# EXPLORING FOUNDATION MODELS FINE-TUNING FOR CYTOLOGY CLASSIFICATION

*Manon Dausort , et al.*

ICTEAM, Universite´ Catholique de Louvain, Belgium

ILIA, Universite´ de Mons, Belgium
CMMI, Universite´ Libre de Bruxelles, Belgium

ISBI 25

# Introduction

- Cytology slides are vital for diagnosing and staging cancer, offering detailed views of abnormal cells to guide treatment decisions.
- Analyzing these slides is labor- intensive and costly, leading to delays in the reporting process, making automation essential to improve both classification efficiency and accuracy.

- An effective approach to automate classification involves using deep learning models. However, the performance of these models is highly dependent on the quantity and diversity of data.

=> ways to utilize limited labeled data by few-shot learning, pretrained-model and fine-tuning

# Introduction

- They evaluate five foundational models pre-trained on natural, biomedical and histopathological images on four cytology classification on 4 benchmark datasets.

- Fine-tuned the foundational models with LoRA in few-shot settings which can reduce trainable parameters.

# Introduction: Dataset

- Mendeley LBC Cervical Cancer

  **963 images** sub-divided into four sets of images representing the four classes of pre-cancerous and cancerous lesions

- Body cavity fluid cytology images

  **160 benign and 533 maglinant** images, Leishman stained, photographed at 40X magnification, then resized to 256 x 192 pixel

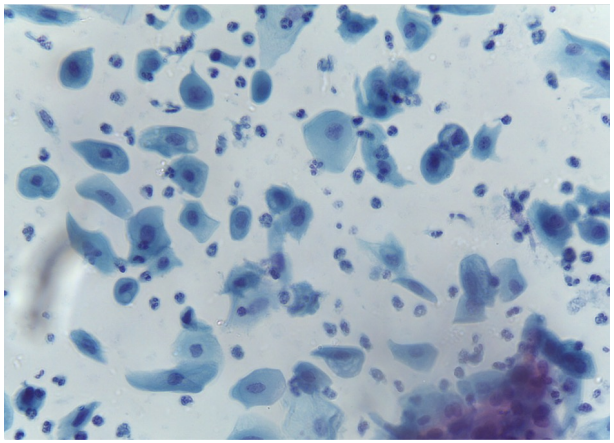
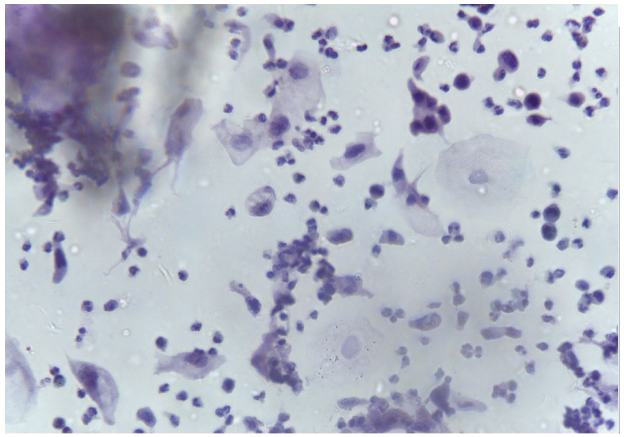
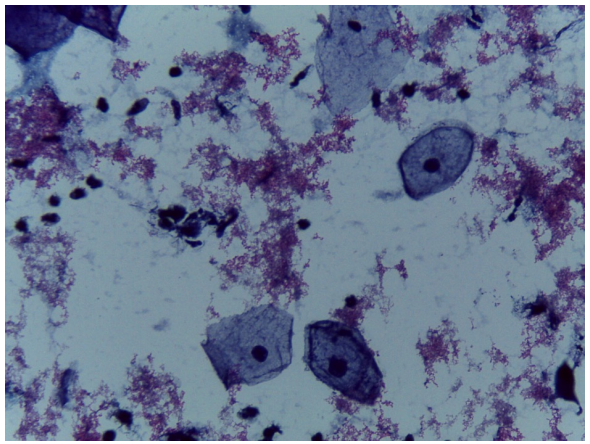
- HiCervix Dataset:

