A Weakly Supervised Deep Learning Strategy for Assisting Malignancy Diagnosis in Urinary Cytology

Yosep Chong[1] , Kwangil Yim[1], Kyung Jin Seo[1], **Hyeongyeol Lim[2]**, Hongjun Yoon[2], Daeky Jeong[2], Won June Cho[2], Mohammad Rizwan Alam[1], Gyungyub Gong[3], Nam Hoon Cho[4], Chong Woo Yoo[5*]

[1]Department of Hospital Pathology, The Catholic University of Korea, College of Medicine, Korea
[2]AI Research Lab, DEEPNOID Inc, Korea
[3]Department of Pathology, Asan Medical Center, Korea
[4]Department of Pathology, Yonsei University College of Medicine, Korea
[5]Department of Pathology, National Cancer Center, Korea

Abstract

Purpose: Urinary cytology is a primary non-invasive screening tool for bladder cancer, but its diagnostic sensitivity remains suboptimal with significant inter-observer variability. This study aimed to develop and validate a weakly supervised deep learning model using multiple instance learning (MIL) for detecting malignancy in urinary cytology whole-slide images (WSIs), eliminating the need for labor-intensive cell-level annotations.

Methods: A total of 3,042 urinary cytology WSIs were collected: 2,135 for training (1,884 benign, 251 malignant) from Uijeongbu St. Mary's Hospital and 907 for testing (600 benign from the same institution, 307 malignant from the Quality Assurance Program of the Korean Society of Cytopathology). The Clustering-constrained Attention MIL (CLAM) framework with a ResNet50 feature extractor was trained using slide-level labels only. Performance was evaluated using AUROC, accuracy, sensitivity, specificity, precision, and F1-score with 95% confidence intervals from bootstrap resampling.

Results: The MIL model achieved an AUROC of 0.941 (95% CI, 0.924-0.957), accuracy of 0.860 (95% CI, 0.833-0.888), sensitivity of 0.824 (95% CI, 0.779-0.868), and specificity of 0.901 (95% CI, 0.862-0.937) on the external test dataset. Attention heatmap visualization confirmed that the model focused on cytologically suspicious regions corresponding to malignant cell clusters, providing interpretable diagnostic rationale.

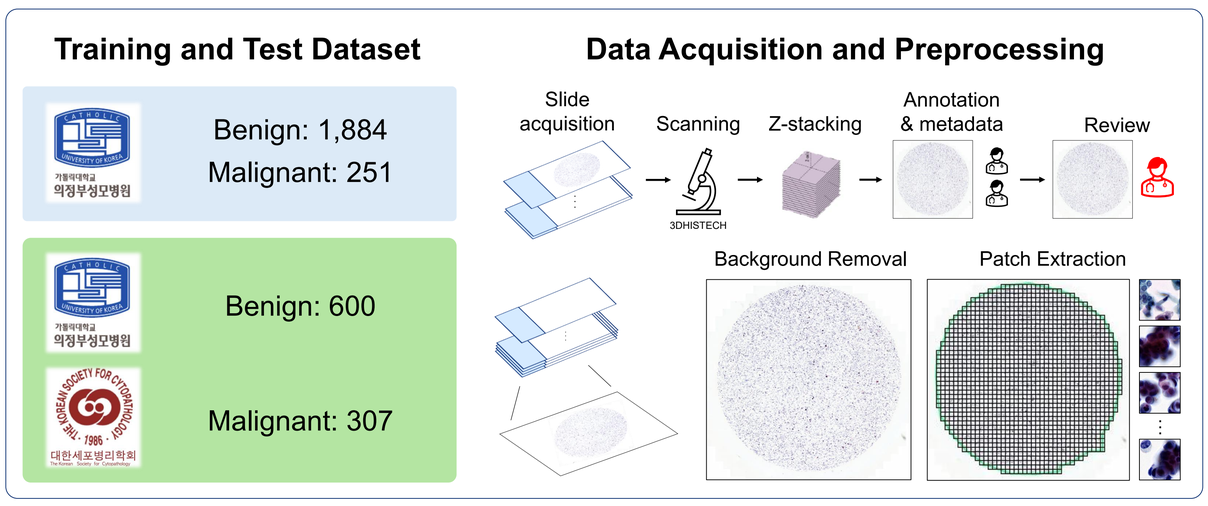
Conclusions: The weakly supervised MIL model demonstrated excellent diagnostic performance for malignancy detection in urinary cytology WSIs without requiring cell-level annotations. This approach offers a practical and scalable framework for AI-assisted cytopathology diagnosis, potentially improving screening efficiency and diagnostic reproducibility in clinical practice.

