Cytological Evaluation of Papillary Thyroid Carcinoma: Deep Learning Analysis of Thyroid Fine-Needle Aspiration Biopsies

Anvith Kakkera, Aneesh Kalla, Anish Malepati

OVERALL WORKFLOW

ABSTRACT

- Research aims to reduce **unnecessary resections** for patients with **indeterminate** histopathology
- The workflow developed is similar to a cytopathologists' workflow and different from traditional deep learning workflows
- **Cell nuclear atypia Convolutional Neural Network (CNN)** classification achieved **66.7% accuracy** and **AUC Score** of **0.72**
- **Cluster nuclear atypia classification** Graph Neural Network(GNN) achieved **78.2%** accuracy and **0.79 AUC**
- **Cluster architectural atypia CNN** achieved **88.1%** accuracy and **0.894 AUC**
- **Giotto-tda Random Forest Classifier** classified **microfollicular** vs. **crowded** with an **accuracy** of **83.6%** and **AUC** of **0.881**

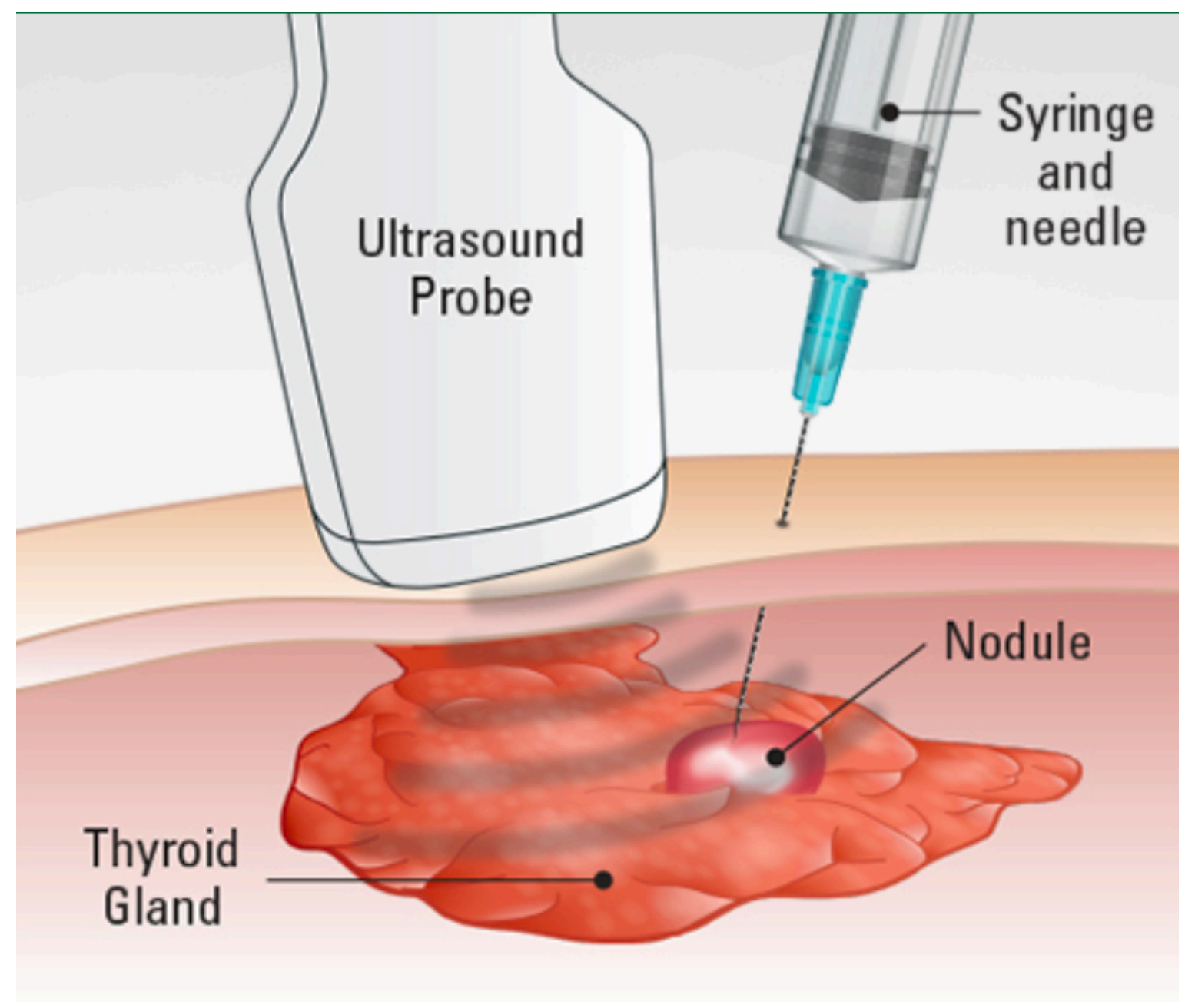
INTRODUCTION

Patient Develops Thyroid Nodule

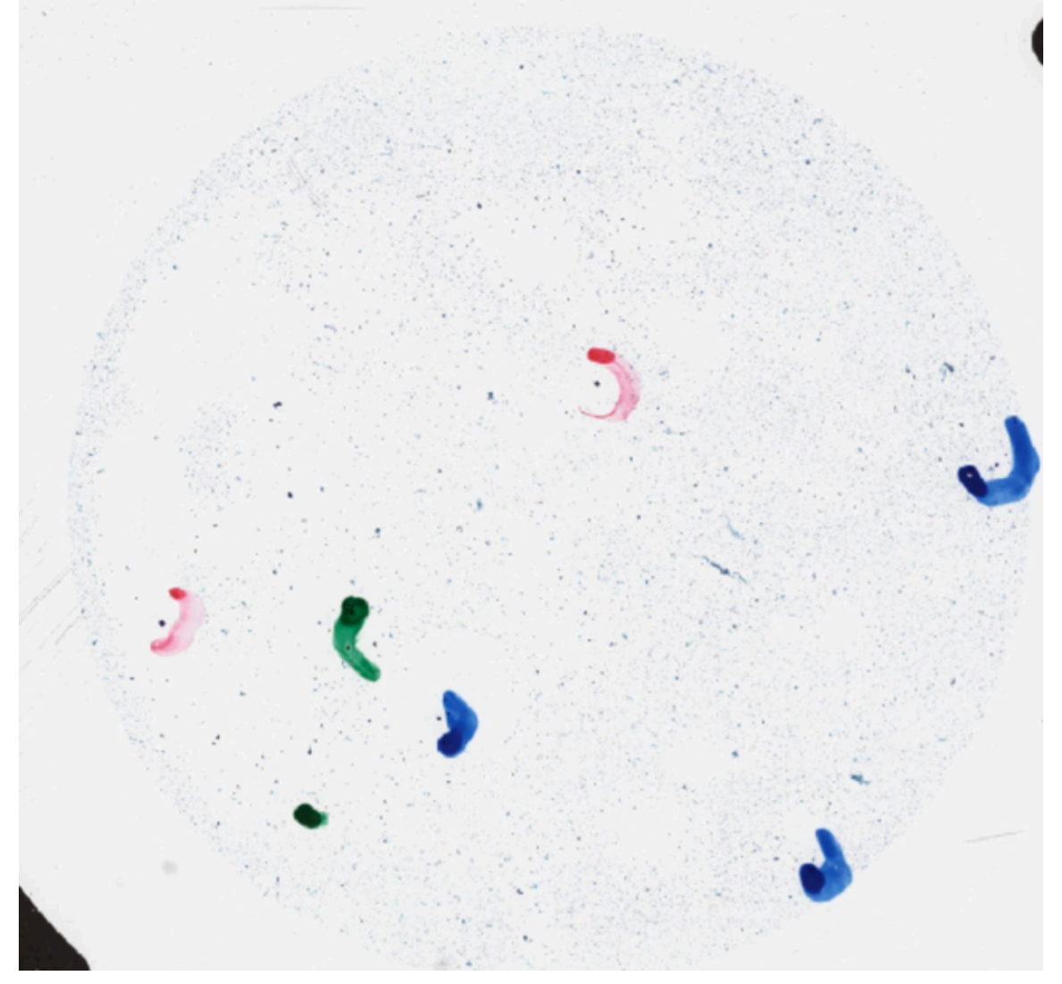
Doctor Screens Patient for Thyroid Cancer via a Fine Needle Aspiration Biopsy (FNA)

