# X-Ray Classification

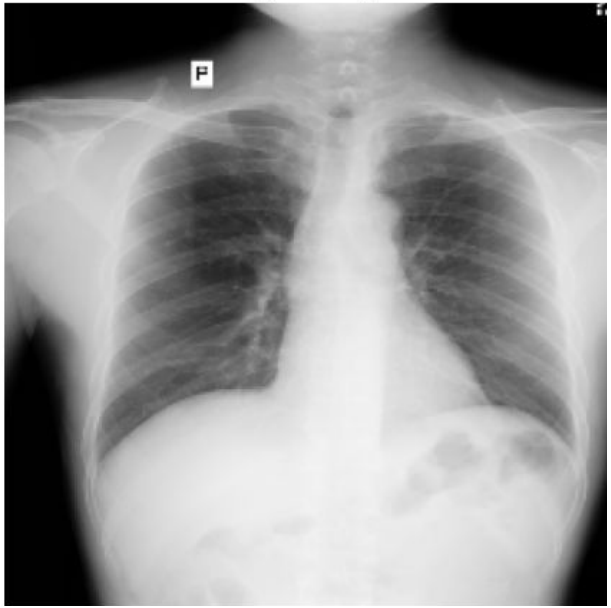
By Tony, Meric, Mohak, Alexander

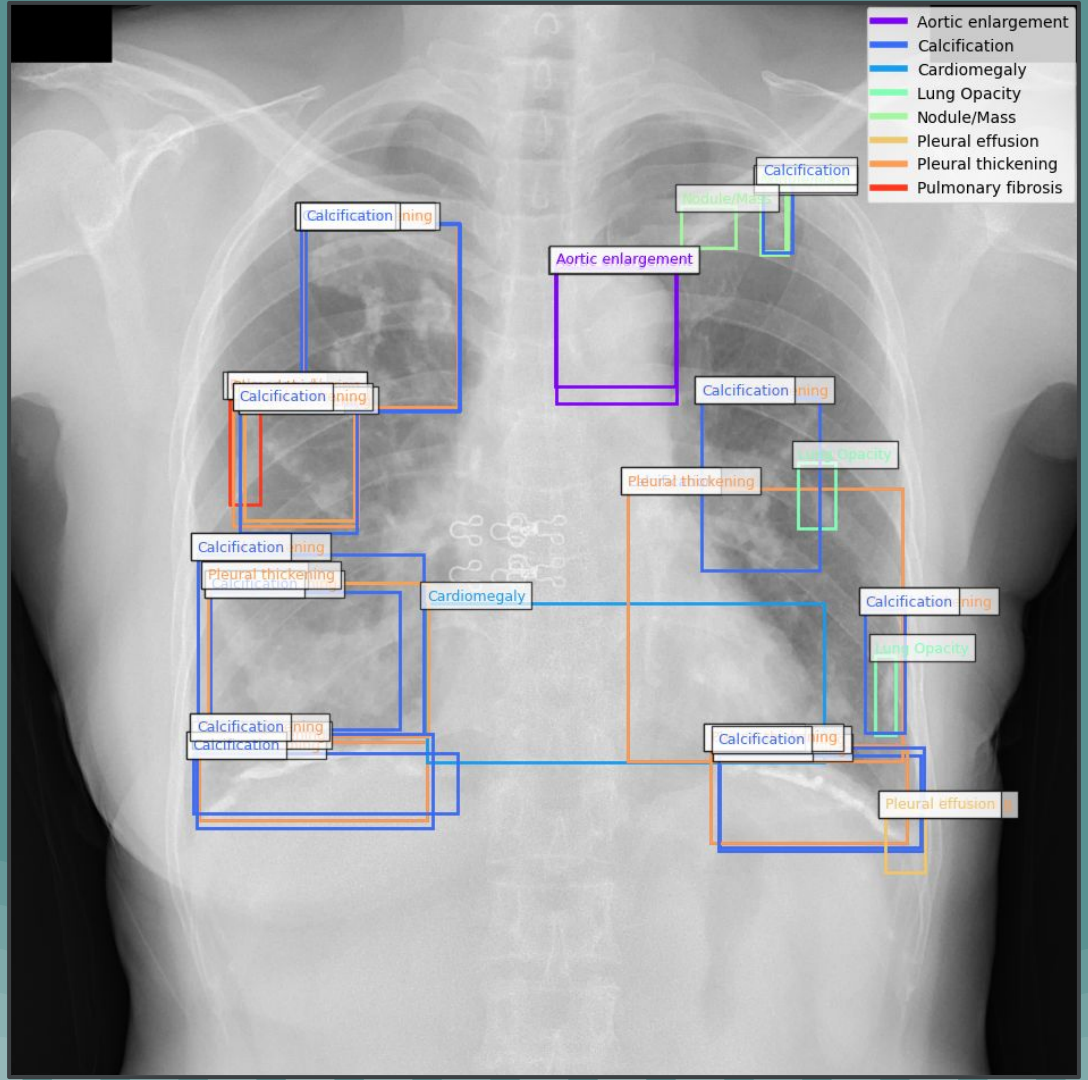
# Background

- Identify different heart/lung/chest conditions via X-rays.
- AI models can identify features that humans miss, especially for fuzzy or low quality images.
    - Reduce misdiagnosis.
- Assist less experienced doctors with diagnosing X-rays, reduce workload for experienced doctors.