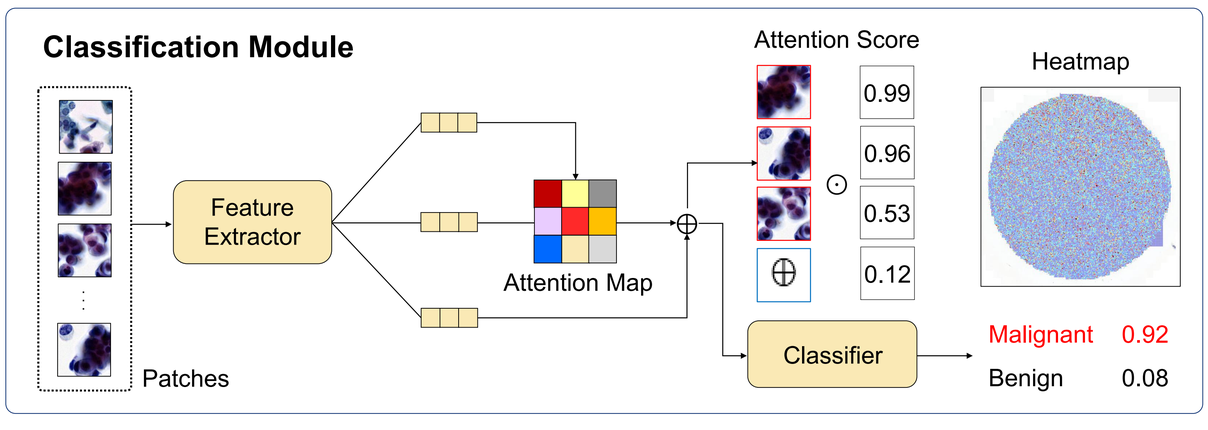
Introduction

Bladder cancer is one of the most common urinary tract malignancies, and urinary cytology remains a primary non-invasive method for its detection and surveillance. However, its diagnostic sensitivity is limited and interpretation is subject to considerable inter-observer variability. AI-based approaches have been applied to address these issues, but most rely on labor-intensive cell-level annotations and have been validated mainly on single-institution datasets.

Materials & Methods

- Urinary cytology whole-slide images were collected from Uijeongbu St. Mary's Hospital for training (UJB;1,884 benign, 251 malignant) and for testing combined with malignant cases from the Quality Assurance Program of the Korean Society of Cytopathology (QAP-KSC; 600 benign, 307 malignant). All slides were scanned with 3DHISTECH scanners and labeled at the slide level through double reading by three board-certified pathologists.
- Patches and features were extracted from each slide and trained using the clustering-constrained attention multiple instance learning (CLAM) method with a ResNet50 feature extractor and slide-level labels only.



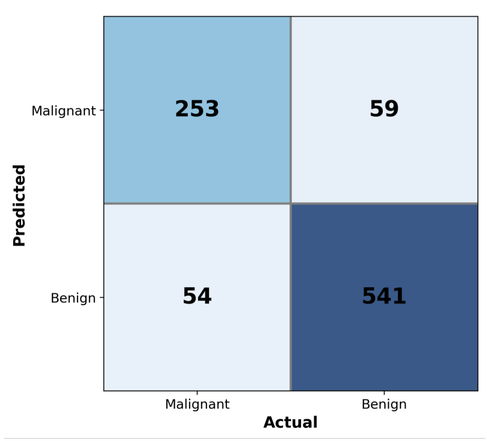


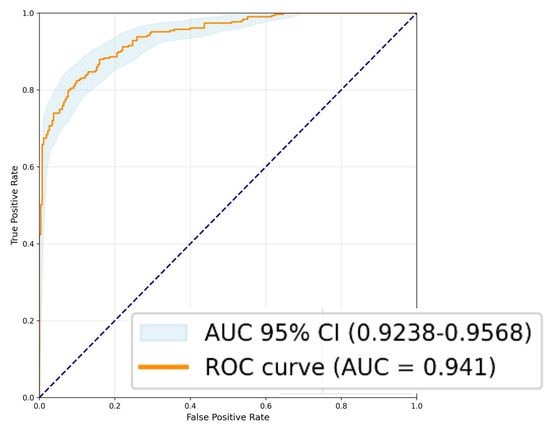
Dataset Characteristics Table: Training and Test Dataset