Image the FNA To Make A Whole Slide Image (WSI)

Pathologist classifies FNA using Bethesda System



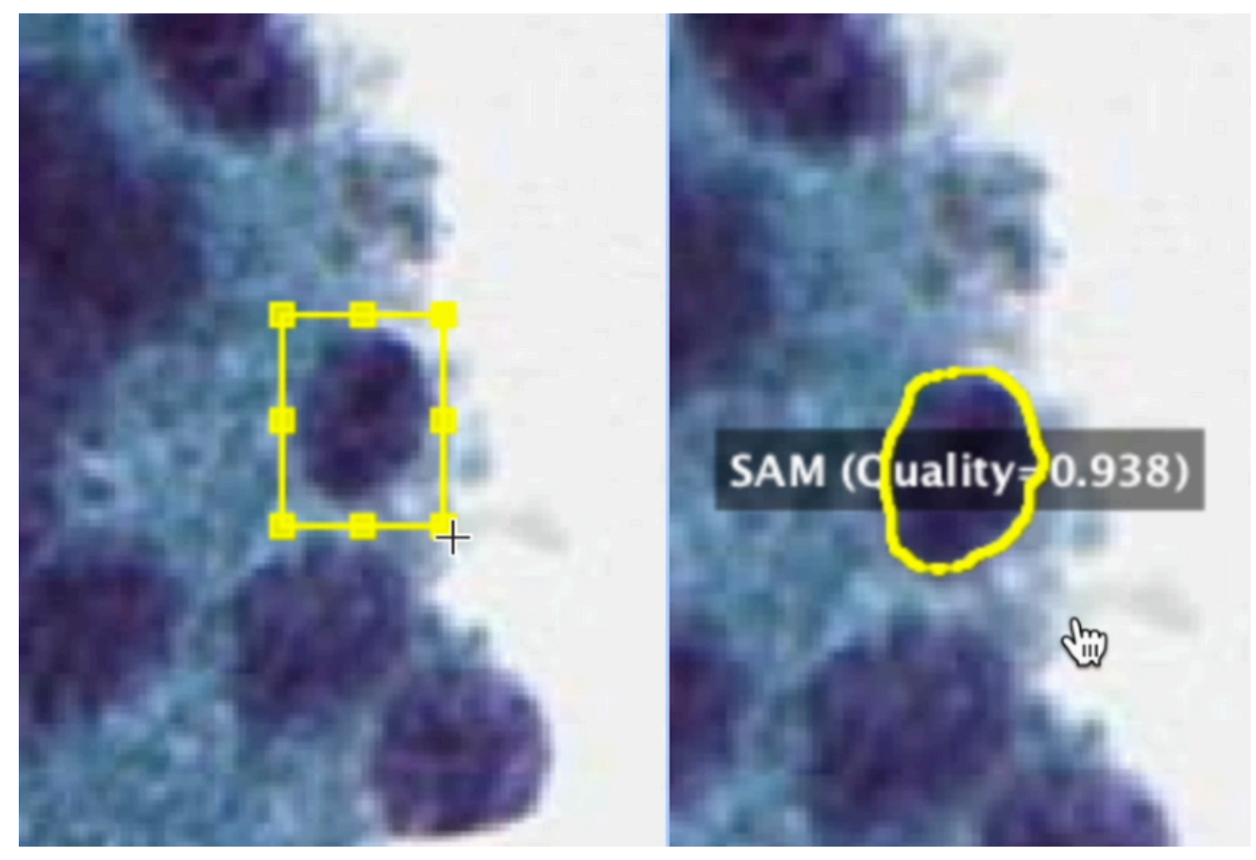
Visualization of a Fine Needle Aspiration Biopsy (FNA)
https://flushingvascular.com/thyroid-nodules-biopsy/



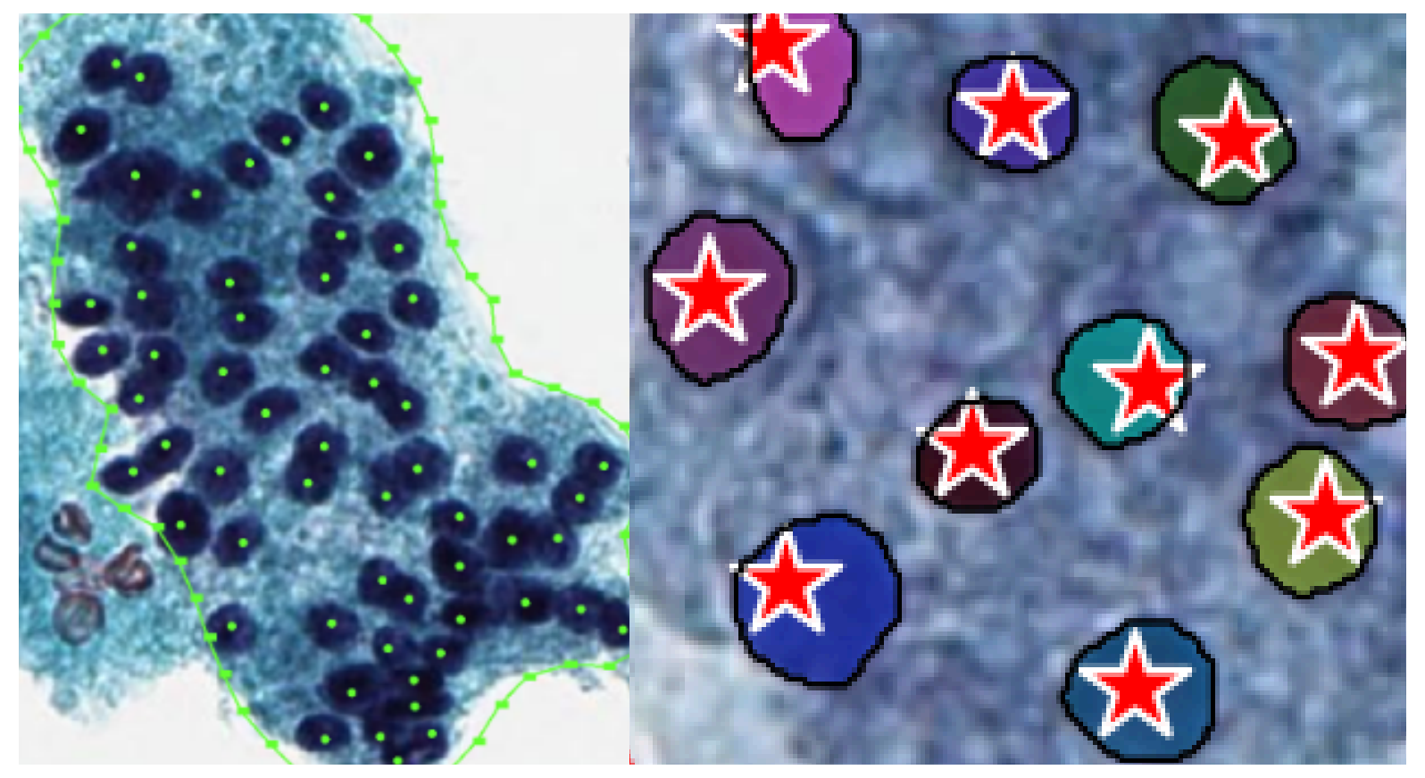
Whole Slide Image of Thyroid FNA biopsy

CELL MASK GENERATION

- To train cell detection models, **cell annotations** were made on crowded cluster images
- Cluster images were loaded using QuPath, and the **Segment Anything** vision transformer model (**SAM**) was used
- A bounding box was drawn over a cell and SAM made a mask
- Annotated objects were saved as GeoJSON files
- **YOLOv8** was **trained with these annotations**



