Team 8

Efficient X-Ray Representations For Classifying Diseases

UBC Medicine Datathon 2025

BACKGROUND

Chest X-rays are a primary diagnostic tool for identifying thoracic diseases.

KEY CHALLENGES

Data Scale

Usable representation

Annotation Constraints

OBJECTIVES

To create an **efficient** and **accurate** representation of x-ray images using a combination of feature extraction and dimensionality reduction

PIPELINE

Feature Extraction

Dimensionality Reduction

Classification

Average 3.64
X-Ray images/patient
(112GB)

PyTorch

EfficientNet (CNN)

Pinecone

112k x
2048
features
(2.7 GB)

PCA

AE

300
Features
(0.3GB)

XGBoost

SVC

RF

Model (RF) Prediction

Label

Predicted
label

Disease labels

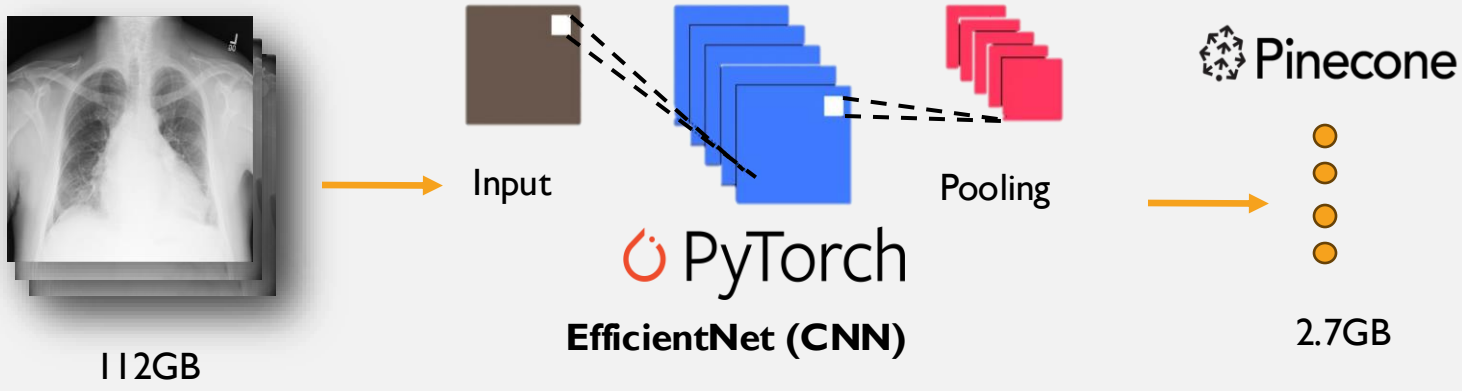
One-hot
encoding

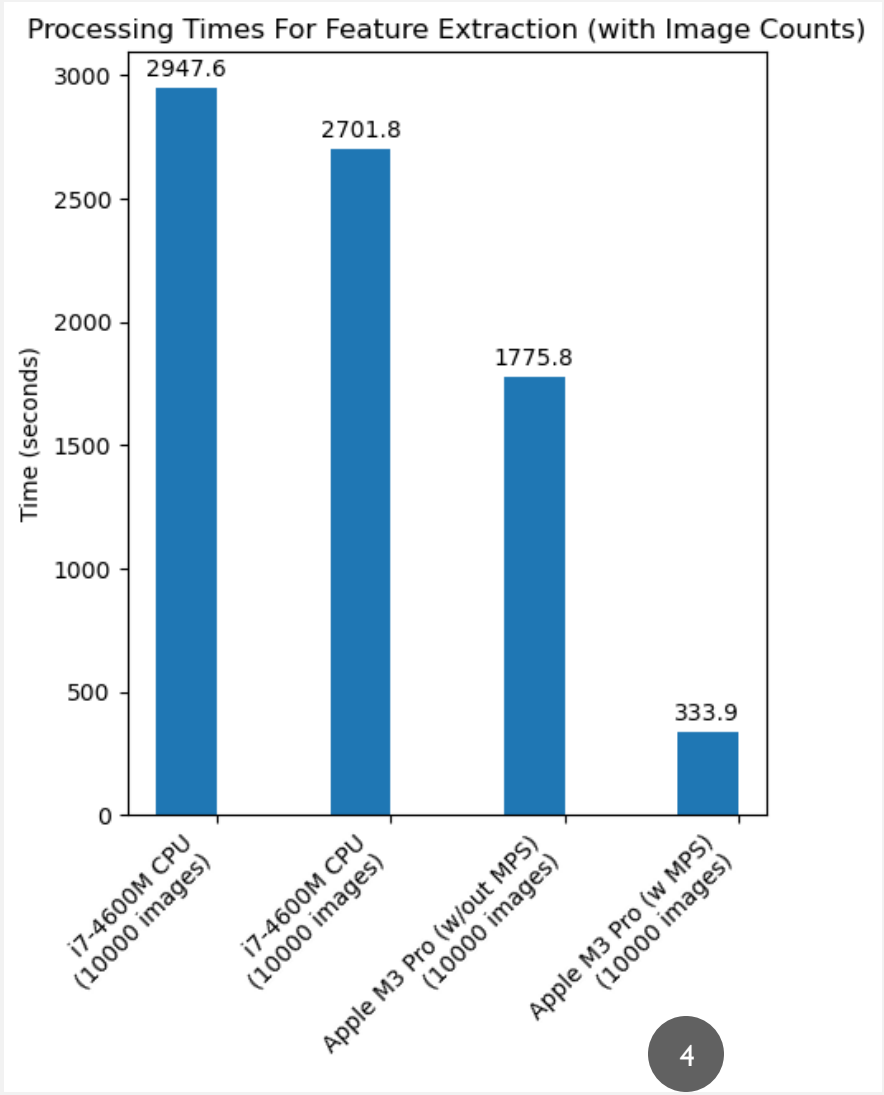
FEATURE EXTRACTION

Challenge

Data Scale: 112GB → 2.7GB

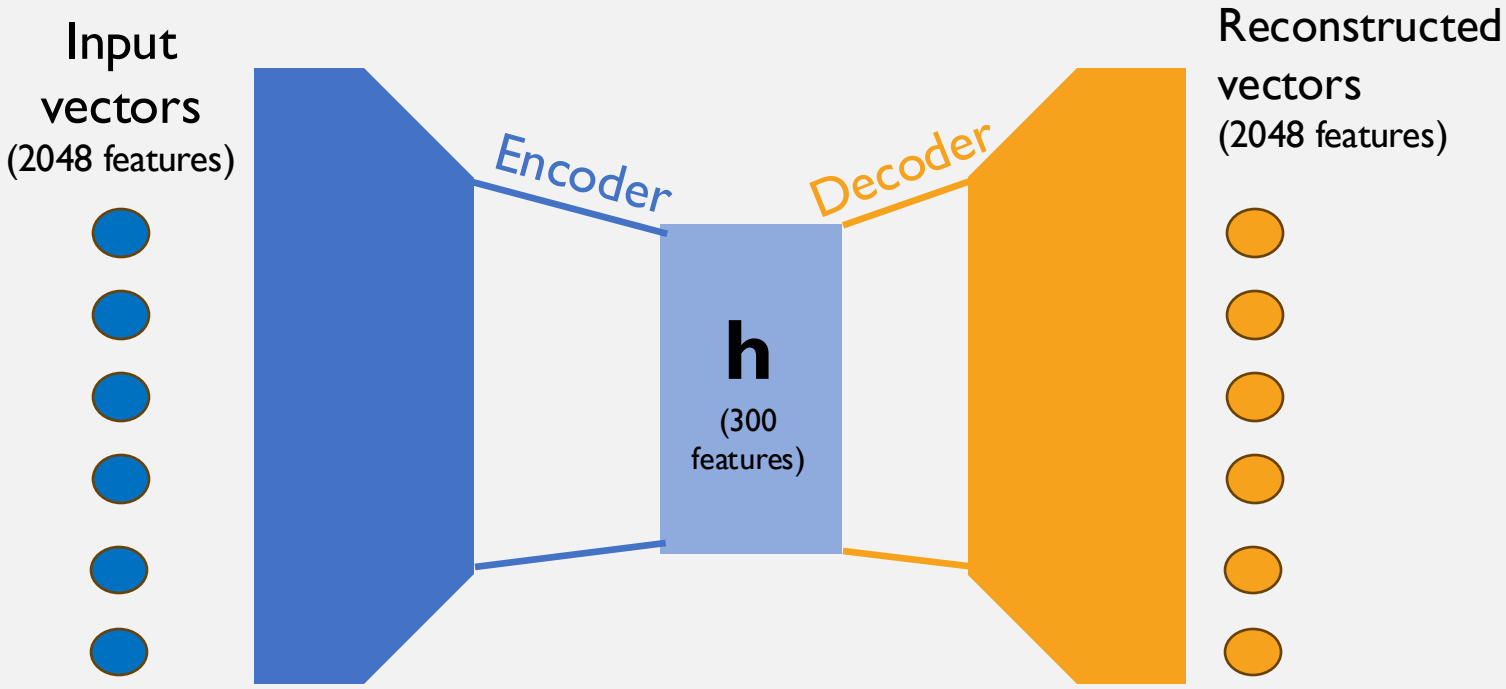
Reduction: 40x

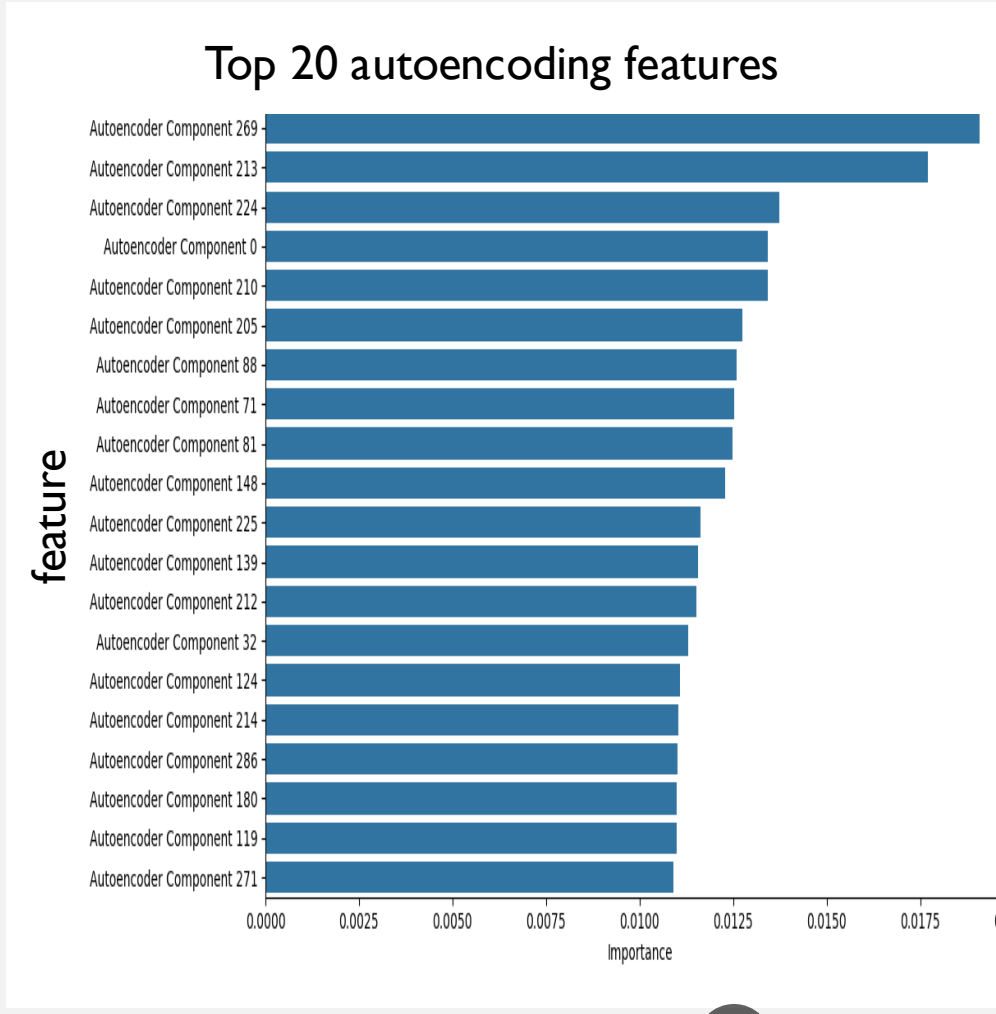


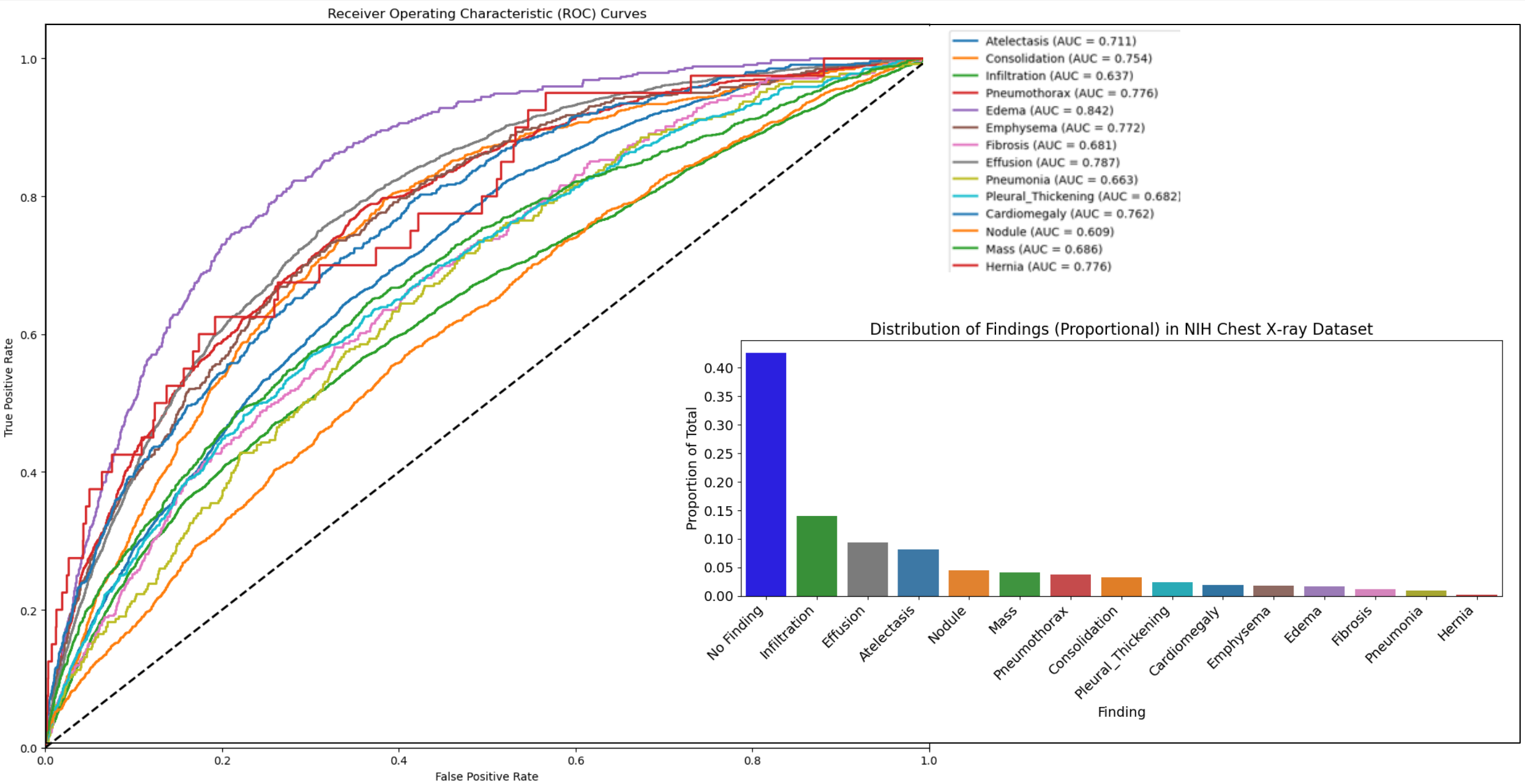


DIMENSIONALITY REDUCTION

Challenges: Usable Representation



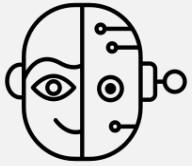



Receiver Operating Characteristic (ROC) Curves
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Atelectasis (AUC = 0.711)
Consolidation (AUC = 0.754)
Infiltration (AUC = 0.637)
Pneumothorax (AUC = 0.776)
Edema (AUC = 0.842)
Emphysema (AUC = 0.772)
Fibrosis (AUC = 0.681)
Effusion (AUC = 0.787)
Pneumonia (AUC = 0.663)
Pleural_Thickening (AUC = 0.682)
Cardiomegaly (AUC = 0.762)
Nodule (AUC = 0.609)
Mass (AUC = 0.686)
Hernia (AUC = 0.776)
Distribution of Findings (Proportional) in NIH Chest X-ray Dataset
Proportion of Total
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
No Finding
Infiltration
Effusion
Atelectasis
Nodule
Mass
Pneumothorax
Consolidation
Pleural_Thickening
Cardiomegaly
Emphysema
Edema
Fibrosis
Pneumonia
Hernia
Finding

