Dynamically Ranking Urothelial Cells by Malignancy Using Multiple Instance Learning and Attention

Tanay Panja, Joshua Levy, Louis Vaickus,
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

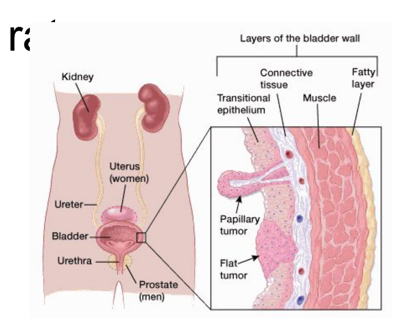
ABSTRACT

Urothelial carcinoma's (UC) heterogeneity complicates diagnosis and treatment. Used AutoParis-X features, including cell morphology and deep-learning features. **Multiple-instance learning** was applied with slides as bags and cells as instances. An attention mechanism ranked cells by malignancy relevance. **Results for the Attention Model:** 79% accuracy, 0.76 AUROC.Results for the Baseline Model: 67% accuracy, 0.66 AUROC.
The Attention Model outperformed the Baseline Model, showing promise for improving UC diagnostics.

INTRODUCTION

Urothelial carcinoma such as Bladder Cancer is the 5th most common non-cutaneous cancer in the U.S., with a 31% recurrence rate. Current cytology is cost-effective but suffers from:

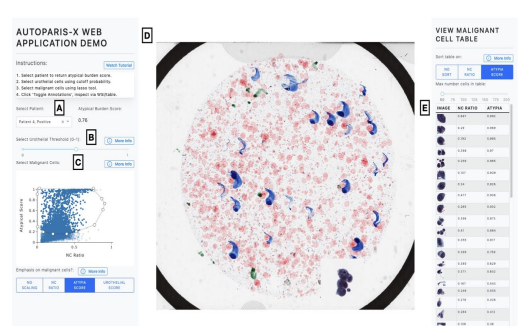
- **Labor-Intensive**: Time-consuming and error-prone.
- **Screening Bias**: Variability and missed patterns.
- **Limited Scalability**: Struggles with high test volumes.



Automating cytology could improve efficiency, accuracy, and consistency while reducing costs and detecting subtle patterns.

AutoParis-X

- Software application with several deep learning models to analyze urine cytology
- Extracts cell and cluster-level features including NC ratio, atypia score, and morphological measures
- ***Does not rank relative malignancy of cells***



Current Approaches

- RankNet, introduced by Burges et al. in 2005, employs neural networks to learn a ranking function by comparing pairs of examples, which can result in computational inefficiency, particularly when working with large datasets.
- Sanghvi et al. developed a semi-autonomous diagnostic decision aid for bladder cancer using deep learning to rank cells based on their likelihood of malignancy, but it faces challenges such as reliance on high-quality labeled data and limited interpretability of the model's predictions.
- Butke et al. proposed an end-to-end multiple instance learning approach for whole-slide cytopathology of urothelial carcinoma, which learns to rank regions of interest within a slide. However, this approach faces challenges such as increased complexity in model training and potential issues with false positives or negatives.

What is Multiple Instance Learning (MIL)?

MIL is a machine learning approach where only bag-level labels are known, not individual instances within the bags.

Why is MIL Useful?

1. **Handles Incomplete Labeling:** Useful for datasets with only overall labels, such as medical images where individual labels are impractical.
2. **Identifies Relevant Instances:** Focuses on important regions within bags, improving detection accuracy, like spotting malignancy in pathology slides.

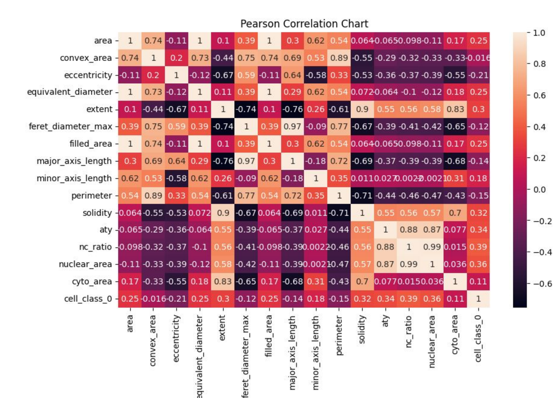
METHODS

Data Collection

- **Slide Acquisition:** Dataset of cytology slides from DH
- **Cell Extraction:** AutoParis-X extracted features
  - Morphological features (e.g., cell shape, size, nuclear morphology).
  - Deep learning-extracted features (e.g., atypia score, NC ratio)
  - Limited to 3000 cells per slide

Feature Selection

- Features were selected based on Pearson correlation and literature review
- Features involving NC ratio and cytoplasmic area showed greater correlation with slide malignancy classification
- Features that shared high correlations had only one feature selected for dimensionality reduction