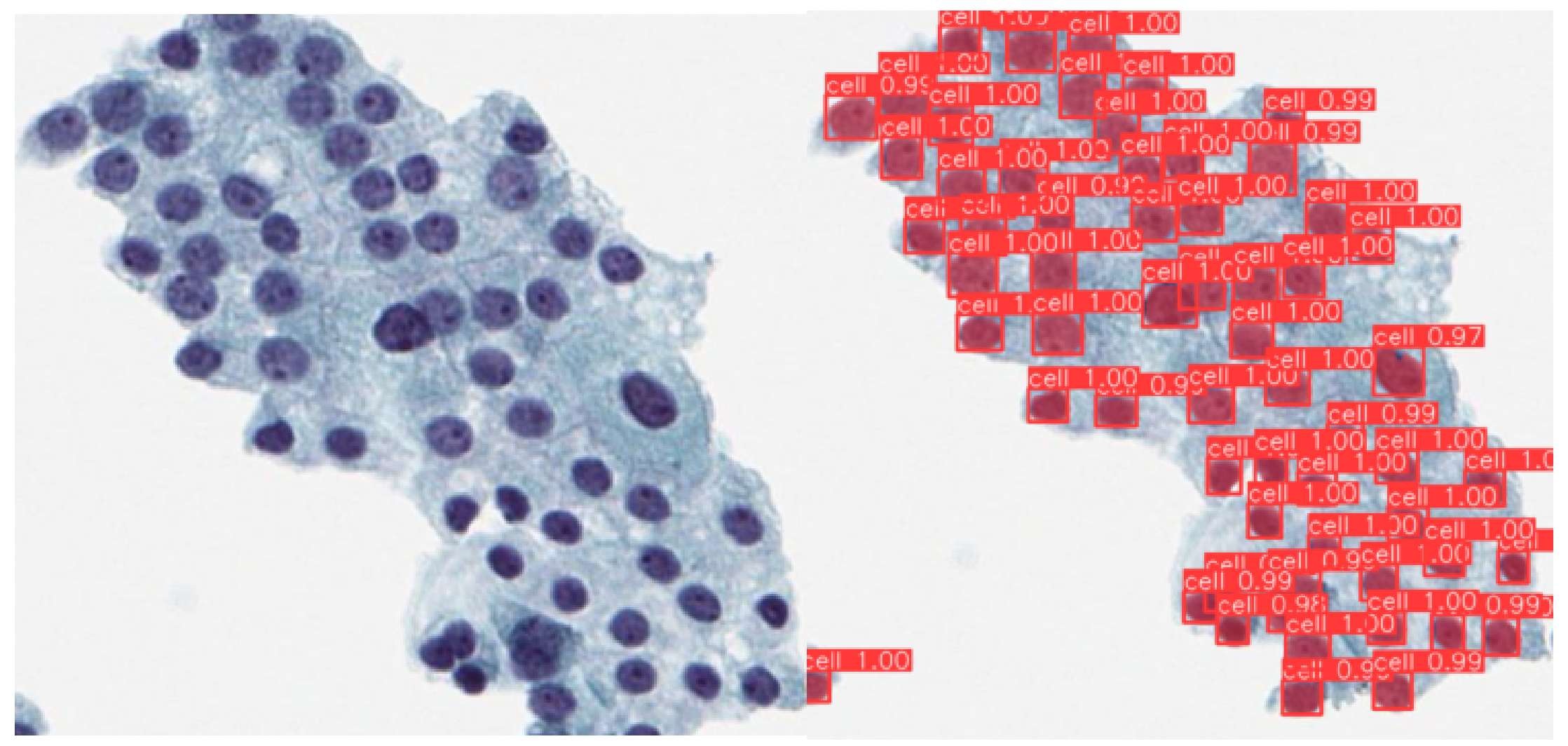
The semi-automated approach initially used. A bounding box is manually drawn, and then SAM automatically segments the cell.



The fully automated approach using SAM used dot annotations (left) to automatically generate cell masks (right).

YOLOV8

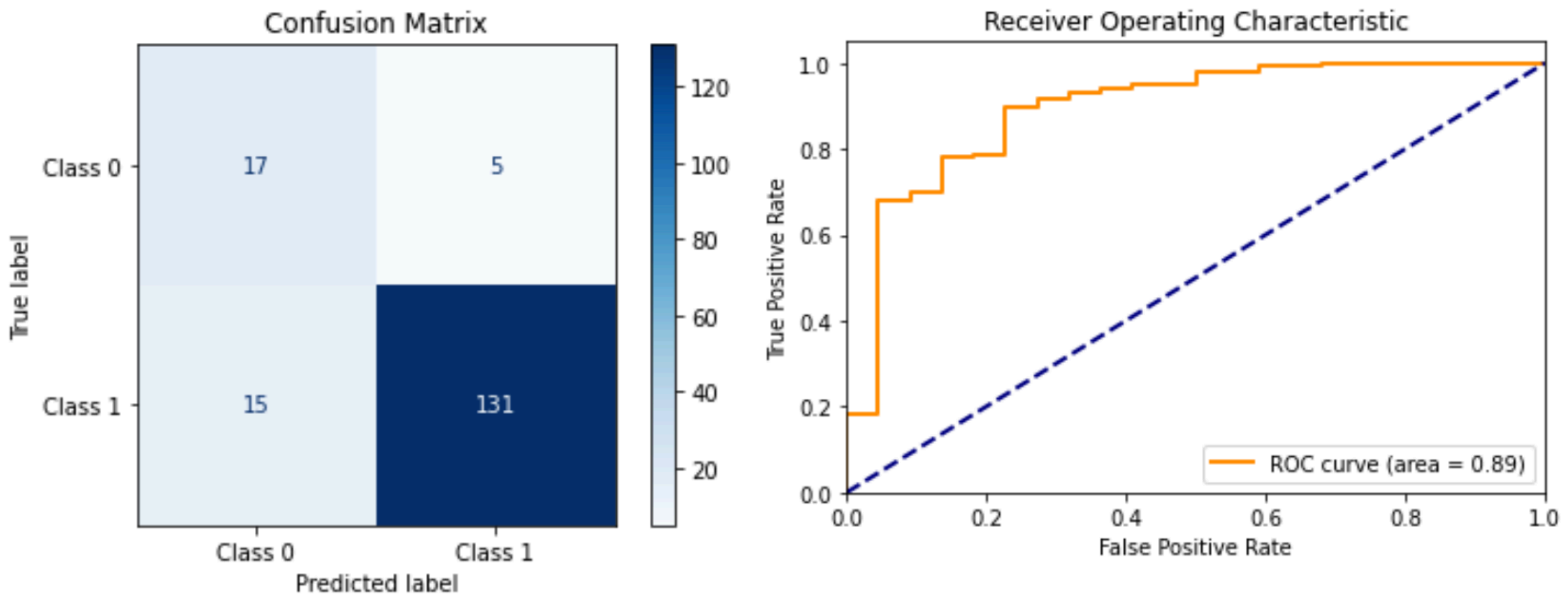
- YOLOv8 segmented cells from a cluster image
- Used pretrained weights from COCO
- Yolov8 achieved scores of...
  - **mAP50:** 0.8551
  - **mAP50-95:** 0.50695
  - **Precision:** 0.7377
  - **Recall:** 0.8733
- IOU threshold: 0.6 - Confidence threshold: 0.90