# Dataset

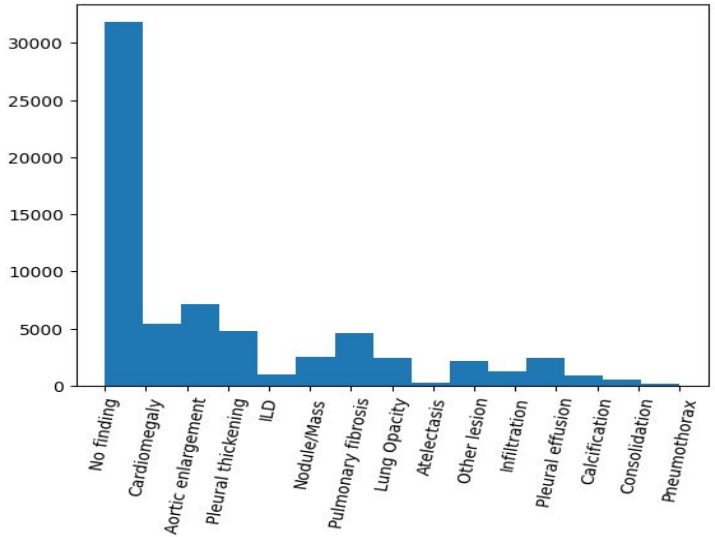
- Vingroup Big Data Institute Chest X-ray Abnormalities Detection from Kaggle
  - Image ID, condition name/id, radiologist id, box coordinates, image dimensions
- 15000+ images
  - Two Vietnamese hospitals (2018–2020)
  - 50000+ annotations by almost 20 radiologists
- 14 medical conditions (15 classes), and instances of co-occurrences within the same image



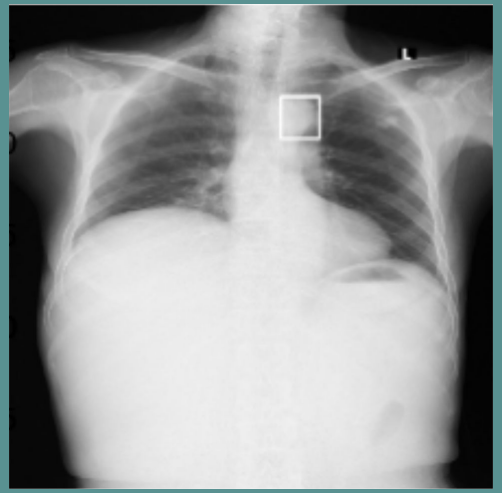

Aortic enlargement
Calcification
Cardiomegaly
Lung Opacity
Nodule/Mass
Pleural effusion
Pleural thickening
Pulmonary fibrosis
Calcification
Aortic enlargement
Cardiomegaly
Pleural effusion

# Preprocessing

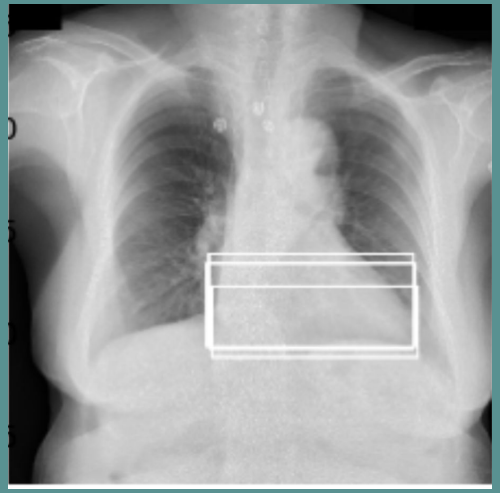
- Choose 5 most abundant classes
  - No Finding, Aortic Enlargement, Pulmonary Fibrosis, Cardiomegaly, Pleural Effusion
- There are ~1000 instances of each class
  - Split in 60:20:20 ratio for train/validation/test
  - Classes within split sets are about the same



