Cervical Cancer Screening: Improving Pap Smear Cell Classification

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

Dartmouth Health

ABSTRACT

- Cervical cancer is the fourth **most common cancer in women**, with ~350,000 deaths yearly; ~90% of deaths occur in underserved regions.
- **Early detection saves lives**, but access to pathologists is limited in many regions.
- This project explores **deep learning (ResNet-18)** with **data augmentation and preprocessing** to classify Pap smear cells.
- **Light and enhanced augmentation strategies** were tested and compared to the original model, evaluated by accuracy, F1, ROC-AUC, and interpretability
- Results show **enhanced augmentation improved Precancer detection (F1 = 0.84, ROC-AUC = 0.978)** while maintaining interpretability (focus on nuclei).
- **Takeaway**: Careful augmentation boosts reliability of AI for cervical cancer screening, with potential impact in underserved regions

INTRODUCTION

- Cervical cancer:
  - Highly **preventable** and **treatable** when **caught early**.
  - Mortality is concentrated LMICs.
- Problem:
  - Pap smears depend on human interpretation, which is **expert-intensive** and **vulnerable to variability** (e.g., stains, illumination, cell morphology).
  - Shortage of pathologists leads to late diagnoses and worst outcomes.
- Why use AI?
  - Convulutional Neural Networks (CNNs) can recognize cell features (nucleus, cytoplasm).
  - An effective cell-level classifier will prioritize suspicious fields for pathologist review to **reduce missed lesions and speed triage**.
  - However, models often fail to generalize due to **staining/ brightness/morphology/variability.**
- Goal of this Project:
  - Test augmentation and preprocessing strategies to **improve accuracy** and **minority-class detection** (Precancer)
  - Provide insights into what augmentations help vs. hurt in medical imaging

METHODS

- Dataset: 3,127 Pap smear cell images, labeled (Normal vs Precancer).
- Preprocessing: resize (256x256), normalization, train/validation split (stratified)
- Model: ResNet-18 CNN, trained with weightened cross-entropy loss to balance Normal vs. Precancer (some types appear less).
- Augmentation Strategies Tested:
  - Original (baseline model): resize, normalization, random flips (H/ V), random rotation (+90 degrees, -90 degrees), random crop
  - Light Augmentation: resize, normalization, random flips, gentle rotation (+10 degrees, -10 degrees), random crop
  - Enhanced Augmentation: same as Light Augmentation, with additional mild color jitter (brightness/contrast/saturation) and slightly larger rotation (+15 degrees, -15 degrees)
  - Evaluation:
  - Metrics: Accuracy, macro-F1, per-class F1; ROC-AUC (how well model tells apart Normal vs. Precancer cells across theresholds)
  - Interpretibility: Captum (integrated gradients, NoiseTunnel) to visualize model focus

RESULTS

- Performance Across Augmentation Strategies:
  - Original (baseline model) achieved macro-F1 = 0.867, with a weaker performance on the minority Precancer class (F1 = 0.779).
  - Light augmentation improved generalization (macro-F1 = 0.887, Precancer F1 = 0.840).
  - **Enhanced augmentation** yielded the **best results** (macro-F1 = **0.905**, Precancer F1 = **0.840**)
- ROC-AUC for the Ehanced augmentation model was **0.978**, indicating excellent separation between Normal and Precancer cell types.
- Atrribution maps confirmed the model focused on the nucleus and nuclear-cytoplasmic boundary, **consistent with cytological features**.

	run	acc	macro_f1	f1_Normal	f1_Precancer
1	original	0.9249201277955271	0.8670534596736661	0.9547641963426372	0.7793427230046949
2	light	0.9361022364217252	0.8872742333387356	0.9614643545279383	0.8130841121495327
3	enhanced	0.9488817891373802	0.9047908745247148	0.9695817490494296	0.84

Table 1: Classification performance of three augmentation strategies on Pap smear cell images, evaluated by Accuracy, Macro-F1, and per-class F1 scores.

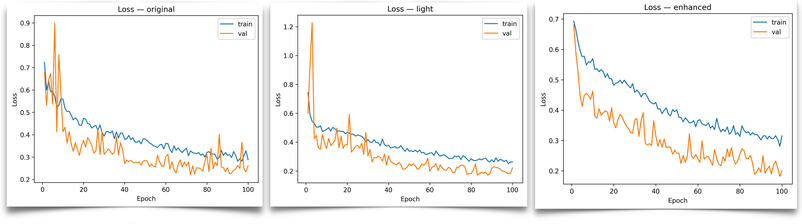

Figure 1: Training and validation loss curves for the original model, light, and enhanced augumentation runs (100 epochs).

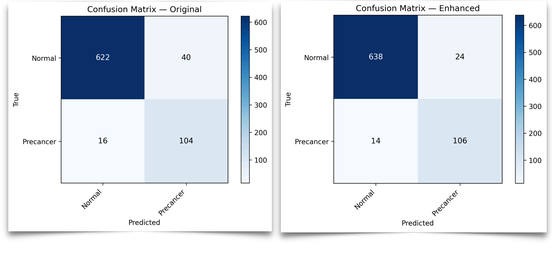

Figure 2: Confusion matrices for Original vs. Enhanced runs, showing reduction in false negatives for the Precancer class after augmentation.

RESULTS

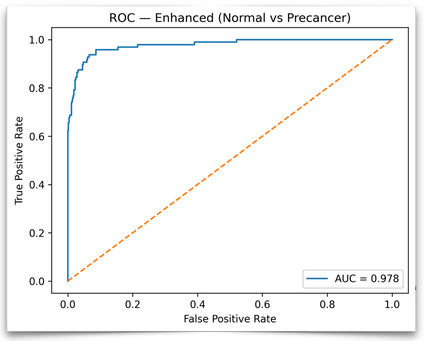

Figure 3: ROC curve for the Ehanced Augmentation run, with ROC-AUC = 0.978, demonstrating strong sepration of Normal versus Precancer cases.

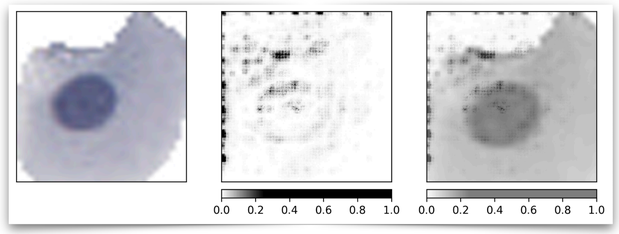

Figure 4: Captum attribution maps (NoseTunnel, Integrated Gradients) for a Precancer example.

CONCLUSION

- Data augmentation strategies **improved the performance of a CNN classifier** for Pap smear cell images.
- Enhanced augmentation (rotation + mild color jitter) yielded the best results: Macro-F1 = 0.905, Precancer F1 = 0.840, ROC-AUC = 0.978
- Interpretability analysis confirmed the model focused on nuclei and nuclear-cytoplasmic boundaries aligning with established cytological features.
- Potential for Clinical Impact:
  - AI-based Pap smear screening could **expand access** to early cervical cancer detection in **low-resource regions** where cytopathologists are scarce.
  - Improved detection of Precancerous cells may **reduce false negatives**, enabling earlier intervention and better patient outcomes.
- Limitations and Future Directions:
  - Dataset was relatively small and limited to two classes (Normal vs. Precancer).
  - Augmentation strategies tested were basic; further work could explore stain normalization, domain adaptation, or transfer learning
  - Validation was limited to a single dataset. Multi-site, multi-population testing is required for clinical translation.
  - Future goal: integration into a web-based tool for global screening access.

edit.