Multiple Instance Learning (MIL) Framework

- **Bags:** Cytology slides
- **Instances:** Urothelial cells
  - **Malignant slides:** Have at least one malignant cell
  - **Benign slides:** Have no malignant cells

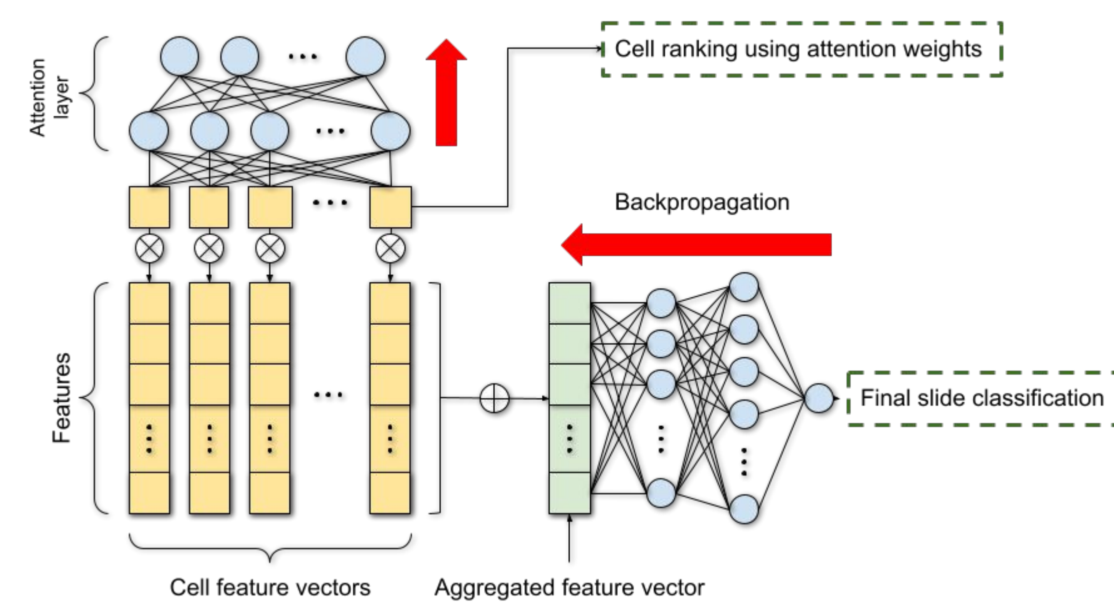


Figure 1: Slide-level Attention Framework

RESULTS

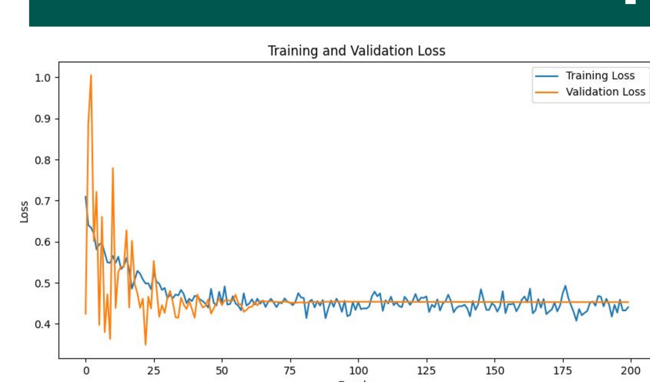


Figure 2: Loss Graph

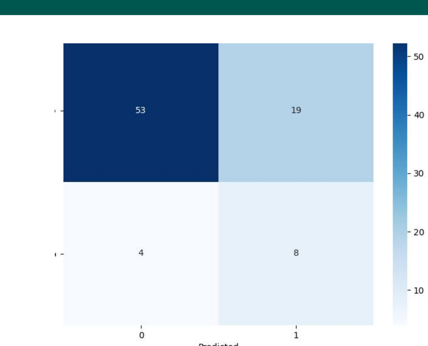


Figure 3: Confusion Matrix

	Accuracy	AUROC	F1
Attention Model	**0.79**	**0.76**	0.36
Baseline Model	0.67	0.66	**0.63**

Table 1:Ablation Study Results

Model Evaluation

Challenge: No predetermined rankings of cells on the slide.
Solution: Our model ranks cells without supervised guidance.
Evaluation:
- **Accuracy:** Measures how well the model's ranking matches the slide's classification.
- **AUROC:** Assesses the model's ability to distinguish between different classes.
- **F1 Score:** Evaluates the balance between precision and recall in the ranking.

Objective: Validate the model's effectiveness by comparing its rankings to the known slide classifications.