The image (LEFT) was fed into YOLOv8, and the segmented masks are shown (RIGHT)

ARCHITECTURAL ATYPIA CLASSIFICATION

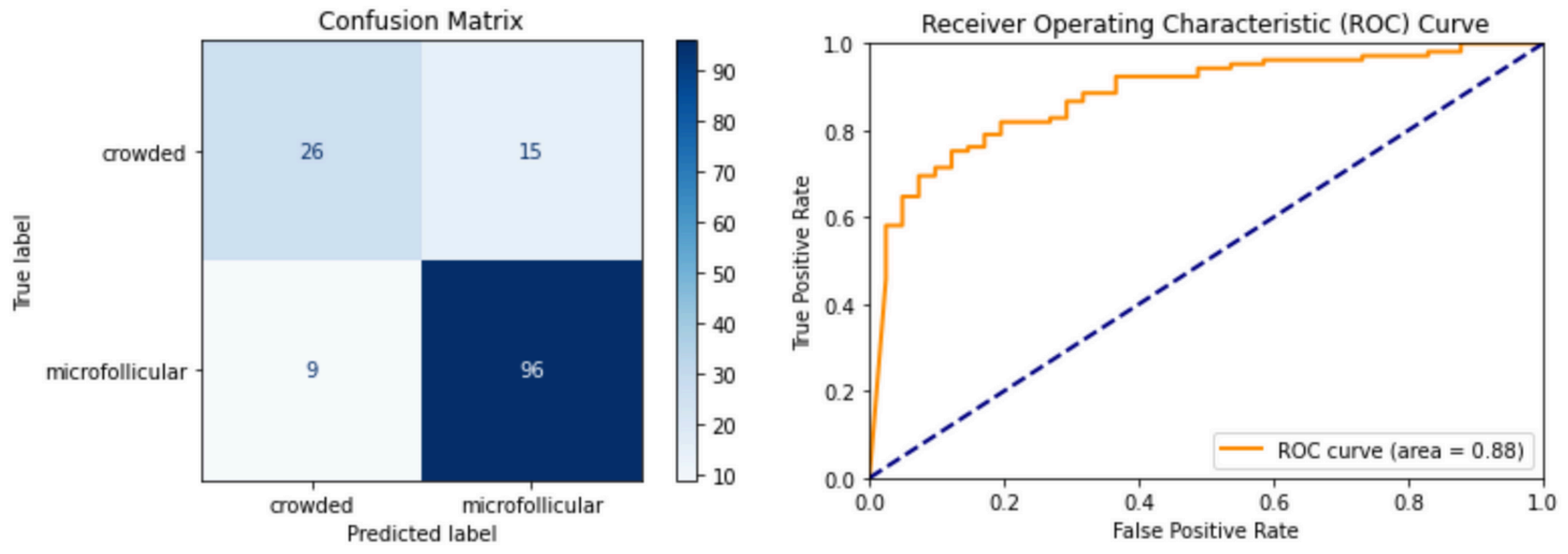
- **Custom PyTorch CNN (ArchCNN)** trained to classify **architectural atypia** (crowded & microfollicular) vs. **non-architectural atypia** (macrofollicular)
- **CNN** achieved an **accuracy** of **88.1%** and **AUC** of **0.894**



Confusion matrix of architectural atypia CNN (LEFT). Class 0 and 1 represents architectural atypia (crowded and microfollicular) and non-architectural atypia (macrofollicular), respectively. ROC Curve for the CNN (RIGHT)

CROWDED VS. MICROFOLLICULAR CLASSIFIER

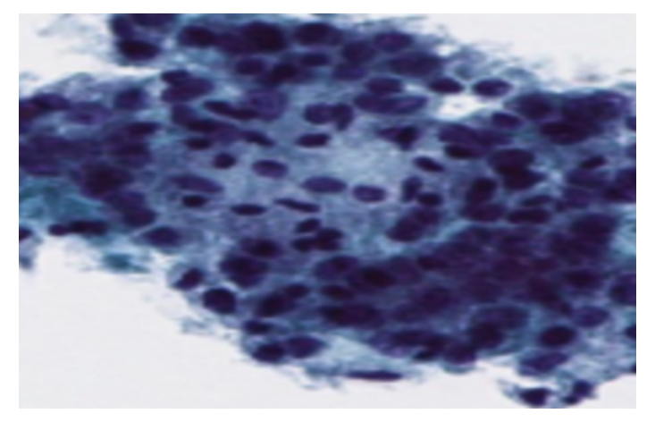
- Clusters marked **positive for architectural atypia** from the ArchCNN were **further classified** either as **crowded** or **microfollicular**
- **YOLOv8** was run on the clusters, and the **(x,y) center points** of each cell mask was stored
- **Vietoris-Rips Persistence** and **Persistence Entropy** via **giotto-tda** was used to **transform the (x,y) points** of each cell cluster
- A **Random Forest Classifier (RFC)** was trained on the **giotto-tda** transformations
- RFC achieved an **accuracy** of **83.6%** and **AUC** of **0.881**

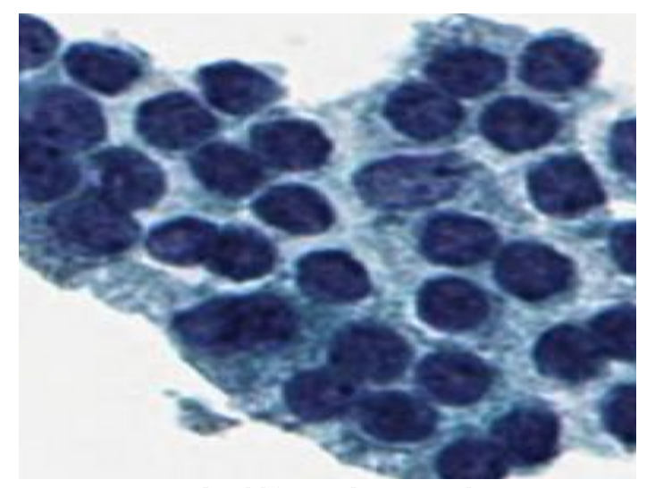


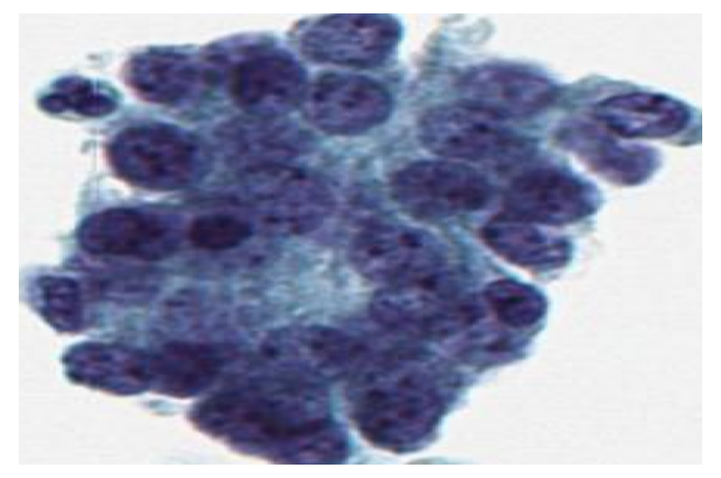
Confusion matrix of crowded vs. microfollicular RFF (LEFT) and ROC Curve for the RFF (RIGHT)