  Includes **4,496 whole slide images** from 40,229 cervical cells, categorized into 29 annotated classes within a three-level hierarchical tree to capture fine-grained subtype information

- Cervical Cancer largest dataset (SipakMed)

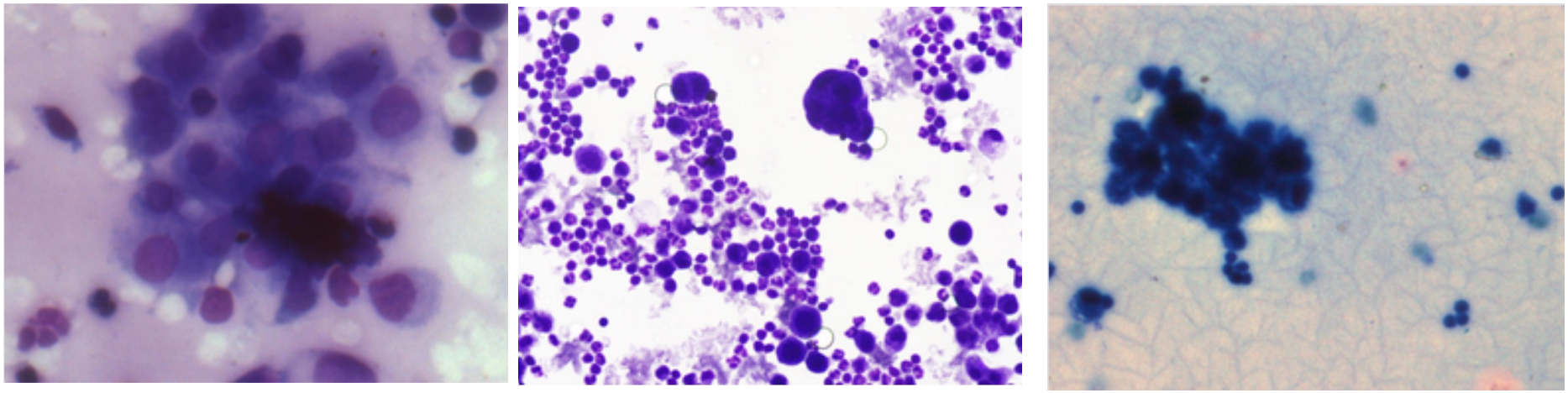
  **4049 images** of isolated cells that have been manually cropped from 966 cluster cell images of Pap smear slides, five categories containing normal, abnormal and benign cells

