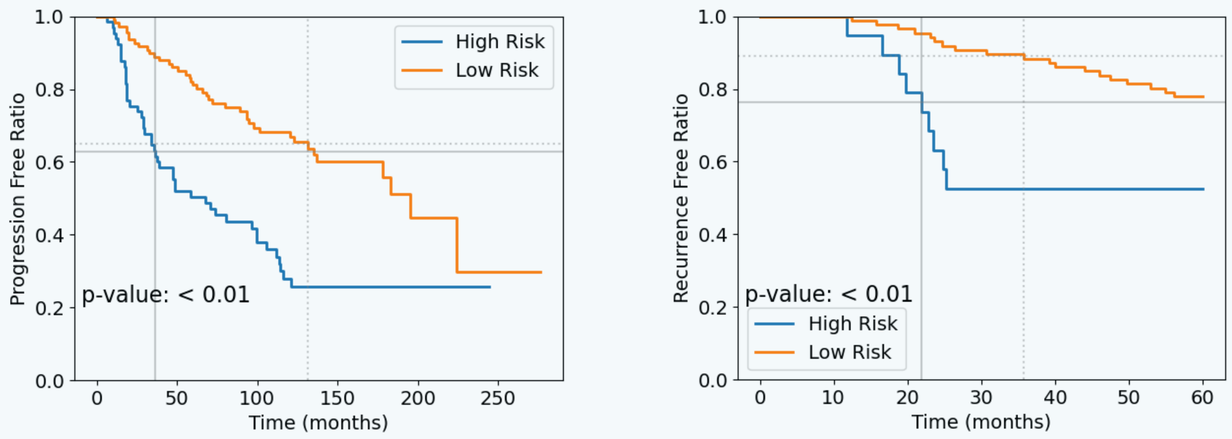
a) The input slide is tiled into **non-overlapping patches**.
b) The patches are fed into a self-supervised encoder to **extract embeddings**.
c) A graph is constructed, and the new local instance-level embeddings are obtained through the **message-passing** process.
d) The **global context** representations in the form of **super-nodes** are extracted utilizing two unsupervised loss functions (R_{minCUT} , $R_{orthognal}$).
e) The fine and coarse feature vectors are aggregated in the **distillation module** to extract representations for **both local and contextual** morphological features.
f) A multilayer Perceptron is deployed to estimate the risk using final resultant vectors.

Hazard Prediction and Patient Stratification Results:

Model	C-index		p-values ↓		Low ↓ - High↑ Median Time		#Parameters
	PCa-AS	PCa-BT	PCa-AS	Pca-BT	PCa-AS	PCa-BT	
Deep Set	0.495 ± 0.017	0.50 ± 0.0	0.837	0.912	67.78 - 71.87	24.62 - 24.89	329K
AMIL	0.544 ± 0.06	0.533 ± 0.060	0.820	0.148	48.99 - 89.10	21.86 - 30.71	592K
DGC	0.522 ± 0.113	0.572 ± 0.150	0.494	0.223	47.61 - 96.66	23.44 - 24.85	626K
Patch-GCN	0.555 ± 0.059	0.541 ± 0.118	0.630	0.981	37.72 - 94.95	23.05 - 25.25	1,302K
ALL-IN + DA (ours)	0.631 ± 0.058	0.596 ± 0.062	< 0.01	< 0.01	37.72 - 115.91	21.86 - 35.77	850K
ALL-IN + MGA (ours)	0.632 ± 0.060	0.589 ± 0.074	< 0.01	< 0.01	47.61 - 101.39	21.86 - 35.77	653K
ALL-IN + MCA (ours)	**0.639 ± 0.048**	**0.600 ± 0.077**	**< 0.01**	**< 0.01**	**36.5 - 131.71**	**21.86 - 35.77**	**653K**

Comparison of our method against baselines and ablation study on policies

Kaplan-Meier Curves:



PCa-AS **PCa-BT**

Ablation Study:

Modules				C-index	
Model	Local-node	Self-supervised encoder	Super node + Distillation	PCa-AS	PCa-BT
Ours	✓	✗	✗	0.584 ± 0.072	0.550 ± 0.109
	✓	✓	✗	0.622 ± 0.055 ↑	0.597 ± 0.045 ↑
	✓	✓	✓	0.639 ± 0.048 ↑↑	0.600 ± 0.077 ↑↑

Summary:

- We introduced a novel model, which utilizes interactions at **different scales** for improved risk stratification.
- Our results suggest that the proposed model is capable of separating patients into statistically significant risk groups with **actionable clinical utility**.