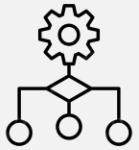
IMPLICATIONS

Disclaimer: AIs are not doctors nor do they play them on TV.

However, there needs to be studies on the complex of human-robot interaction.



Using algorithms to **assist** preliminary decisions

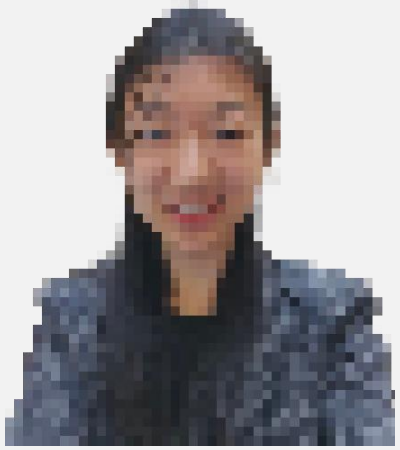


Like any other model, the model is vulnerable to concept drift.

TEAM

Ethan Rajkumar

Joel Bonnie



Charity Grey

Pushya Jain

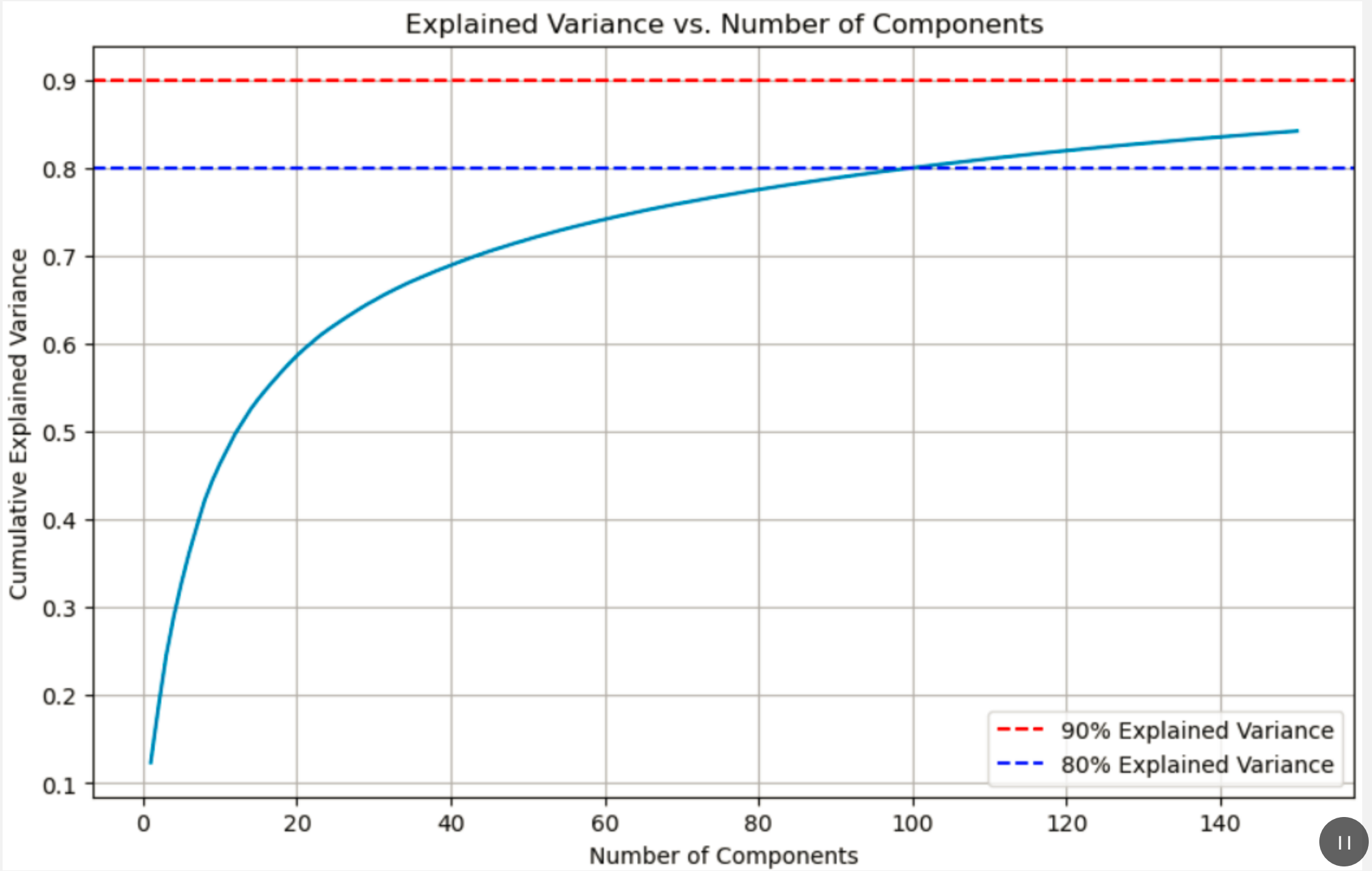
Vivaan Jhaveri

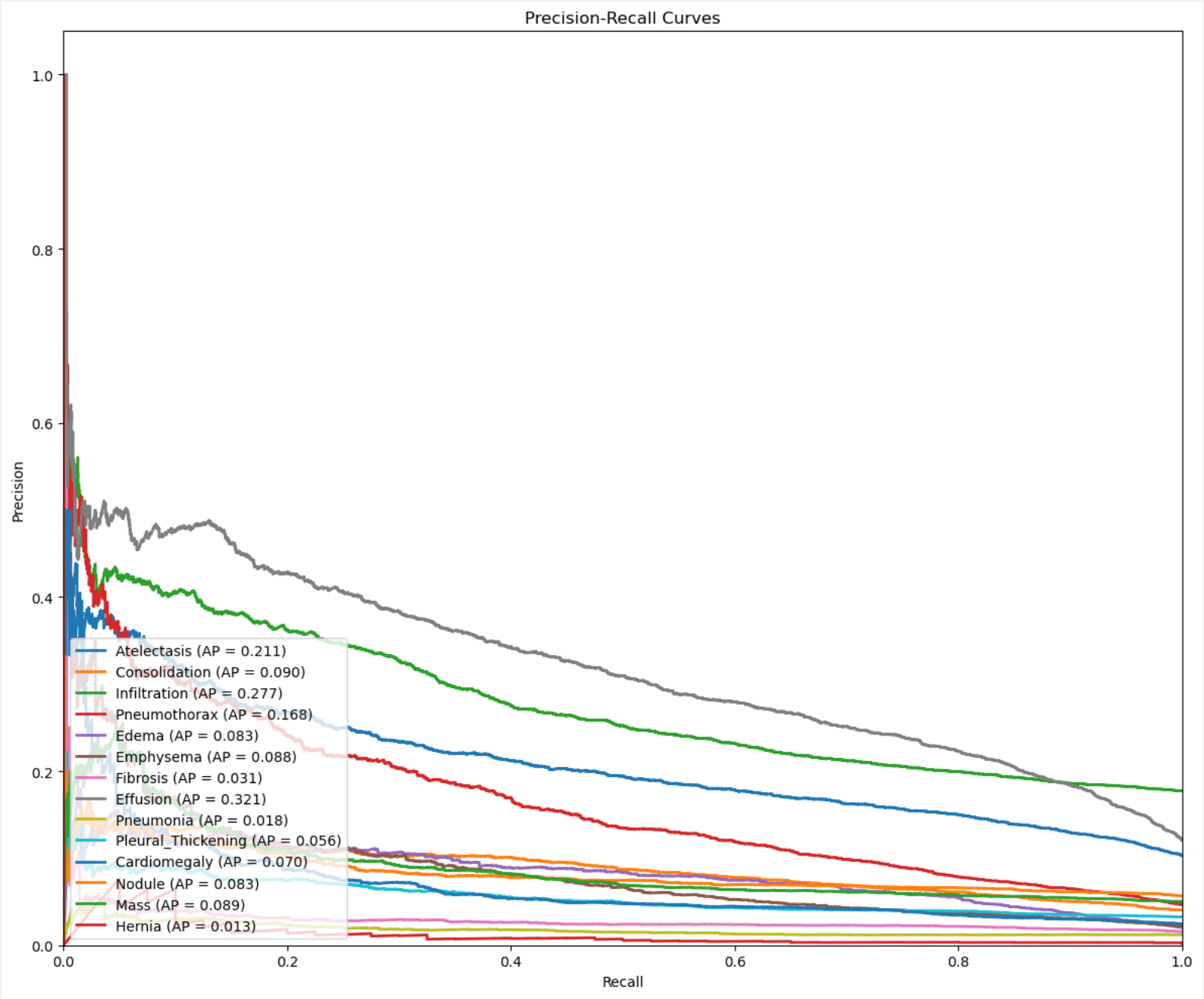
Erhan Javed