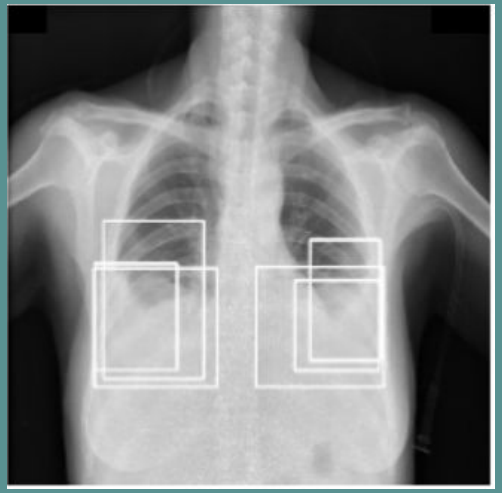
Aortic Enlargement



Cardiomegaly



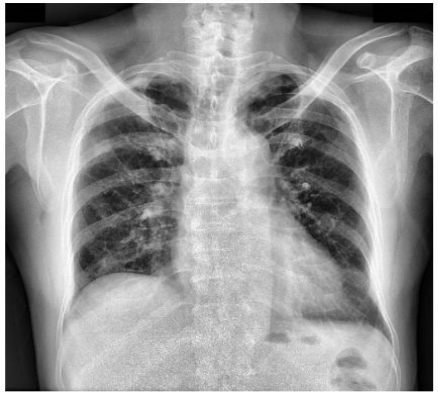
Pleural Effusion



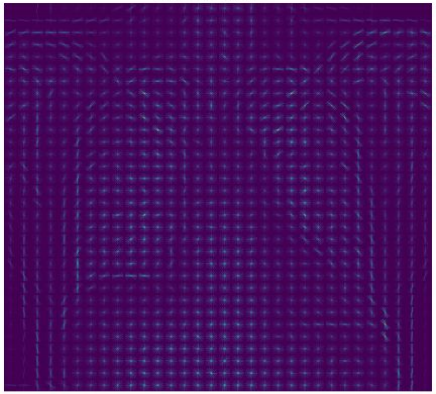
# Features

| | |
|---|---|
| Histogram Of Oriented Gradients (HOG) | - Captures localized orientation patterns and edge gradients.<br>- Highly useful for detecting structural changes in the lungs such as fibrosis and opacities |
| Fourier Transform | - Highlights global frequency structures<br>- Helpful in identifying patterns in texture and calcifications across the entire X-Ray |
| Local Binary Pattern | - Capture the subtle inconsistencies in texture caused by these diseases<br>- Useful for identifying diffuse disease patterns like interstitial disease or pleural thickening |
| Spatial Features | - Extracts positional relationships between pathologies (distances, angles, overlaps) as numerical values<br>- These spatial measurements are incorporated as additional features in our models to improve classification accuracy |

