Automated Bladder Cancer Screening with Deep Learning Algorithms

Joshua Wang, Siddhartha Vatsa

Emerging Diagnostic and Investigative Technologies , Department of Pathology , Dartmouth Hitchcock Medical Center

Introduction

220,000+
global deaths from bladder cancer every year

~70%
10-year recurrence rate for non-muscle-invasive disease

- **Urine cytology,** a standard screening method, is normally done manually: a pathologist inspects each cell under a microscope
  - Time consuming and subjective
- **Automating the screening process**, without giving up accuracy, is the problem this project addresses.

MANUAL VS. AUTOMATED SCREENING

Manual urine cytology
Time-consuming, cell-by-cell review by a cytopathologist

Automated urine cytology
Can improve screening consistency, efficiency, and accuracy

RESEARCH AIM

- Combine U-Net segmentation with SVM classification to automate and optimize urine cytology
- Extract quantitative features from each cell and investigate their predictive power
- Benchmark the proposed pipeline against a multi-Otsu thresholding baseline

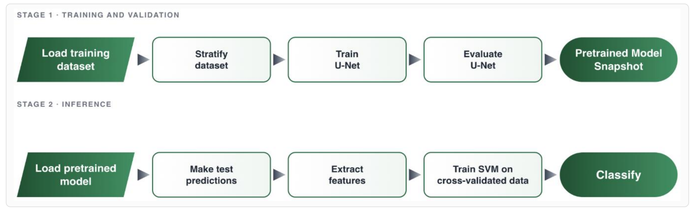
Why it matters

An automated pipeline for urine cytology screening can promote early detection of cancer, improving patient outcomes.

Methods

Data: 191 labeled training images, 100 test images with class labels; only the training set has ground-truth segmentation masks.
Baseline: multi-Otsu thresholding on percentile-normalized, grayscaled images.

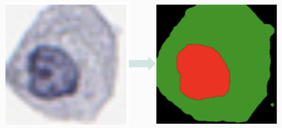
OUR U-NET + SVM PIPELINE



Training: the labeled set is split 80% train / 20% validation. The U-Net trains with online augmentation to increase data diversity, then is evaluated.

EXAMPLE SEGMENTATIONS

OUR U-NET PIPELINE



Raw image → U-Net predicted mask

MULTI-OTSU BASELINE

Ground truth vs. predicted mask

CELL FEATURES EXTRACTED

Nucleus area | Cytoplasm area | N/C ratio | Circularity | Solidity

Combined into one feature dataframe for classification

MODEL VALIDATION & CLASSIFICATION

5-fold cross-validation: the model trains on 4 folds and validates on the held-out fold, rotating until every fold has served as validation. The cross-validated features then train a SVM that sorts each cell into one of four diagnostic categories: negative, atypical, suspicious, or positive.

Why this design

Stratified splits, online augmentation, and cross-validation all target the same risk: overfitting to a small, blurry dataset.

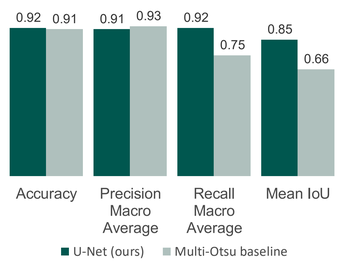
Results

+18%
higher mean IoU than the multi-Otsu baseline (0.846 vs. 0.662)

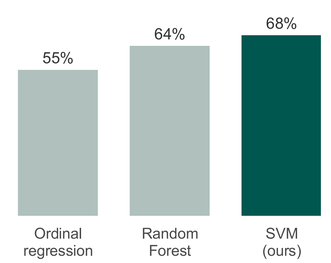
3×
nucleus recall vs. baseline (0.90 vs. 0.32)

68%
SVM classification accuracy vs. 50% baseline

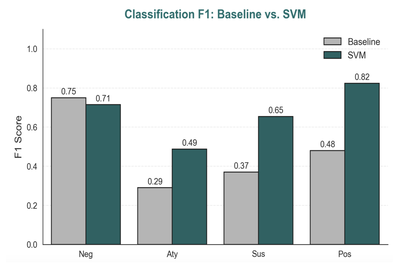
SEGMENTATION QUALITY



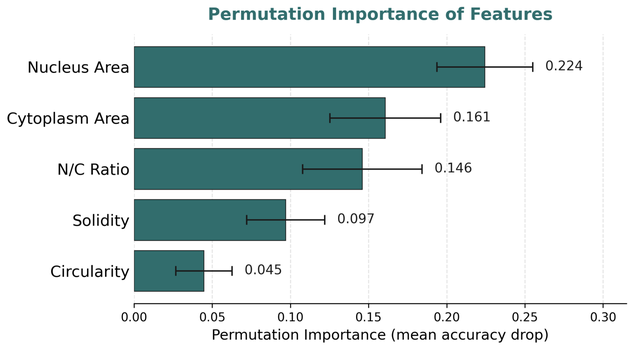
CLASSIFICATION ACCURACY BY MODEL



CLASSIFICATION F1 BY CATEGORY



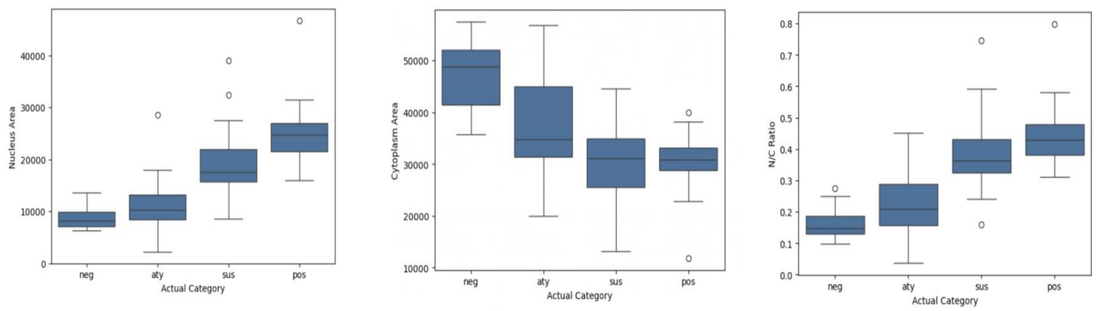
WHICH FEATURES PREDICT MALIGNANCY?



Permutation importance (left) shows how important each feature is to the model's classification accuracy.
Spearman correlation shows each feature's own relationship with actual category:

Nucleus area ρ = +0.595
N/C ratio ρ = +0.592
Cytoplasm area ρ = −0.434
Solidity ρ = −0.405
Circularity ρ = −0.381

CELL MORPHOLOGY SHIFTS WITH DIAGNOSTIC CATEGORY



Categories: neg = negative · aty = atypical · sus = suspicious · pos = positive

Conclusion

DL & ML algorithms are more accurate and consistently outperform traditional methods, offering greater accuracy and enhanced utility for urine cytology.

Limitations

Blurry cell images and a small sample size (n = 291 total) limit accuracy.

Future work

Collect more high-quality images, extract additional features without introducing noise, and build an end-to-end classifier to prevent error propagation between stages.