# Introduction: Dataset

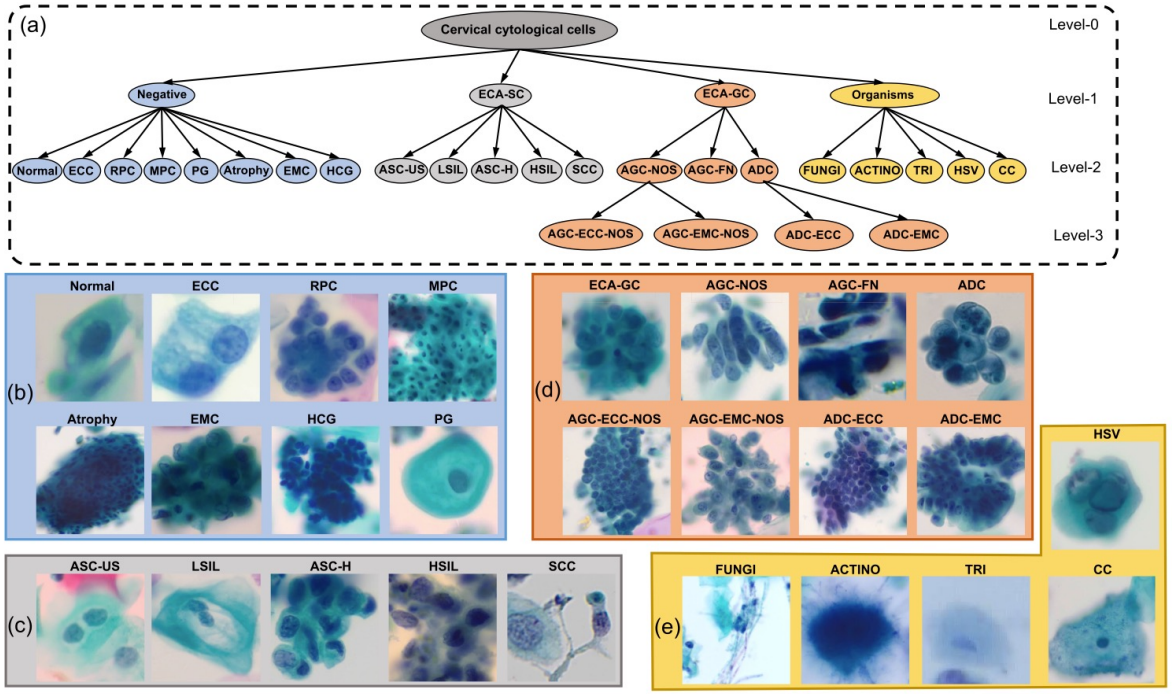
**Mendeley LBC Cervical Cancer**

# Introduction: Dataset



**Body cavity fluid cytology images**

# Introduction: Dataset

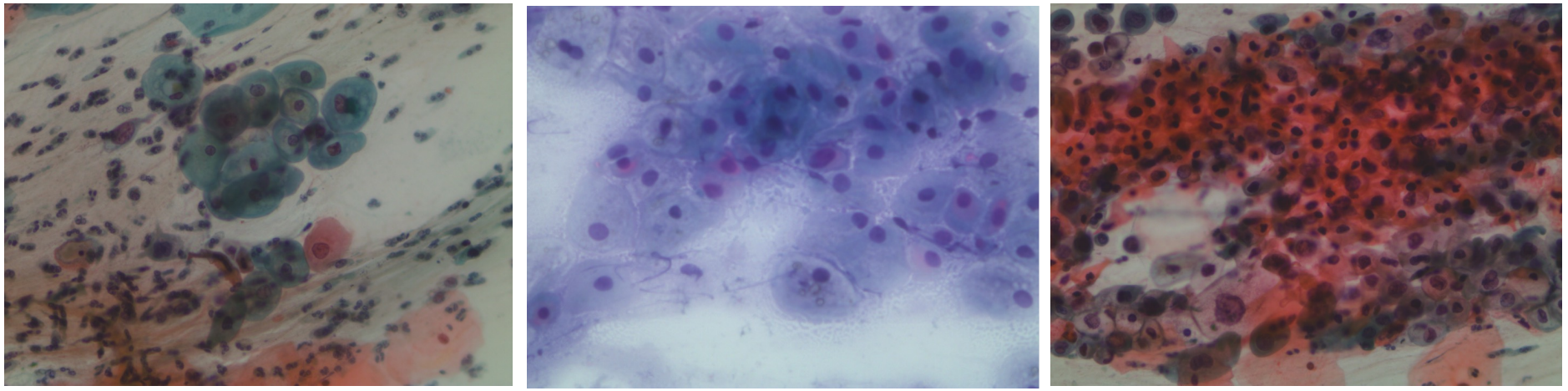



D. Cai *et al.*, "HiCervix: An Extensive Hierarchical Dataset and Benchmark for Cervical Cytology Classification," in *IEEE Transactions on Medical Imaging*, vol. 43, no. 12, pp. 4344-4355, Dec. 2024

## HiCervix Dataset

The three-level hierarchical structure of the cervical cytological cells in HiCervix. The root node of cervical cytological cells is at level-0 and the granularity goes from coarse- to fine-grained in a top-down way.