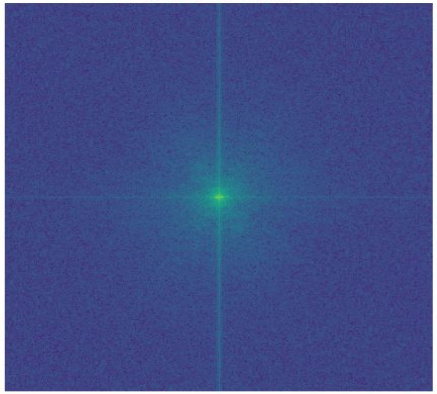
Original Image



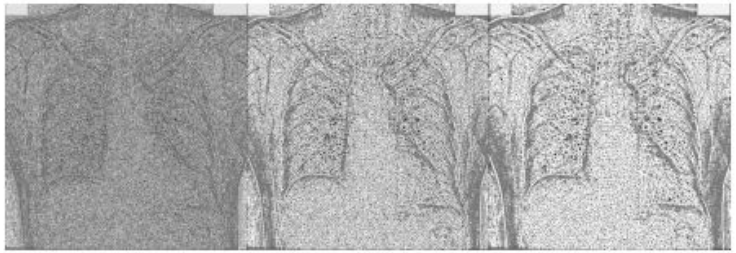
HOG Features



Fourier Spectrum



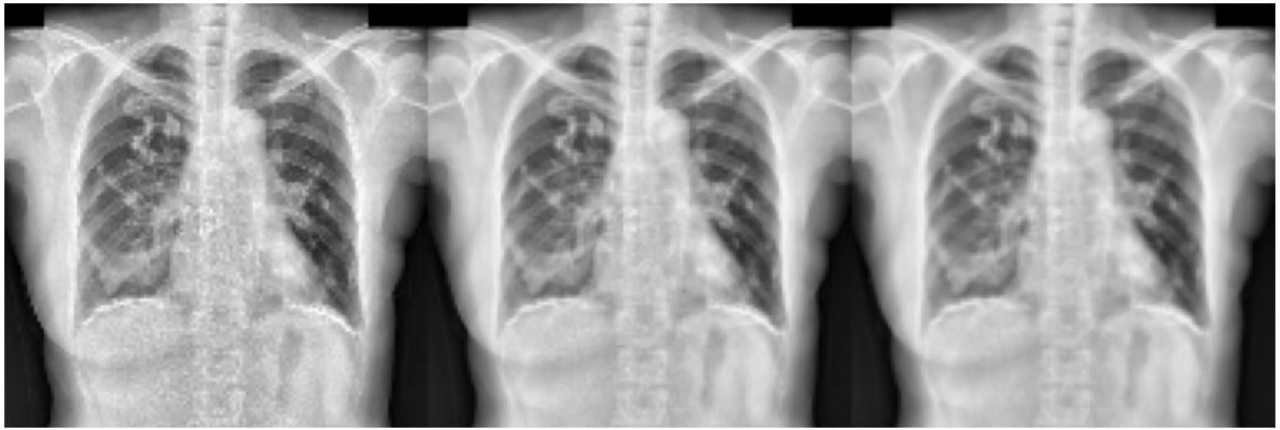
Local Binary Pattern



# Feature - Image Pyramids

| | |
|---|---|
| **Image Pyramid Decomposition** | • Builds multiscale representations of images using Gaussian pyramids<br>• Preserve both fine and coarse structural features across different scales |

Image Pyramid



# Feature - Transfer Learning Models

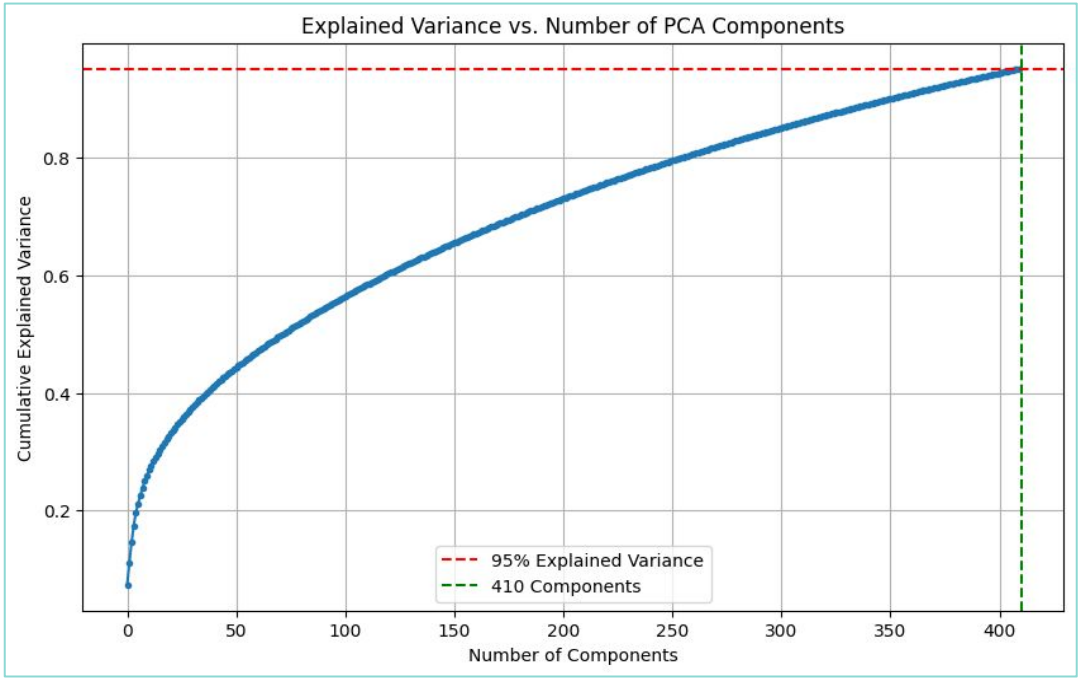
Our project utilizes **pre-trained convolutional neural networks** as complex features:

- **DenseNet121**: Dense connectivity pattern with 121 layers
  - Each layer connects to every other layer in a feed-forward fashion
  - Effective feature reuse and reduced parameter count
- **ResNet50**: Deep residual network with 50 layers
  - Skip connections help solve vanishing gradient problem
  - Enables training of much deeper networks
- **EfficientNetB0**: Optimized network architecture
  - Balanced depth, width, and resolution using compound scaling
  - Achieves state-of-the-art performance with fewer parameters

These models, pre-trained on ImageNet (1.2M images), capture hierarchical visual features from edges and textures to complex anatomical structures relevant for medical image classification.