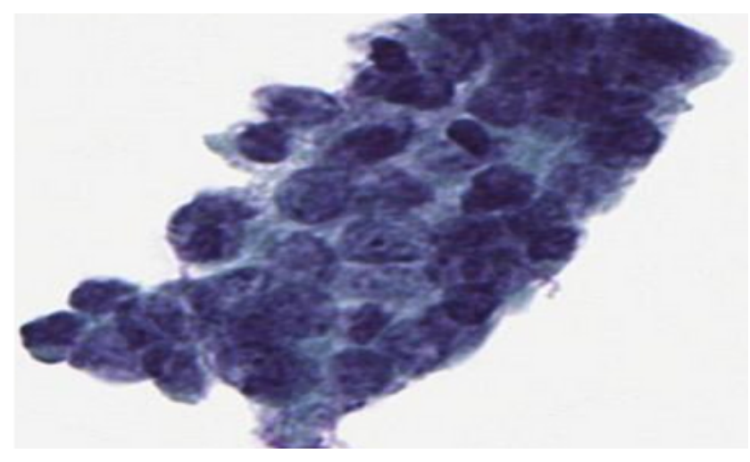
THYROID CYTOLOGY INTRODUCTION

- **Nuclear** and **architectural atypia** are cancer predictors
- **Nuclear atypia** has **atypical clusters** and cells
- **Architectural atypia** deals with **crowded** and **microfollicular,**
- **Non-architectural atypia** deals with **macrofollicular**


Crowded Cluster

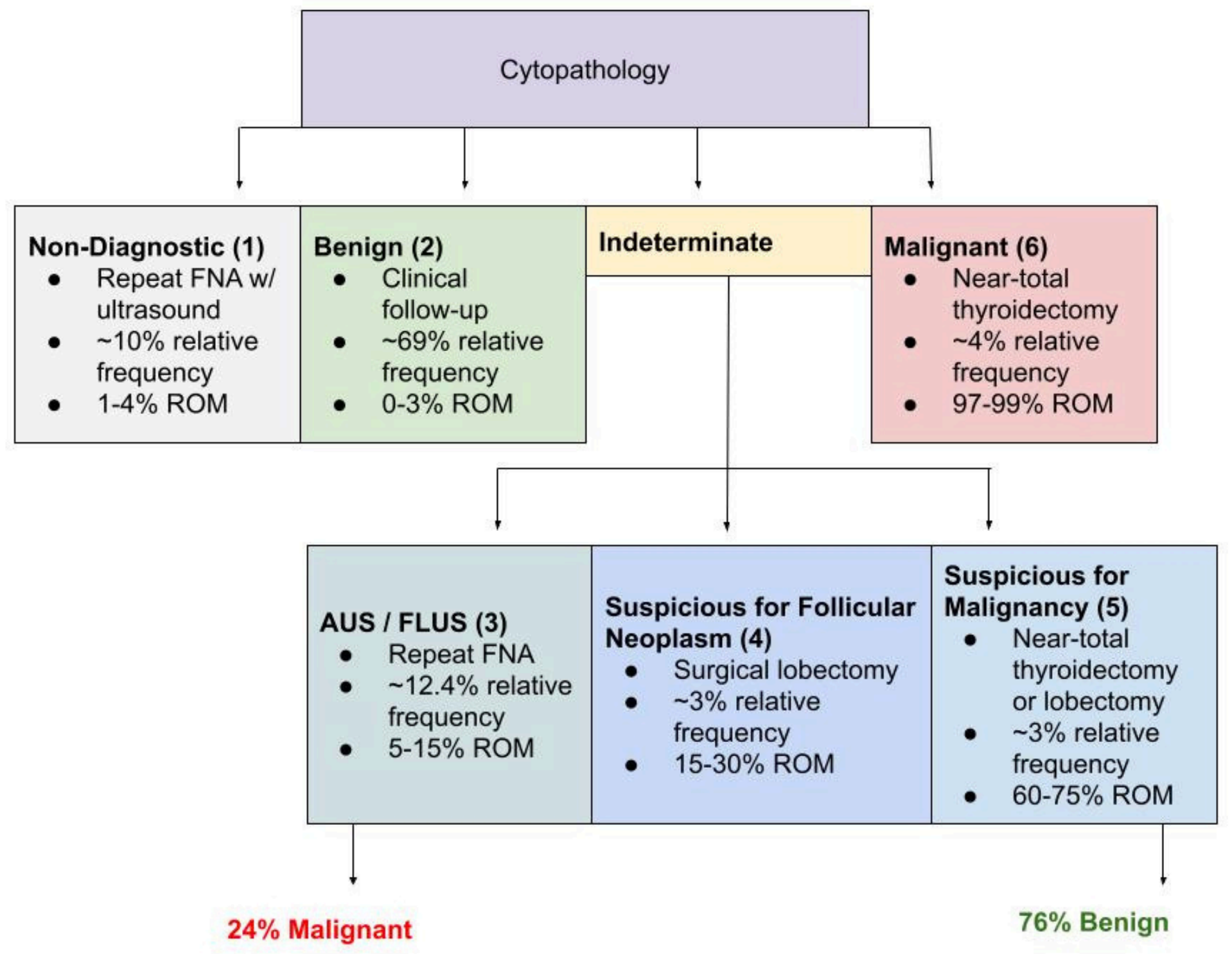

Macrofollicular Cluster


Microfollicular Cluster


Atypical Cluster

PROBLEM

- Category 3/4/5 are indeterminate, where doctors cannot diagnosis the FNA
- Indeterminate cases are generally treated with **thyroidectomy or lobectomy surgeries**
- **62%** of surgeries on indeterminate cases were **unnecessary**
- **Surgery complications** include: vocal cord damage, infection, bleeding, esophageal/nerve injury, and wasted healthcare resources/costs


Cytopathology

Non-Diagnostic (1)
- Repeat FNA w/ ultrasound
- ~10% relative frequency
- 1-4% ROM

Benign (2)
- Clinical follow-up
- ~69% relative frequency
- 0-3% ROM

Indeterminate

Malignant (6)
- Near-total thyroidectomy
- ~4% relative frequency
- 97-99% ROM

AUS / FLUS (3)
- Repeat FNA
- ~12.4% relative frequency
- 5-15% ROM

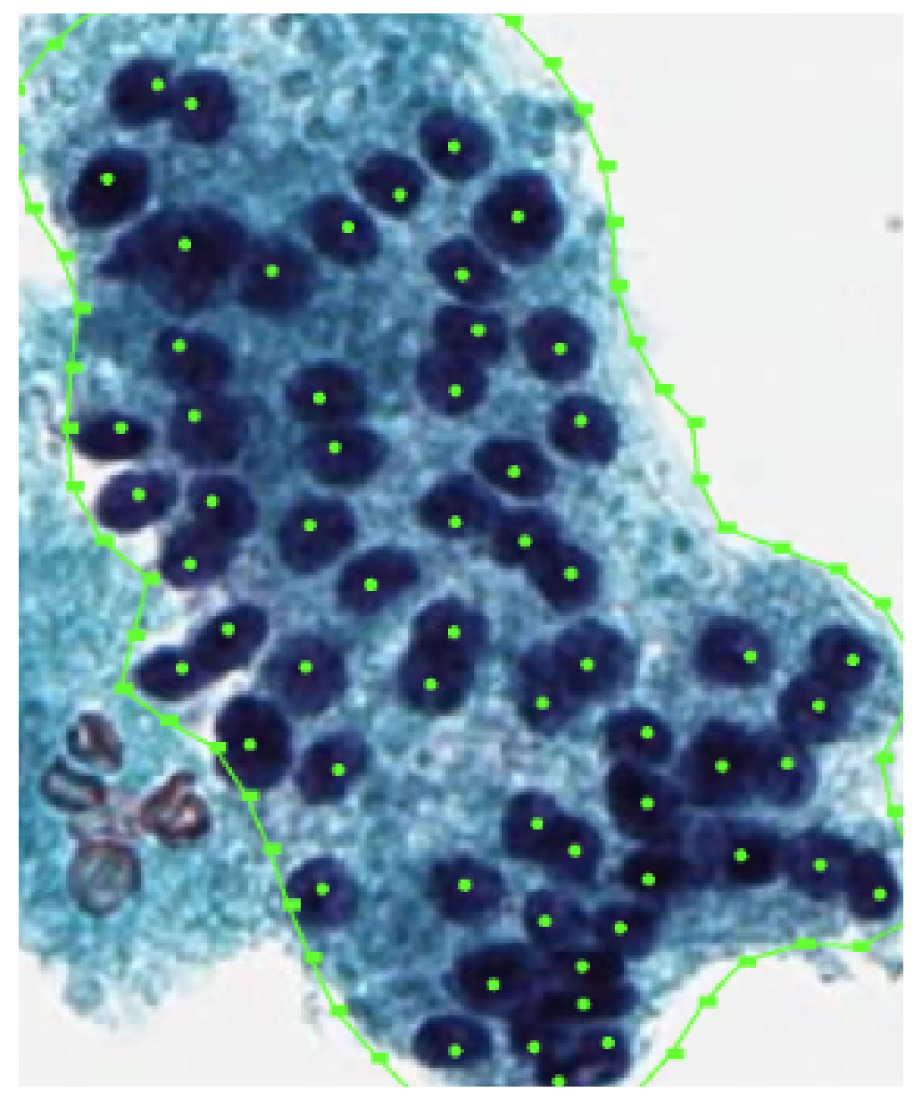
Suspicious for Follicular Neoplasm (4)
- Surgical lobectomy
- ~3% relative frequency
- 15-30% ROM