# Introduction: Dataset

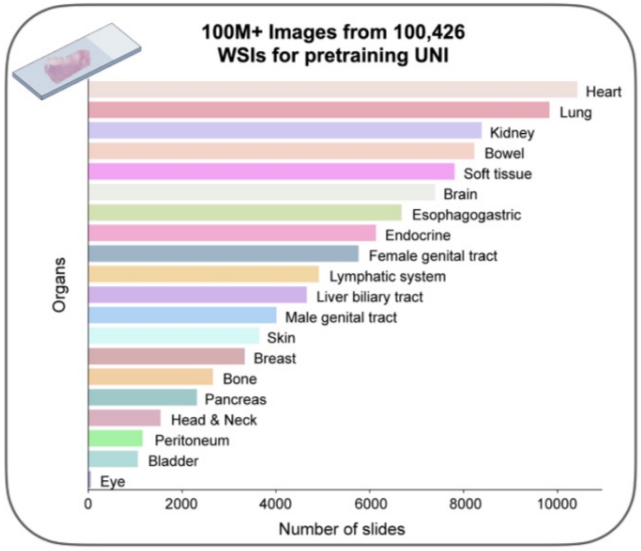


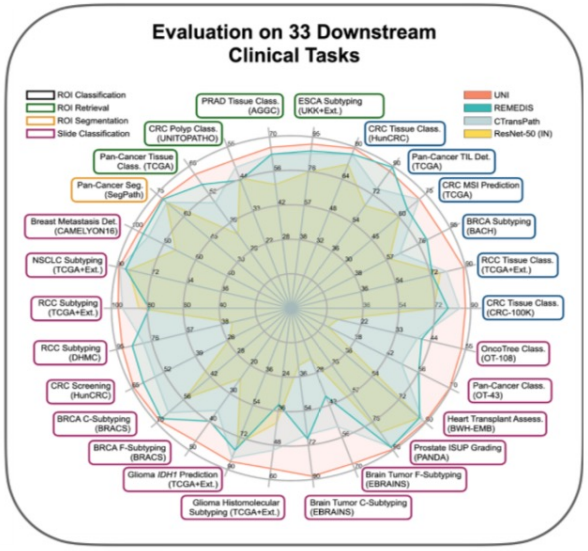
**Cervical Cancer largest dataset (SipakMed)**

# Introduction

Foundation models:

- CLIP
- ViT
- BiomedCLIP
- QUILT
- UNI




Xu, H., Usuyama, N., Bagga, J. *et al.* A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181–188 (2024).

# LoRA (Low rank Adaptation)

- **Linear classifier:**

  Assume we have a pretrained model with parameter $\theta$, and feature extractor $z = f_\theta(x)$. To do classification, learn a projection $W$: $y = Wz$

  $z \in \mathbb{R}^n$, $W \in \mathbb{R}^{c \times n}$.

- **LoRA:**

  Update weight: $W \in \mathbb{R}^{m \times n}$ -> $W = W + \Delta W$, $\Delta W \in \mathbb{R}^{m \times n}$

  Represent weight update with submatrices: $W = BA$, $B \in \mathbb{R}^{m \times r}$, $A \in \mathbb{R}^{r \times n}$

  The output: $h = Wx + \gamma \Delta W x = Wx + \gamma BAx$ , $\gamma$ is the scaling factor

  Assume m>n, the the trainable parameter: $\boldsymbol{mn \rightarrow r(m+n)}$.

  $mn > n^2 > 2rm > r(m+n)$

  If $r \ll n$, $\boldsymbol{2rm \ll n^2}$.

# Few-shot learning

- Train: The support set contains K labeled training samples for each of the N classes. The model uses these support samples to learn generalized representations for each class.