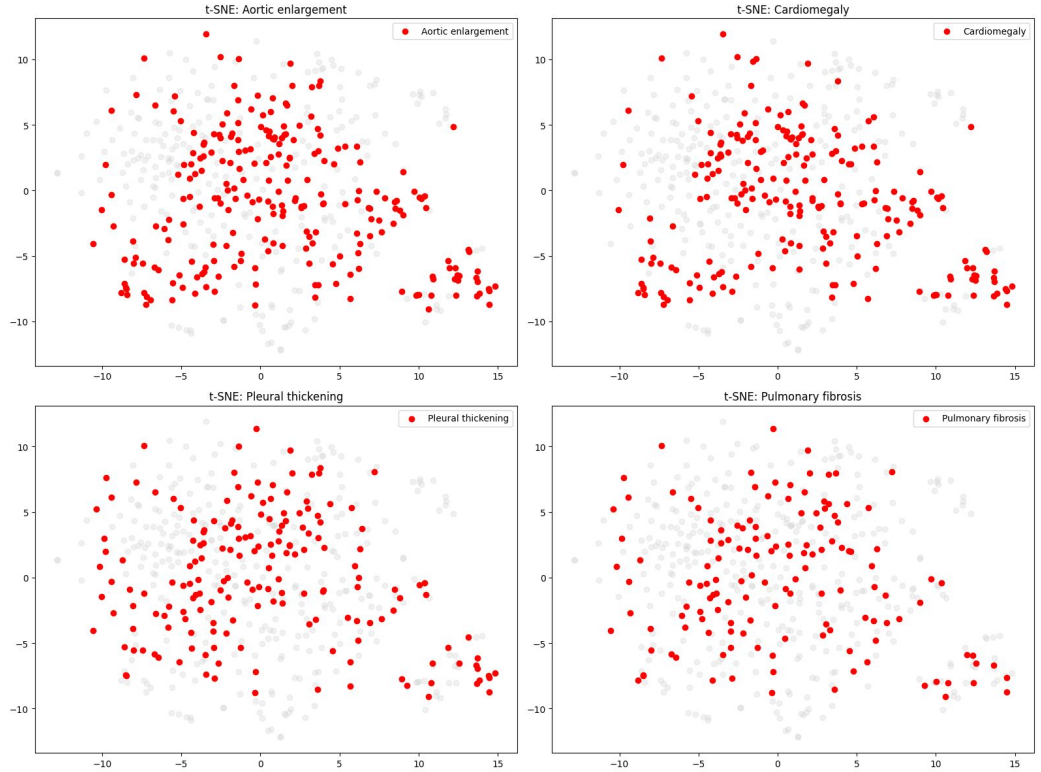
We fine-tune these networks using our chest X-ray dataset while preserving their ability to recognize generalizable visual patterns.

# Dimensionality Reduction - PCA




- **Principal Component Analysis (PCA)** was used to retain 95% of the variance across samples.
- This PCA step was useful to reduce the dimensionality of the engineered features, making the logistic regression and SVM models more efficient and less prone to overfitting

# Dimensionality Reduction - t-SNE




Observations:

- We don't see clear, isolated clusters for any condition, which suggests these diseases have overlapping feature representations
- The similar distribution patterns across different conditions suggest potential correlations between these diseases

# Models

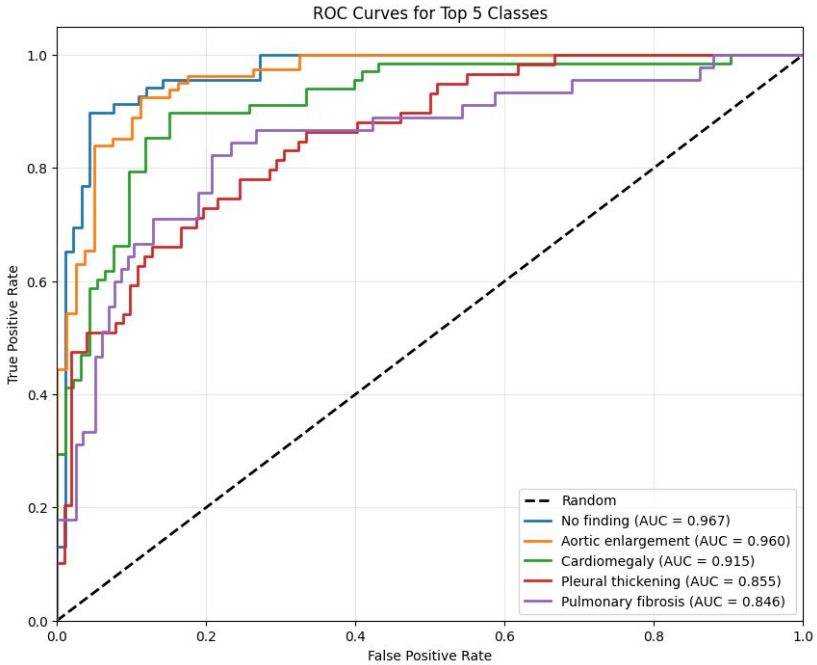
| | |
|---|---|
| Logistic Regression | • Uses binary one-vs-rest approach, training a separate model for each disease class<br>• Hyperparameter tuning is performed using GridSearchCV with cross-validation<br>• The model operates on PCA-reduced engineered features, which include HOG, Fourier, edge detection, LBP, pyramid, and spatial features |
| Support Vector Machine | • Uses binary one-vs-rest approach, training a separate model for each disease class<br>• GridSearchCV optimizes both C (regularization parameter) and gamma (kernel coefficient)<br>• Like the logistic regression model, it operates on PCA-reduced engineered features |
| Convolutional Neural Network | • Uses a transfer learning approach with systematic hyperparameter tuning via Keras Tuner (learning rate, dropout rate, #fine tuning layers, dense units in classification head)<br>• The base model is DenseNet121 (pre-trained on ImageNet), with fine-tuning of top layers<br>• uses a multi-label output with sigmoid activation for binary classification of each disease |