Suspicious for Malignancy (5)
- Near-total thyroidectomy or lobectomy
- ~3% relative frequency
- 60-75% ROM

24% Malignant

76% Benign

The Bethesda System Classification

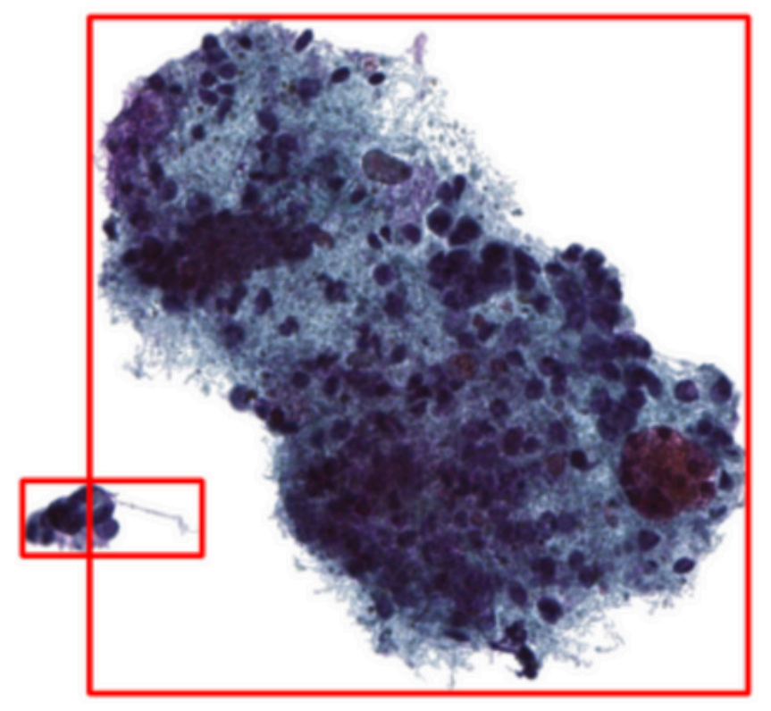

Dot annotations received in dataset

DATA COLLECTION

- All data used was collected from the **Dartmouth Hitchcock Medical Center's private dataset**
- 84 **Whole Slide Images** (WSIs) stained with **ThinPrep Pap Test**
- Each WSI consists of **clusters** and within clusters are **cells**
- Data received had **dot annotations** for cells within clusters
- Models were trained on **1000 clusters** and **40,000 cells**
- **29 patients** had a diagnosis with a detailed diagnosis report, and were used for Patient Diagnosis portion of the project

CLUSTER ISOLATION

- **Connected Component Analysis** isolates potential clusters within Whole Slide Images(WSIs)
- YOLOv8 **cell segmentation** models ensure clusters have **follicular cells**, eliminating **dead space**

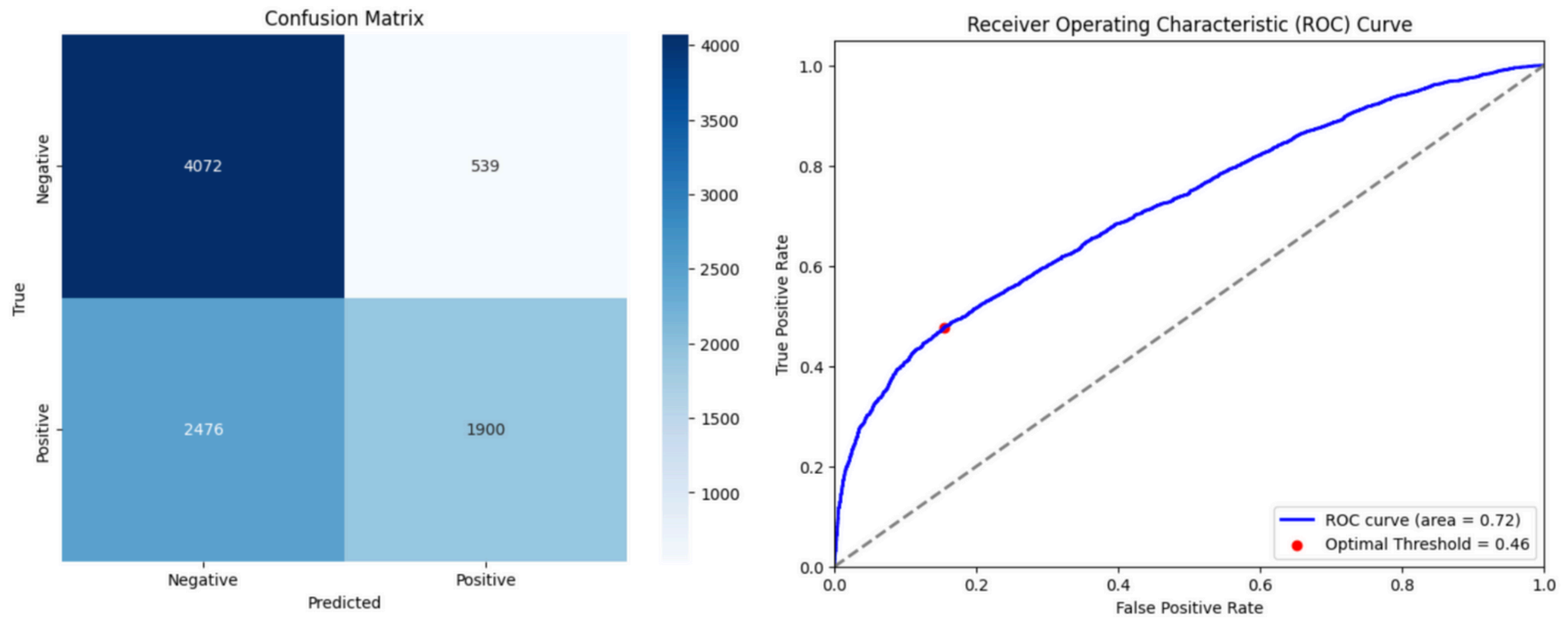

Connected Component Analysis(CCA)


