Classification of Fluid Cytology Whole-Slide Images Using Pretrained Vision Models and Multiple-Instance Learning

Afya Shaikh

INTRODUCTION

- Fluid cytology is used to evaluate cells collected from cytologic specimens for evidence of malignancy.
- Automated analysis of fluid cytology whole-slide images is difficult because diagnostically important cells may be sparse, unevenly distributed, and mixed with large areas of background or artifact.
- Most retrospective datasets also lack detailed cell-level annotations, making slide-level learning approaches especially important.

RESEARCH QUESTION

Can pretrained vision models and multiple-instance learning classify fluid cytology whole-slide images into four diagnostic categories using only slide-level labels?

BACKGROUND

- Whole-slide images contain large amounts of non-diagnostic background.
- Cell-level annotation is time-consuming and was not available in this dataset.
- Multiple-instance learning is useful when only slide-level diagnoses are known.
- Pretrained vision models may capture useful morphology even in small datasets.

MATERIALS & METHODS

- Dataset:
  - 80 de-identified fluid cytology WSIs
    → 4 unreadable slides excluded
    → 76 slides total
- Patch Selection:
  - 24 candidate 512 × 512 patches/slide
    → ranked by sharpness, contrast, and stained area
    → 8 patches/slide retained
    → 608 patches total
- Models:
  - ResNet-18
  - DINO-v2
  - Phikon-v2
  - Attention-based MIL
- Evaluation:
  - 5-fold slide-level cross-validation
  - Accuracy, balanced accuracy, precision, recall, macro F1

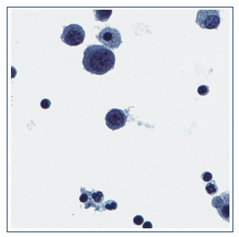

Figure 1: A microscopy image showing scattered stained cells on a mostly clear background, with a few small cell clusters.

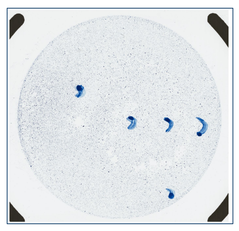

Figure 2: A whole-slide view of a fluid cytology specimen with scattered stained material.

RESULTS

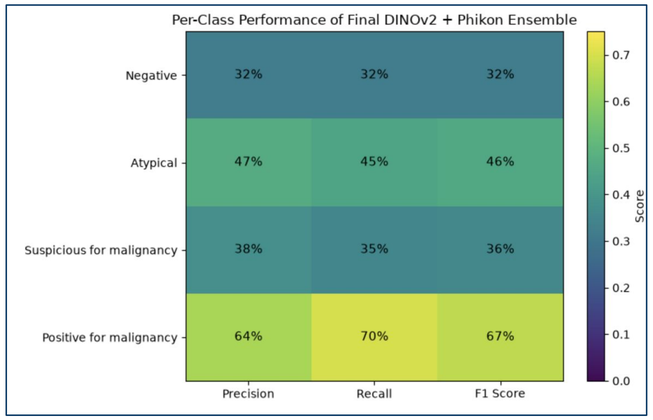

Figure 3: Class-specific performance of the final ensemble.
Positive-for-malignancy slides showed the strongest performance, with 70% recall and a 67% F1 score.

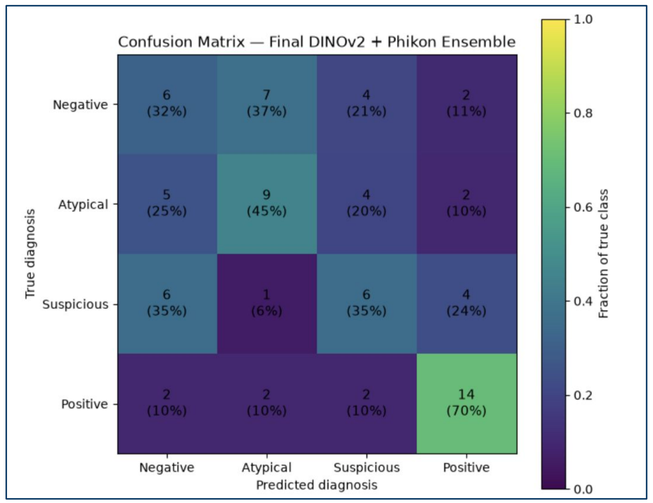

Figure 4: Confusion matrix for the final DINOv2 and Phikon-v2 ensemble. Fourteen of 20 positive-for-malignancy slides were correctly classified.

Model	Accuracy	Macro F1
ResNet-18 baseline	35.5%	34.1%
Original DINOv2	31.6%	30.1%
Quality-controlled DINOv2	38.2%	37.8%
DINOv2 + Attention MIL	43.4%	41.9%
Phikon-v2 + Attention MIL	31.6%	30.3%
DINOv2 + Phikon-v2 ensemble	46.1%	45.2%

Figure 5: Comparison of slide-level classification performance across model experiments.

DISCUSSION

- Quality-controlled patch selection improved DINOv2 performance, suggesting that reducing blurred and low-information regions helped the model. Attention-based MIL improved performance further by allowing some patches to contribute more strongly than others to the slide-level prediction.
- Positive-for-malignancy slides showed the strongest performance, while negative, atypical, and suspicious-for-malignancy slides were harder to separate. This may reflect greater morphologic overlap among the intermediate diagnostic categories and the fact that only slide-level labels were available.
- The final ensemble produced the highest exploratory performance, but the small dataset, sparse sampling of only eight patches per slide, and lack of cell-level annotations limit how broadly these results can be interpreted. Independent validation is needed before any clinical use.

CONCLUSION

Pretrained vision models and attention-based MIL captured diagnostic signal from fluid cytology WSIs using only slide-level labels. The best ensemble reached 46.1% accuracy, with strongest performance in positive-for-malignancy slides, but larger annotated datasets and external validation are needed.

REAL WORLD APPLICATIONS

- Potential future use as a decision-support or triage tool
- May help identify slides more likely to contain malignant morphology
- Future work should include:
  - larger datasets
  - broader slide coverage
  - cell-level or cluster-level annotations
  - external validation

KEY REFERENCES

1. [1] Jiang, H., Zhou, Y., Lin, Y., Chan, R. C. K., Liu, J., & Chen, H. (2023). Deep learning for computational cytology: A survey. *Medical Image Analysis*, 102691. https://doi.org/10.1016/j.media.2022.102691
2. [2] Lu, M. Y., Williamson, D. F. K., Chen, T. Y., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, *5*(6), 555–570. https://doi.org/10.1038/s41551-020-00682-w
3. [3] Kim, D., Lee, J., Jung, M., Yim, K., Hwang, G., Yoon, H., Jeong, D., Cho, W. J., Alam, M. R., Gong, G., Cho, N. H., Yoo, C. W., Chong, Y., & Seo, K. J. (2025). Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning. *Lung Cancer (Amsterdam, Netherlands)*, *204*, 108552. https://doi.org/10.1016/j.lungcan.2025.108552