  For example, the dataset for a *3-way-2-shot* classification task contains 3 classes of images and provides 2 examples of each.

- Test: The query set contains one or more new examples for each of the N classes. Using representations learned from the support set, the model predicts classification for each example in the query set.

# Few-shot learning


*Few-shot learning*

Example:
Suppose the support set composed of K examples per category. We want to learn the representation for each category: “CAR”, “BIRD”, “ANIMAL”, “FLOWER”.

At test time, examples from query set are used for prediction with representation learned.

# Experiments

Experiment 1: Fine-tune the linear classifier layer while freezing the backbone
(Results are averaged over 3 seeds. Top-1 accuracy is evaluated on the validation set)

**Table 1.** Mean accuracy of fine-tuned classifiers, evaluated on five models and four datasets.

| | Models | | | | |
|---|---|---|---|---|---|
| **Datasets** | CLIP | QUILT | BiomedCLIP | UNI | ViT |
| BCFC | 94.74 ±0 | **100±0** | 96.17±0 | 99.04±0 | 98.72±0.28 |
| MLCC | 93.95 ±0.52 | 97.95±0 | 95.1±0.52 | **99.2±0.4** | 96.8±0.2 |
| SIPaKMed | 90.5 ±0.16 | 92±0.57 | 87.62±0.33 | **93.33±0.16** | 89.71±0.49 |
| HiCervix | 55.2±0.2 | 54.64±0.08 | 46.39±0.1 | **64.07±0.25** | 57.64 ±0.14 |
| **Average** | 83.59 | 86.15 | 81.32 | **88.91** | 85.72 |

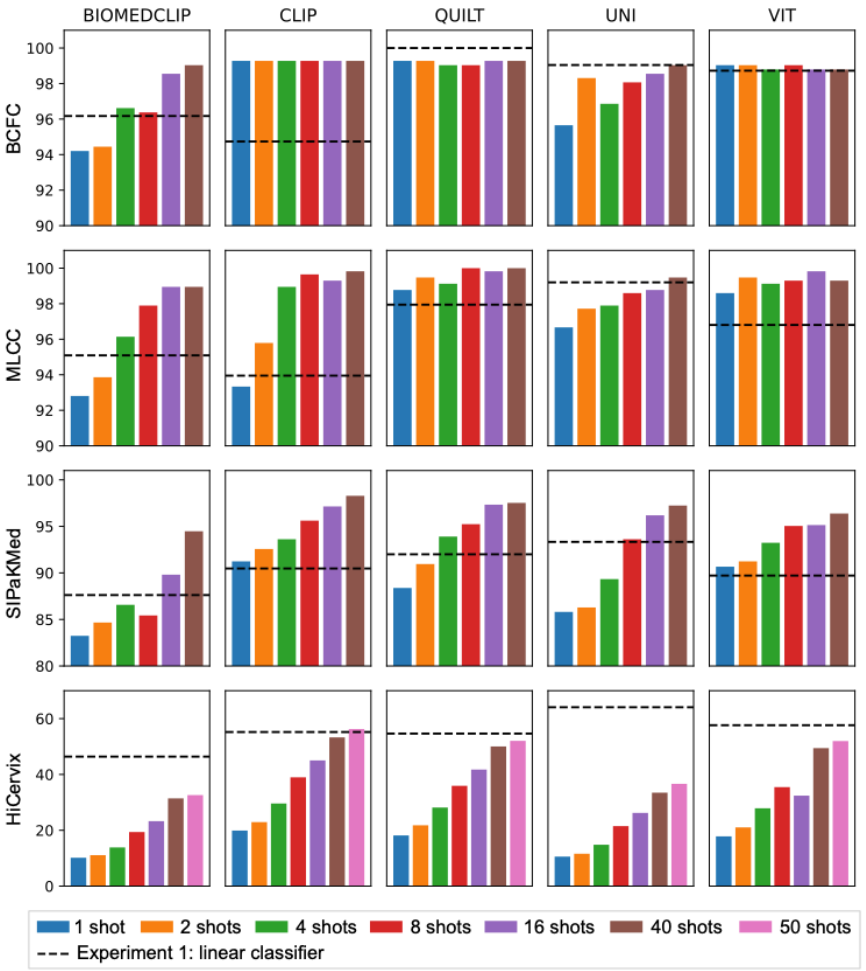
Histopathology-pre-trained FMs are better to extract discriminant features for cytology tasks. (UNI, QUILT)