Cell Segmentation models detecting follicular cells

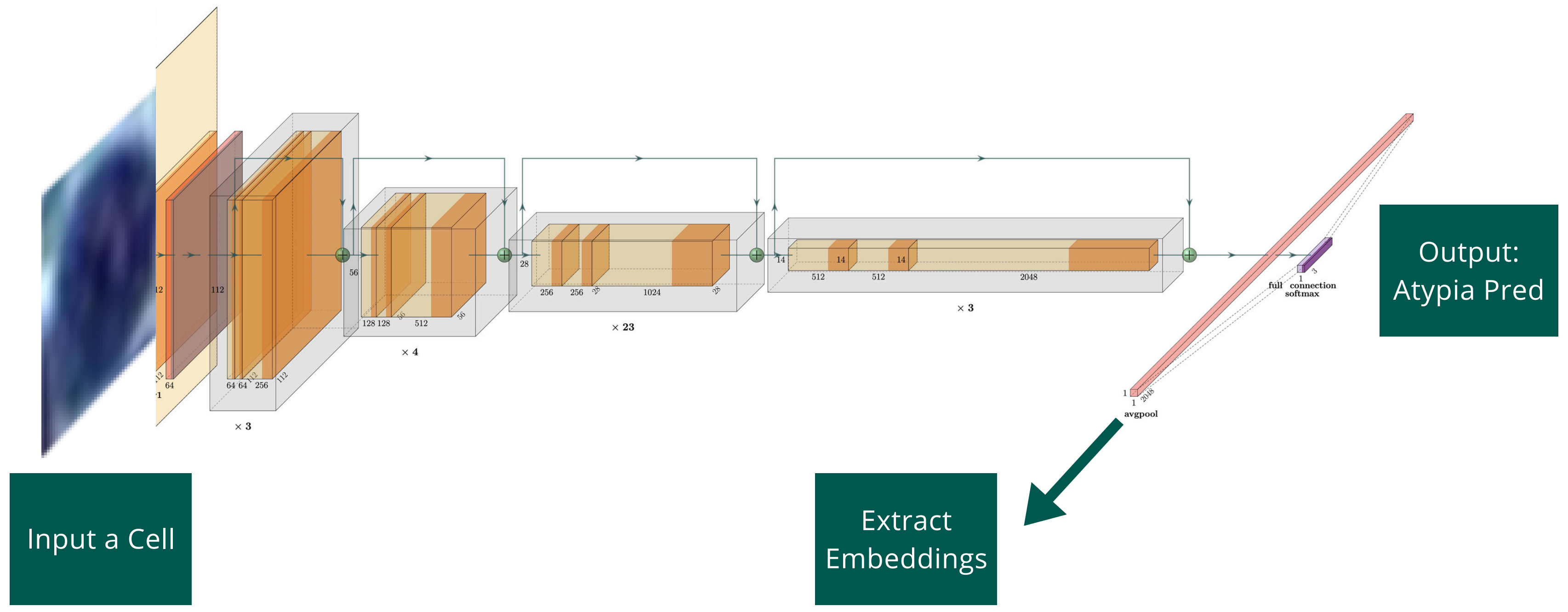

The dead space filtered out

CELL ATYPIA CLASSIFICATION

- **Custom PyTorch** Convolutional Neural Network (**CNN**) trained to identify **atypical vs. non-atypical cells**
- Achieved **66.7%** and **AUC Score** of **0.72**
- Only 3 out of 4 residual layers locked from training
- **Embeddings** of Cell Atypia CNN from **2D Pooling Layer** were **fed** into the **Graph Convolutional Network (GCN)**



Confusion matrix of atypical vs. non-atypical cells via the AtypiaCNN (LEFT) and ROC Curve (RIGHT)