# Results- model x f1 score for each class

| Model | No Finding | Aortic Enlargement | Pleural Thickening | Pulmonary Fibrosis | Cardiomegaly | Total F1 |
|---|---|---|---|---|---|---|
| Logistic Regression | 0.888889 | 0.895706 | 0.698413 | 0.678571 | 0.845070 | 0.801 |
| SVM | 0.809917 | 0.864865 | 0.764706 | 0.624000 | 0.782609 | 0.769 |
| CNN | 0.413333 | 0.437086 | 0.385093 | 0.257426 | 0.512500 | 0.401 |

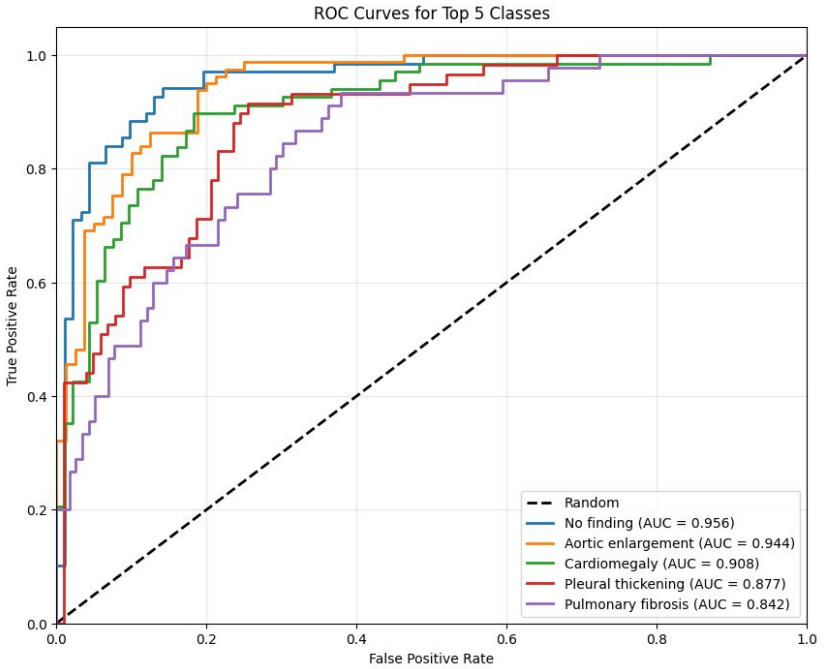
# Results -Logistic Regression

- "No finding" (AUC 0.967) and "Aortic enlargement" (AUC 0.960) show exceptional classification performance
- "Pleural thickening" and "Pulmonary fibrosis" show good but lower performance (AUC 0.855 and 0.846)
- Effectively capture key diagnostic patterns even with this simple model