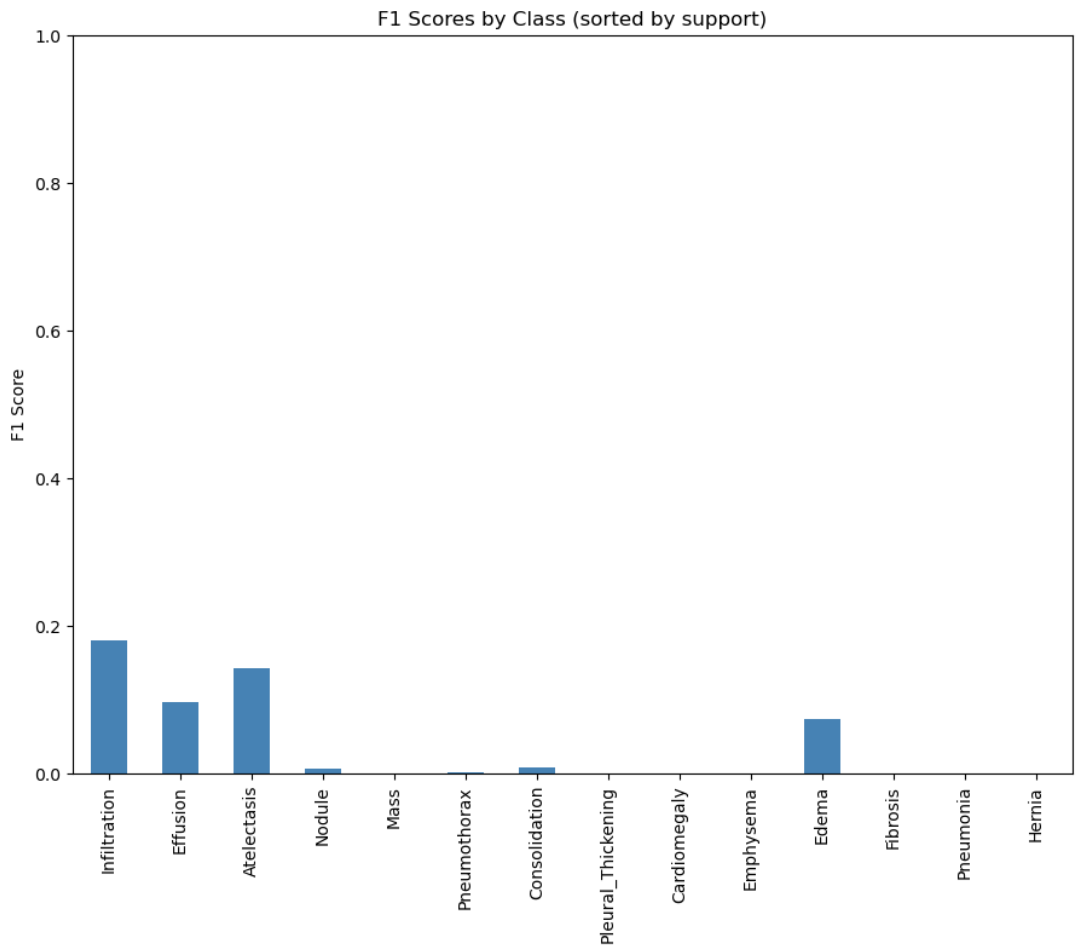
REFERENCES

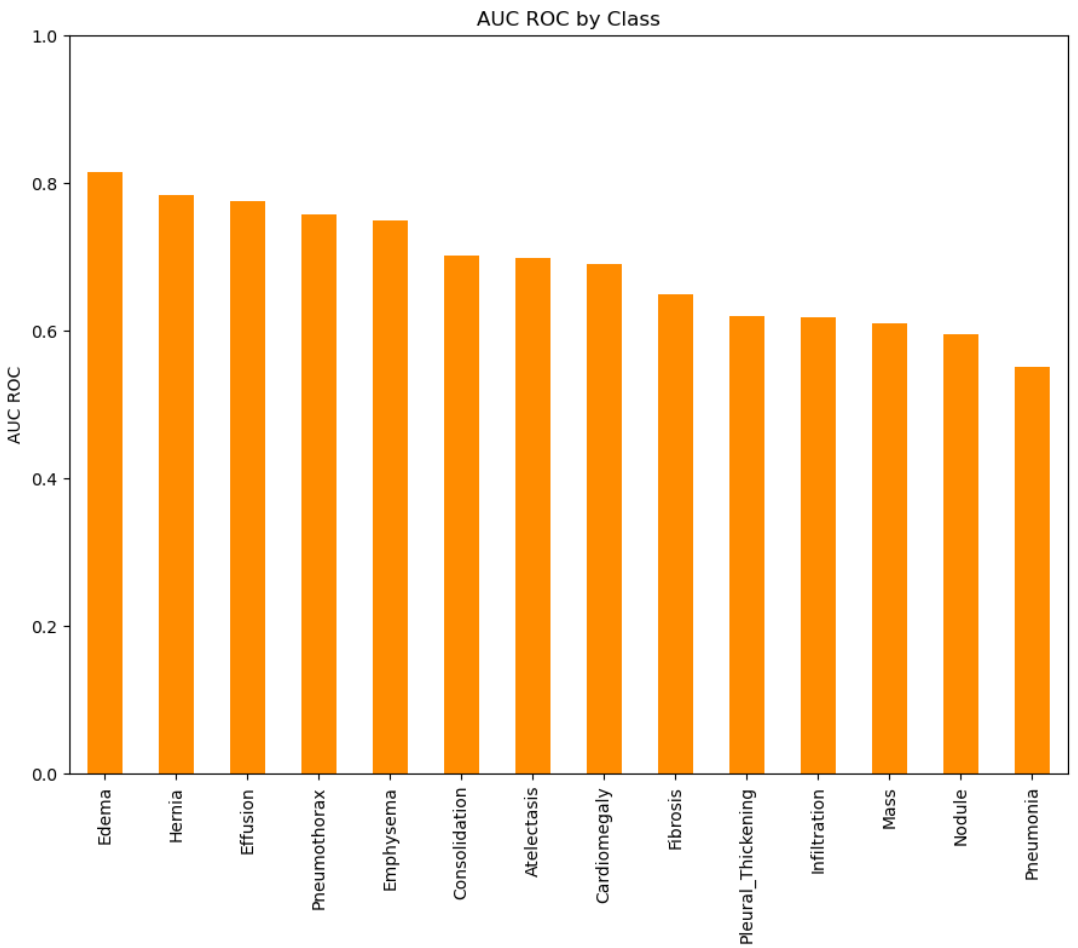
- Wang, Xiaosong, et al. "Chestx-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 3462–3471, https://doi.org/10.1109/cvpr.2017.369.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