# Experiments

## Experiment 2: LoRA few-shot adaptation

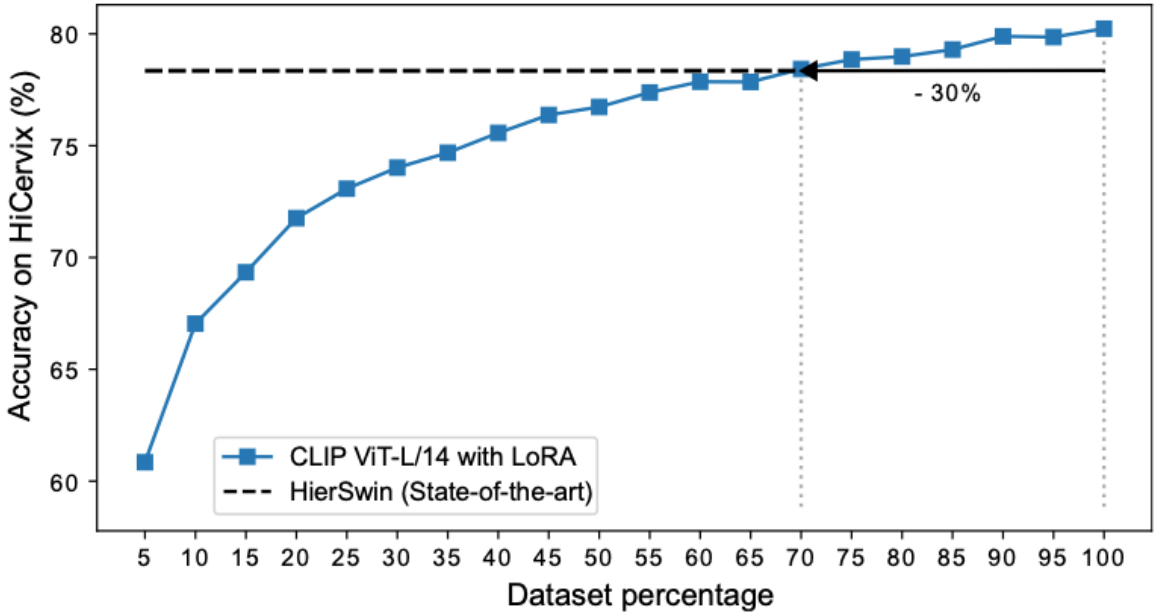
Mean accuracy of fine-tuned foundation models, evaluated on four datasets in a few-shot setting with a varying numbers of shots. The black horizontal line reports the mean accuracy from Experiment 1.

- Models pre-trained on medical or histopathological images require more data to reach comparable performance to those trained on more diverse datasets.
- Fine-tune backbone > only classification head (apply LoRA to the query and value matrices of each attention block in the visual encoder)