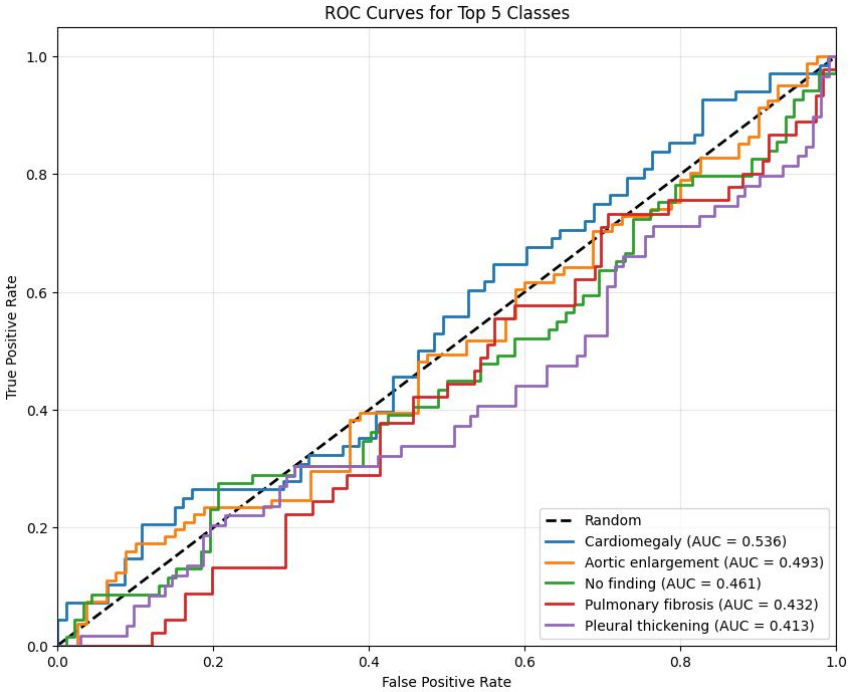
# Results - SVM

- "No finding" maintains strong performance (AUC 0.956), slightly lower than logistic regression
- SVM shows improved detection of "Pleural thickening" (AUC 0.877) compared to logistic regression
- Classification patterns remain consistent across models with same disease ranking

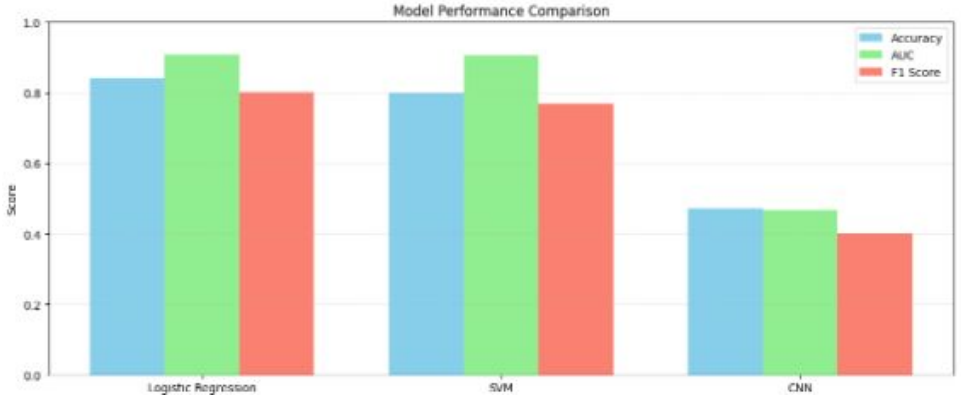


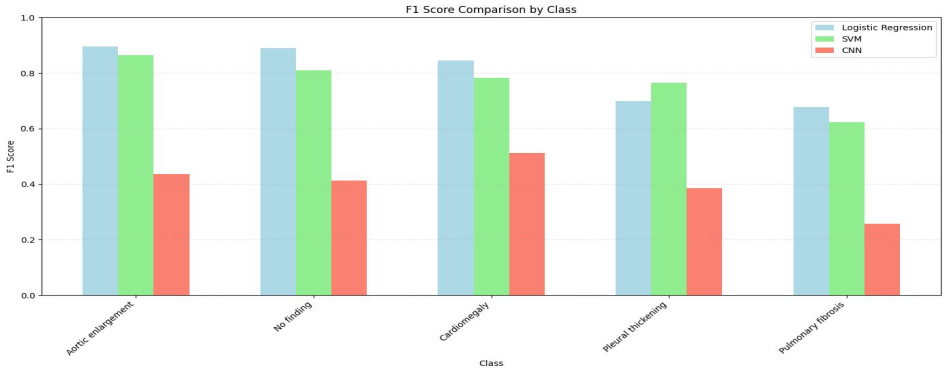
# Results - CNN

- All disease categories show poor classification performance (AUC values 0.413-0.536)
- Transfer learning from natural images (ImageNet) clearly failed to adapt to medical X-ray domain
- Curves hover near or below the random classification line, indicating ineffective learning
- "Cardiomegaly" achieves marginally better results (AUC 0.536) but still barely above random
- Suggests CNN model failed to capture relevant features or experienced training issues
- Dramatic performance drop compared to engineered feature approaches (LogReg, SVM)