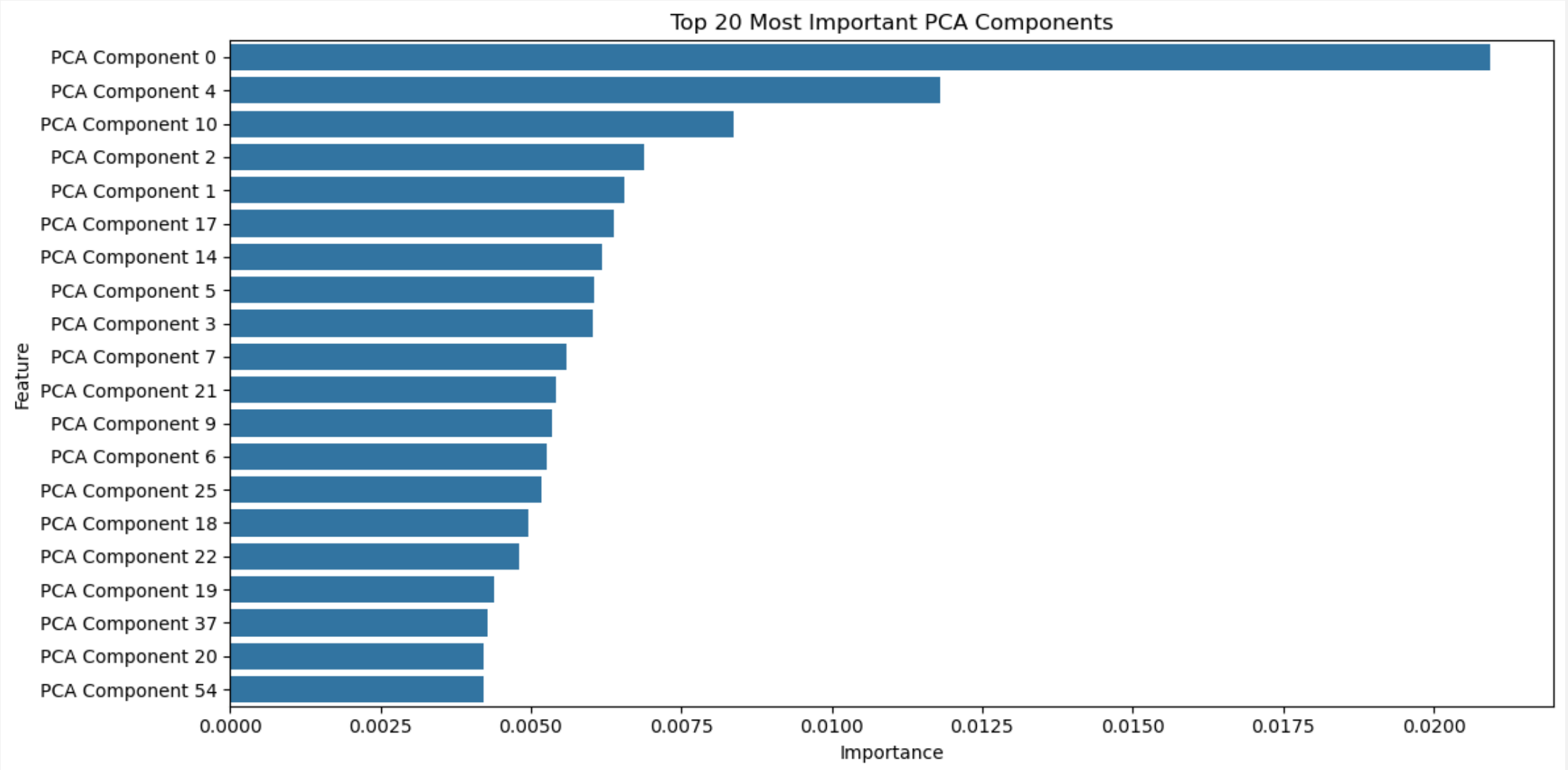
APPENDIX

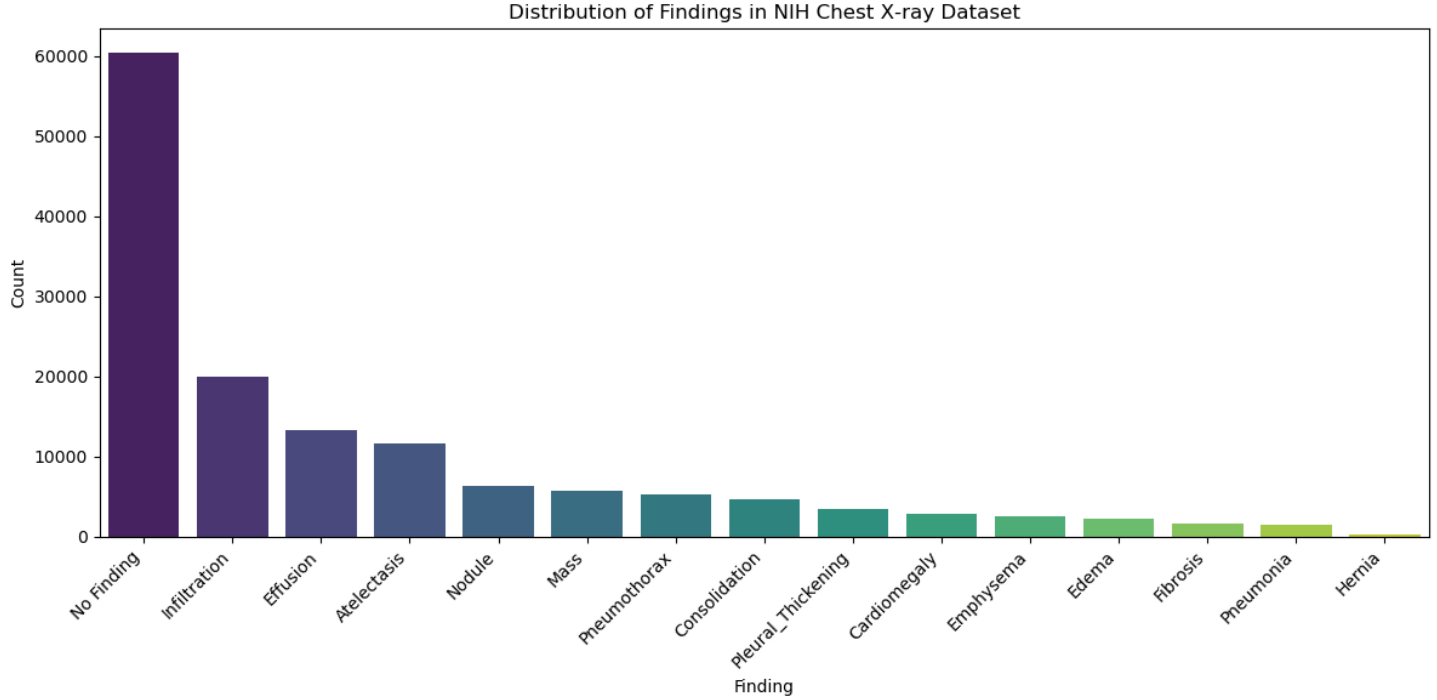

Explained Variance vs. Number of Components
Cumulative Explained Variance
Number of Components
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100
120
140
90% Explained Variance
80% Explained Variance

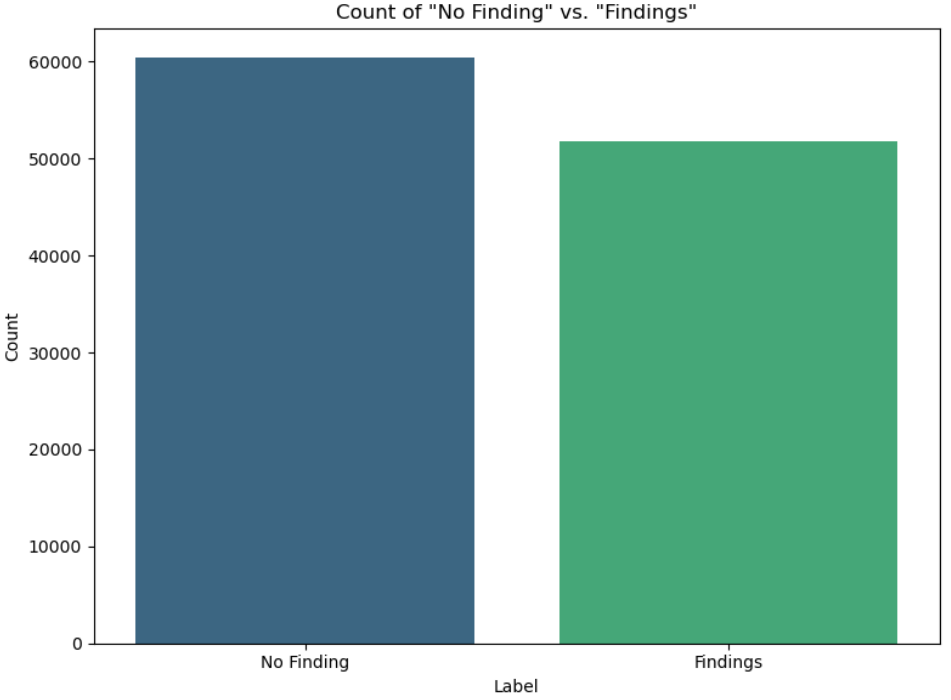

Precision-Recall Curves
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Atelectasis (AP = 0.211)
Consolidation (AP = 0.090)
Infiltration (AP = 0.277)
Pneumothorax (AP = 0.168)
Edema (AP = 0.083)
Emphysema (AP = 0.088)
Fibrosis (AP = 0.031)
Effusion (AP = 0.321)
Pneumonia (AP = 0.018)
Pleural_Thickening (AP = 0.056)
Cardiomegaly (AP = 0.070)
Nodule (AP = 0.083)
Mass (AP = 0.089)
Hernia (AP = 0.013)


F1 Scores by Class (sorted by support)
F1 Score
1.0
0.8
0.6
0.4
0.2
0.0
Infiltration
Effusion
Atelectasis
Nodule
Mass
Pneumothorax
Consolidation
Pleural_Thickening
Cardiomegaly
Emphysema
Edema
Fibrosis
Pneumonia
Hernia


AUC ROC by Class
AUC ROC
1.0
0.8
0.6
0.4
0.2
0.0
Edema
Hernia
Effusion