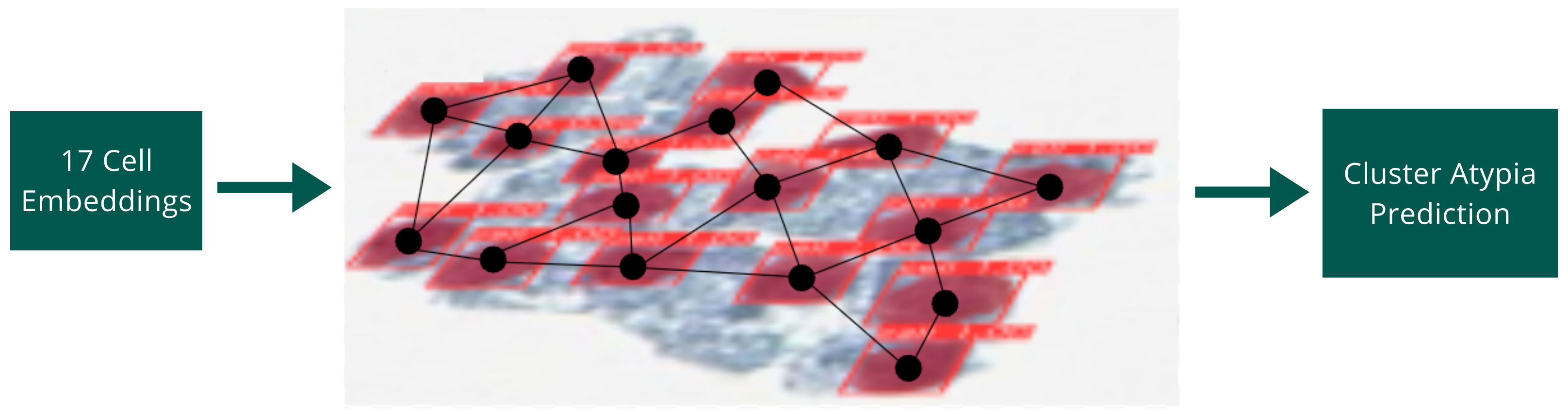

Output: Atypia Pred

Input a Cell

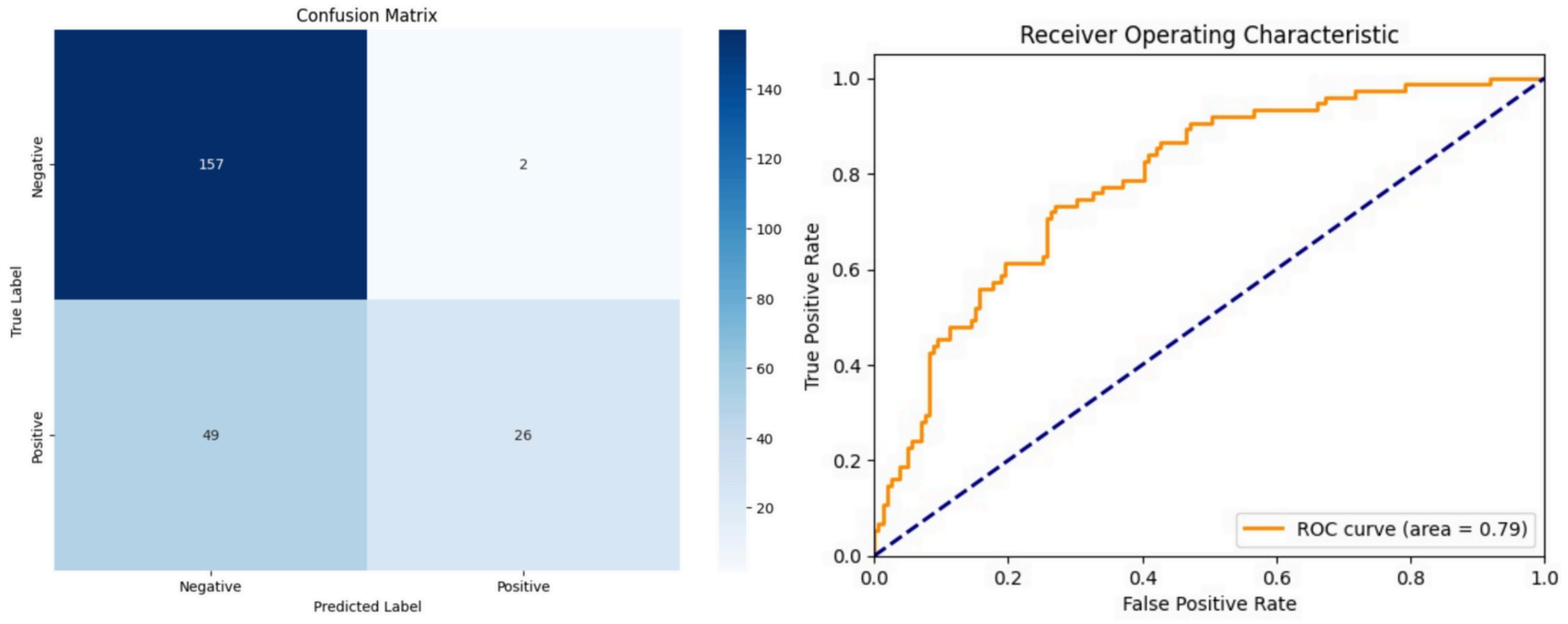
Extract Embeddings

GRAPH NEURAL NETWORK (GNN)

- **Embeddings** from the **Cell Atypia CNN** are **fed** into a Graph Neural Network (**GNN**) for **every cell** in a **cluster**
- Using **every cells' embedding**, the **GNN classifies** the **cluster** as **atypical** or **non-atypical**
- **CNN-GNN** are trained **end-to-end**
- **GNN** achieved an **accuracy** of **78.2% AUC Score** is **0.79**


17 Cell Embeddings

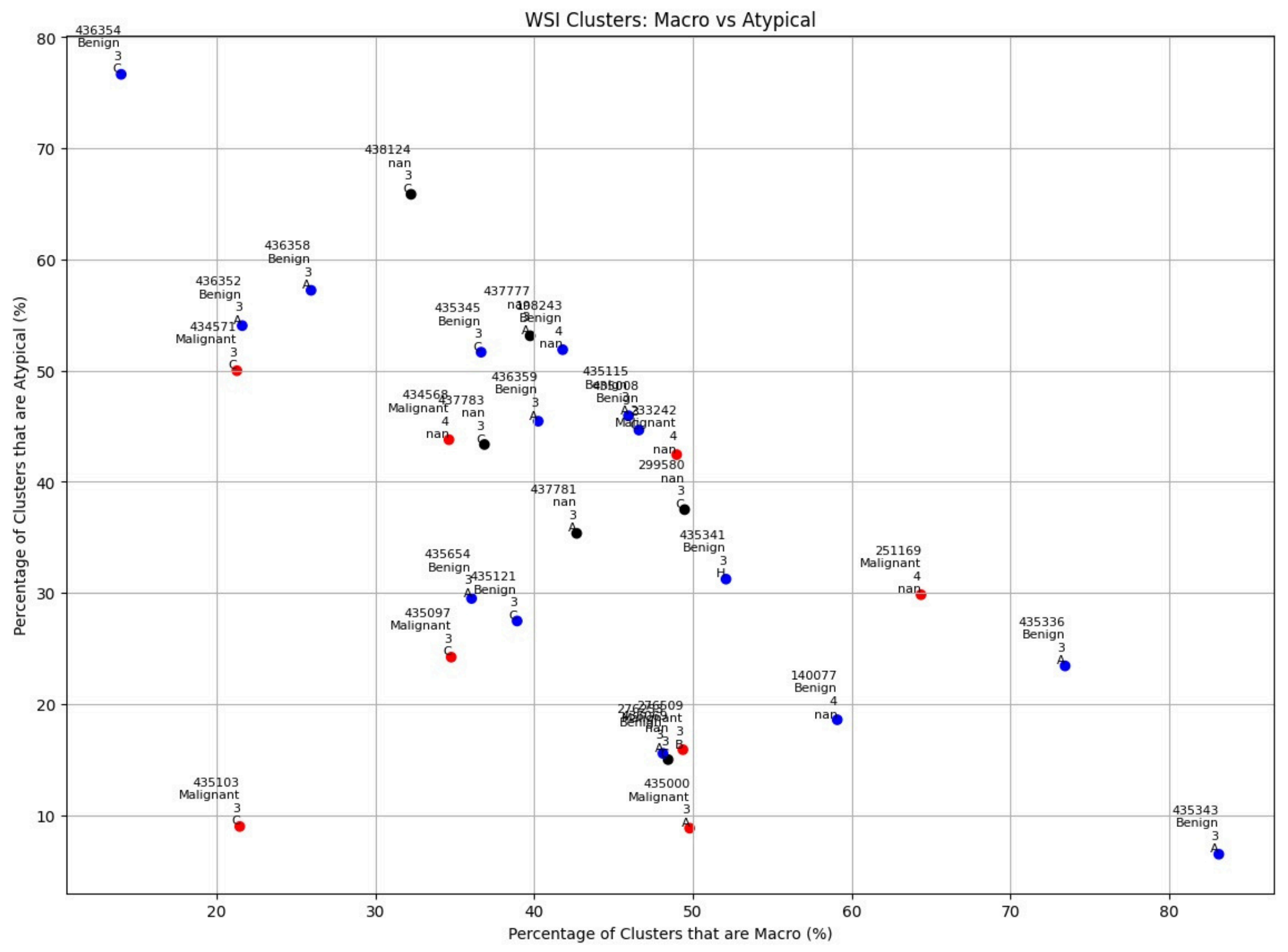
Cluster Atypia Prediction



A confusion matrix (LEFT) and ROC curve (RIGHT) pertaining to the prediction of atypical clusters (positive) vs. non-atypical clusters (negative) by the CNN-GCN model

PATIENT DIAGNOSIS

- Used **CNN-GNN** and **ArchCNN** to find the **classification of clusters**
- **Plotted percentage** of each **cluster** found in WSI on a **2D plane**, where **each axis** represents a **cluster type**
- Each **dot** represents a **WSI/patient** who were Bethesda Categories 3 & 4
- **Malignant: Red --- Benign: Blue**



2D Plot of the percentage of each cluster type shown for each WSI

RESULTS

- **YOLOv8** segmented cells with scores of **mAP50:** 0.8551, **mAP50-95:** 0.50695, **Precision:** 0.7377, **Recall:** 0.8733
- **AtypiaCNN** predicts **cell-nuclear atypia** with **66.7% accuracy** and **0.72 AUC Score**
- The **cell CNN embeddings** were **fed into** a **GNN** for **cluster atypia** predictions and resulted in an **78.2% accuracy** and **AUC Score** of **0.79**
- The **ArchCNN** classified with an **88.1% accuracy** and **AUC Score** of **0.894**
- The **RFC** predicted **microfollicular vs. crowded** with **83.6% accuracy** and **AUC Score** of **0.881**
- The **2D Plot** for a **patient/WSI** visually **didn't have a correlation** with any **cluster types**

CONCLUSION

- Many **benign indeterminant cases** can potentially be **accurately identified** by this model, reducing **unnecessary surgeries**
- **Connected Component Analysis** can be **improved** to ensure that **only follicular cells** are **fed into** the **Patient Diagnosis**
- **Improve nuclear atypia** classification **CNN-GNN model** to have a **higher accuracy**
- **More patient data** must be acquired to be able to effectively train and validate patient level results
- Cluster atypia **GNN** and Architectural atypia **CNNs** can **highlight clusters of interest** to **aid pathologists** with **complex and time-consuming cases**
- **Model's** workflow is **similar** to a **pathologist's workflow**, providing more **explainability to the results** than conventional approaches, meaning **pathologists** are **more likely to trust this model's results**

References: https://tinyurl.com/Thyroid-Carcinoma-EDIT-Project