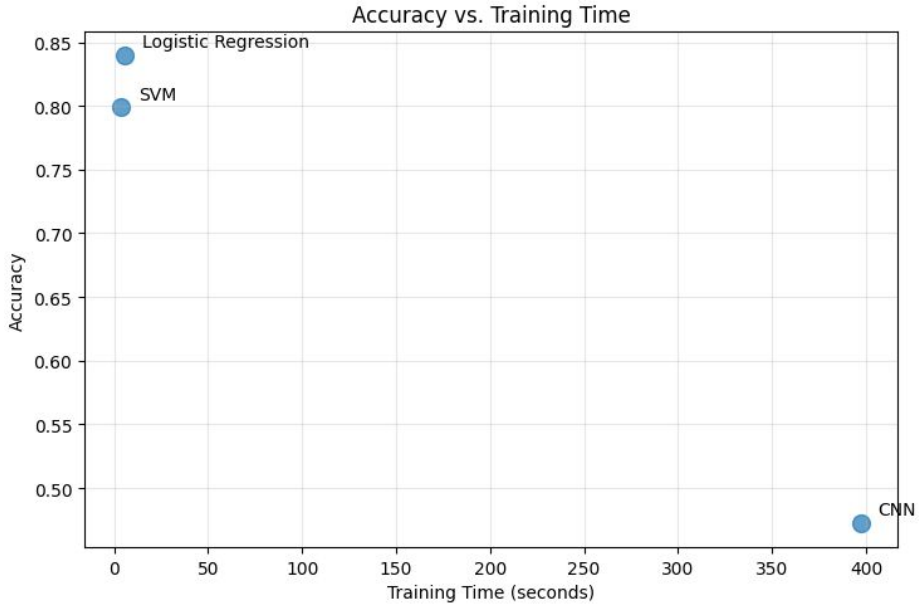


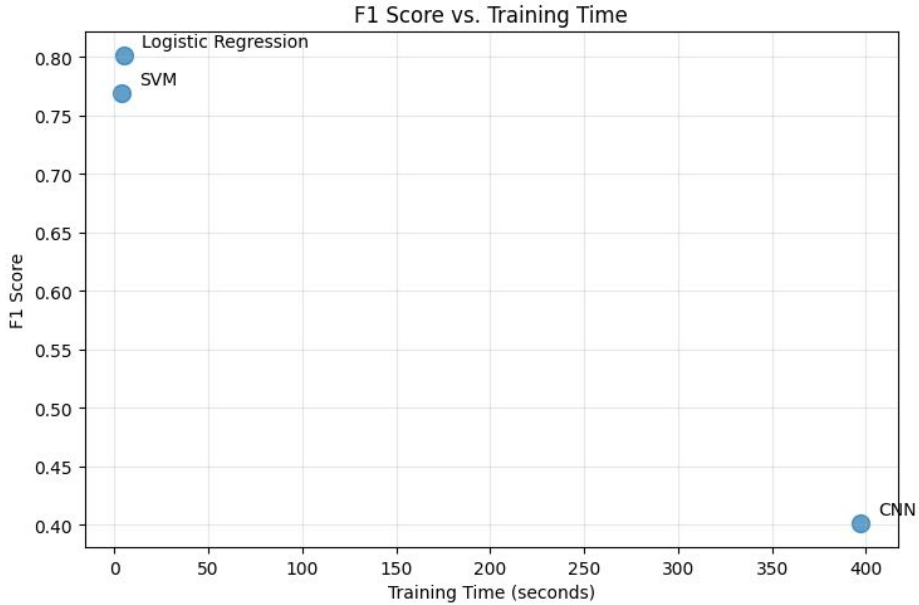
# Discussion - Model Comparison



# Discussion - Training



# Conclusion

- Compared 3 models for thoracic disease detection
  - Aortic Enlargement, Cardiomegaly, Pleural Thickening, Pulmonary Fibrosis, No Finding
  - Logistic Regression / SVM / CNN
- Logistic Regression performed the best across all metrics (AUC, F1, accuracy)
  - Additionally is was the most computationally efficient
- Lack of efficacy of CNN models

# Thank you!

# Reference links for data and stuff

Past Powerpoint:
https://docs.google.com/presentation/d/1k5_-dGR8qoV-HKyayk4uby9PWyVnEO7V7mANLOyxVVI/edit?slide=id.g349d5c2eaf0_0_10#slide=id.g349d5c2eaf0_0_10

Colab:

(Old) https://colab.research.google.com/drive/1fSQw8jvgEeiBmfblqfPsI6Dmn7FUEHz2?authuser=1

(Latest)
https://colab.research.google.com/drive/1wrqoyukZCSysKNUlPPAUm0G1ycrzBKLX?authuser=1

Rubric:
https://bcourses.berkeley.edu/courses/1540266/files?preview=90391386

# Transfer Learning Model Performance Challenges

Despite using these powerful architectures, we observed poor performance (F1 scores ~0.13-0.33):

- **Domain shift**: Models pre-trained on natural images (ImageNet) may not transfer well to medical imaging
- **Limited compute capacity constrained training data**: Our dataset size (1000 samples) is potentially insufficient for fine-tuning deep networks
- **Class imbalance**: Medical conditions have inherent imbalance affecting model training
- **Subtle visual patterns**: Radiological findings can be extremely subtle compared to ImageNet object classes
- **Inter-class similarity**: Different pathologies may have similar visual appearances
- **Mutual exclusivity constraints**: The relationship between "No finding" and other classes creates unique learning challenges

We explored multiple approaches to address these issues including custom loss functions, various fine-tuning strategies, and custom architectures but did not see any promising results.

# Misclassification Deep Dive

- Cardiomegaly is most commonly misclassified as

# Misclassification Deep Dive