Pneumothorax
Emphysema
Consolidation
Atelectasis
Cardiomegaly
Fibrosis
Pleural_Thickening
Infiltration
Mass
Nodule
Pneumonia


Top 20 Most Important PCA Components
PCA Component 0
PCA Component 4
PCA Component 10
PCA Component 2
PCA Component 1
PCA Component 17
PCA Component 14
PCA Component 5
PCA Component 3
PCA Component 7
PCA Component 21
PCA Component 9
PCA Component 6
PCA Component 25
PCA Component 18
PCA Component 22
PCA Component 19
PCA Component 37
PCA Component 20
PCA Component 54
Feature
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
Importance


Distribution of Findings in NIH Chest X-ray Dataset
Count
60000
50000
40000
30000
20000
10000
0
No Finding
Infiltration
Effusion
Atelectasis
Nodule
Mass
Pneumothorax
Consolidation
Pleural_Thickening
Cardiomegaly
Emphysema
Edema
Fibrosis
Pneumonia
Hernia
Finding


Count of "No Finding" vs. "Findings"
Count
60000
50000
40000
30000
20000
10000
0
No Finding
Findings
Label

RANDOM FOREST

SVC → bad ROC score < 50%
Gradient boosted → slow
Random Forest → best

RESULTS- TYPES OF DISEASE