# Experiments

Experiment 3:Pushingmodelfine-tuninglimits

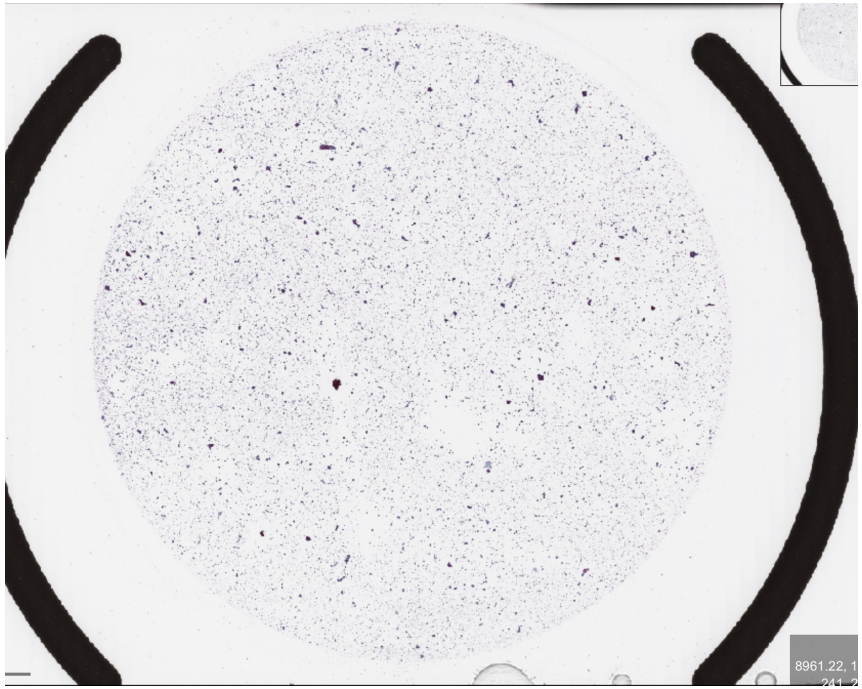


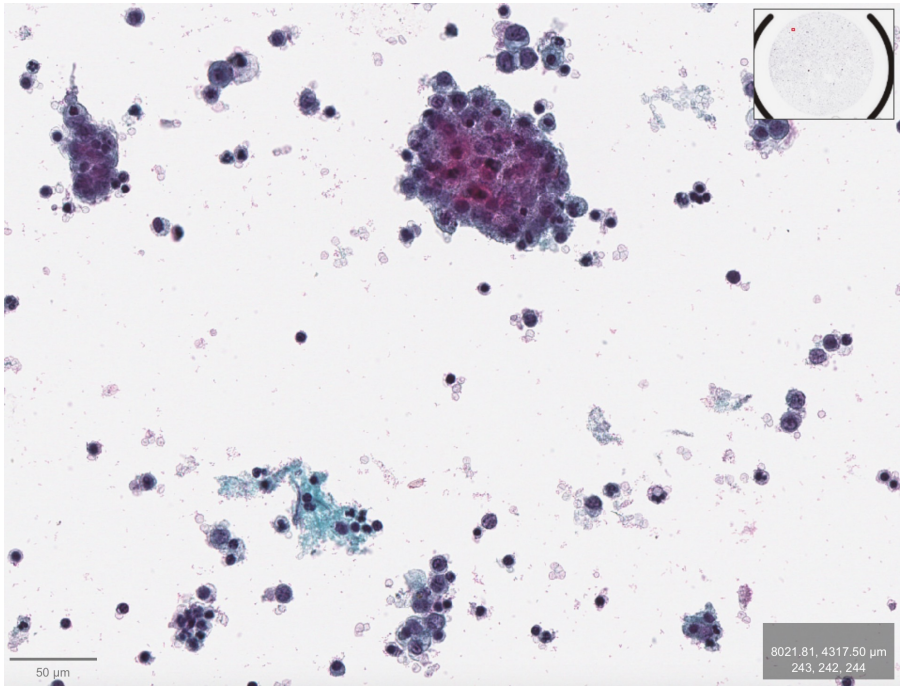

Mean accuracy of CLIP fine-tuned with LoRA and a ViT-L/14 backbone on different percentage of the HiCervix dataset. The black horizontal line represents the state-of-the- art accuracy (HierSwin).

Model fine-tuned on 70% of the HiCervix dataset equals HierSwin in terms of both parameter efficiency and data usage.

Trained with 100% -> outperforms HierSwin

# Mayo cytology slide





# THANK YOU!