Characteristics	No. of cases (%)		Total
	Training Dataset	Test Dataset	
	UJB (N = 2,135)	UJB + QAP-KSC (N = 907)	
Age (Median)	15-96 (68.00)	13-97 (69.00)	13-97 (68.00)
Sex - Male	1,184 (55.46%)	487 (53.69%)	1,671 (54.93%)
Sex - Female	951 (44.54%)	420 (46.31%)	1,371 (45.07%)
	2,135	**907**	**3,042**
Class - Benign	1,884 (88.24%)	600 (66.15%)	2,484 (81.66%)
Class - Malignant	251 (11.76%)	307 (33.85%)	558 (18.34%)
	2,135	**907**	**3,042**

Results

Performance on Test Dataset (N=907)

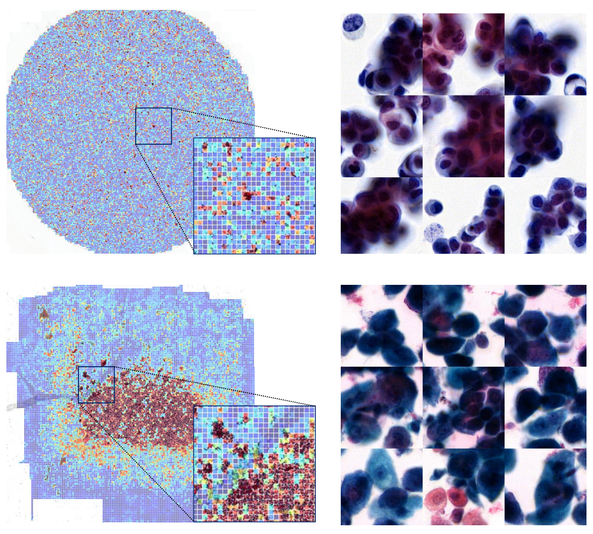




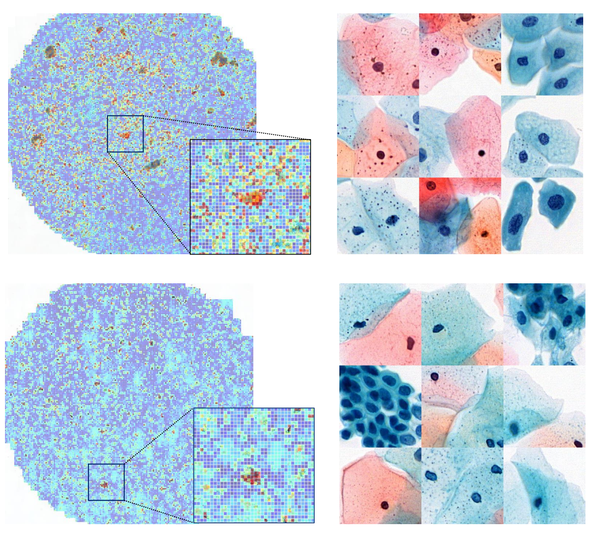
Metric	Value (95% CI)
Accuracy	0.8596 (0.8333 – 0.8877)
Precision	0.8693 (0.8297 – 0.9088)
Recall (Sensitivity)	0.8241 (0.7792 – 0.8679)
Specificity	0.9011 (0.8622 – 0.9366)
F1-score	0.8635 (0.8330 – 0.8919)
AUROC	0.9411 (0.9238 – 0.9568)

Attention Heatmap Visualization on Test Dataset

Class – Benign



Class – Malignant



Conclusion

- A weakly supervised MIL model was developed and externally validated for detecting malignancy in urinary cytology whole-slide images using slide-level labels only, without cell-level annotations.
- The model showed strong diagnostic performance on an external test dataset, and attention heatmap highlighted cytologically suspicious regions, offering interpretable rationale for its predictions.
- Broader validation across institutions with varied preparation protocols and scanners would further support its use, but these results suggest the approach is a practical and scalable option for AI-assisted urinary cytology in clinical practice.