DISCUSSION

Model Performance and Loss Analysis

- Training loss decreased and stabilized, showing effective learning.
- Validation loss had initial variability but eventually converged, suggesting good generalization.
- Model with attention outperformed baseline model in accuracy and AUROC
- Loss graph indicates some instance of overfitting due to stagnation after 50 epochs

Confusion Matrix Insights

- Model shows sensitivity to overlapping features between benign and malignant cells.
- Confusion matrix results indicates the model's main classification weakness lies in the false positives category

Ablation Study Results

- Attention model performed better than baseline in accuracy and AUROC.
- Lower F1 score reflects a trade-off between precision and recall.

FUTURE WORK

- Reduce number of false positives by determining a higher threshold
- Develop saliency map to understand which regions of cells are more responsible for slide classifications
- Experiment with different learning weights and a weight scheduler
- Integrate project into web app

CONCLUSION

Enhanced Diagnostic Accuracy

- The integration of the attention mechanism within our multiple instance learning framework substantially improved the model's diagnostic accuracy and AUROC.
- The attention mechanism effectively highlighted the most relevant cells, refining the model's ability to distinguish between benign and malignant urothelial cells.

Precise Cell Ranking

- The model's capacity to dynamically rank cells based on their malignancy is a significant advancement, offering a more nuanced evaluation compared to traditional methods.
- This approach enables more targeted and accurate assessments, contributing to better diagnostic and treatment strategies for urothelial carcinoma.

REFERENCES : American Cancer Society. (2017). Bladder cancer. In Cancer Facts & Figures 2017 (pp. 1-72). American Cancer Society. Retrieved from https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html

1. Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians, 67*(1), 7-30. https://doi.org/10.3322/caac.21387
2. Babjuk, M., Burger, M., Zigeuner, R., Shariat, S. F., van Rhijn, B. W., Comperat, E., ... & Sylvester, R. J. (2013). EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: Update 2013. *European Urology, 64*(4), 639-653. https://doi.org/10.1016/j.eururo.2013.06.003
3. Murphy, W. M., Soloway, M. S., & Jukkola, A. F. (1984). Urinary cytology and bladder cancer. The cellular features of transitional cell neoplasms. *Human Pathology, 15*(7), 675-684. https://doi.org/10.1016/S0046-8177(84)80003-9
4. Raab, S. S., & Grzybicki, D. M. (2009). Urine cytology and the detection of urothelial neoplasms: A review of the recent literature. *Cancer Cytopathology, 117*(3), 175-181. https://doi.org/10.1002/cncy.20029
5. Compérat, E. M., Burger, M., Gontero, P., Mostafid, A. H., Palou, J., Rouprêt, M., van Rhijn, B. W. G., Shariat, S. F., Sylvester, R. J., Zigeuner, R., & Babjuk, M. (2018). Grading of urothelial carcinoma and the new "World Health Organisation Classification of Tumours of the Urinary System and Male Genital Organs 2016". *European Urology, 73*(2), 240-246. https://doi.org/10.1016/j.eururo.2016.05.041
6. McCroskey, Z., Pambuccian, S. E., Kleitherms, S., Antic, T., Cohen, M. B., & Barkan, G. A. (2015). Accuracy and interobserver variability of the cytologic diagnosis of low-grade urothelial carcinoma in instrumented urinary tract cytology specimens. *American Journal of Clinical Pathology, 144*(6), 902-908. https://doi.org/10.1309/AJCPE1O9YKMRSQKG
7. Burges, C. J. C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., & Hullender, G. (2005). Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 89-96). ACM. https://doi.org/10.1145/1102351.1102363
8. Sanghvi, A. B., Allen, E. Z., Callenberg, K. M., & et al. (2019). Performance of an artificial intelligence algorithm for reporting urine cytopathology. *Cancer Cytopathology, 127*(10), 658-666. https://doi.org/10.1002/cncy.22162
9. Butke, J., et al. (2021). End-to-end multiple instance learning for whole-slide cytopathology of urothelial carcinoma. In *Proceedings of the MICCAI Workshop on Computational Pathology*. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-87237-3_11
10. Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Muti, H. S., ... & Zöllner, F. G. (2022). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer, 3*(4), 395-403. https://doi.org/10.1038/s43018-022-00334-5
11. Sirinukunwattana, K., Raza, S. E. A., Tsang, Y. W., Snead, D. R., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging, 35*(5), 1196-1206. https://doi.org/10.1109/TMI.2016.2525803
12. Ilse, M., Tomczak, J. M., & Welling, M. (2018). Attention-based deep multiple instance learning. In Proceedings of the 35th International Conference on Machine Learning (Vol. 80, pp. 2127-2136). PMLR. Retrieved from https://proceedings.mlr.press/v80/ilse18a/ilse18a.pdf

edit.