# NIH Chest X-ray Multi-label Disease Classification

AI Final Project Report

**Team :**
Ali Seyhan 210717048
Baki Turhan 210717044

# 1 Introduction

The purpose of this project is to develop and evaluate deep learning models capable of detecting multiple thoracic diseases from chest X-ray images using the NIH Chest X-ray dataset. This is a multi-label classification task where each image may contain more than one pathological label.

# 2 Dataset Description

## 2.1 Source:

NIH Chest X-ray Dataset (`https://www.kaggle.com/datasets/nih-chest-xrays/data`)

**Total Images Before Filtering:** 112,120
**Labels (15 classes):**

- Atelectasis
- Cardiomegaly
- Consolidation
- Edema
- Effusion
- Emphysema
- Fibrosis
- Hernia
- Infiltration
- Mass
- No Finding
- Nodule
- Pleural Thickening
- Pneumonia
- Pneumothorax

**Multi-label property:** Each image can have multiple disease labels.

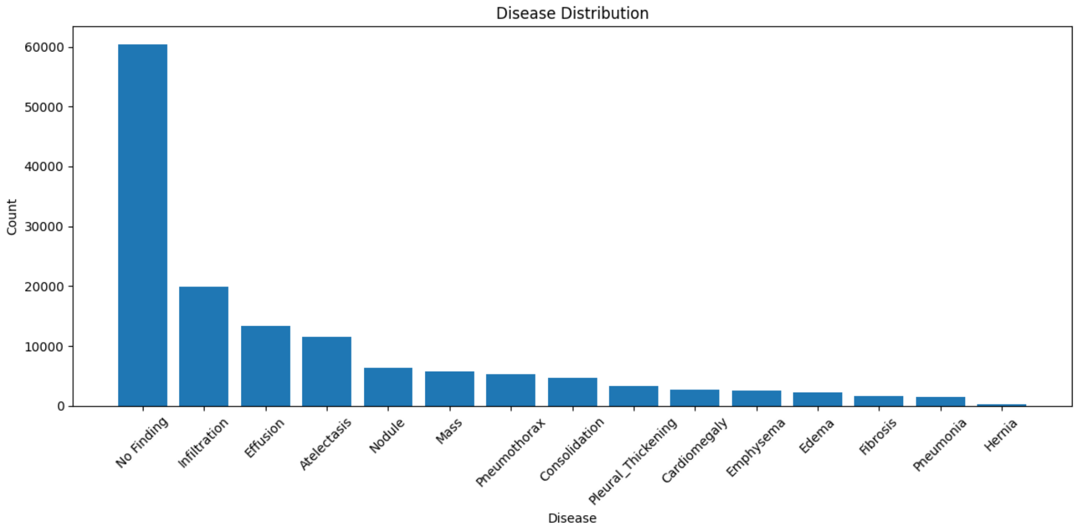


Figure 1: Disease Distribution

## 2.2 Data Preparation Steps:

- Filtering "No Finding" cases: Reduced to 3x per-class target to reduce label imbalance.
- Under-sampling Overrepresented Classes: Targeted 600 images per class after under/oversampling.
- Removed Rare Classes: Classes with less than 500 samples (Hernia) after sampling were removed to reduce noise and data sparsity.
- Final number of images: 9310 images.

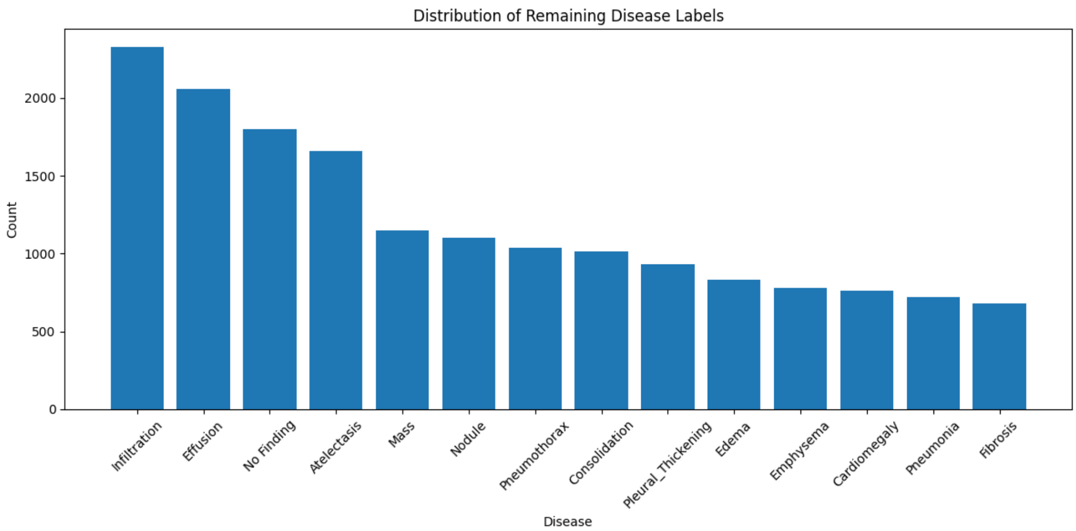


Figure 2: Remaining Disease Distribution

# 3 Data Processing and Augmentation

## 3.1 Image Preprocessing:

Resized to 224x224 for DenseNet121 and VGG19.

## 3.2 Train/Val/Test Split:

70% / 15% / 15% .

## 3.3 Augmentation Applied (Train Only):

- Random Horizontal Flip
- Random Rotation (±10 degrees)
- Color Jitter (brightness, contrast)
- Random Affine Transformations

# 4 Models and Training

## 4.1 DenseNet121 Model

**Pretrained Weights:** ImageNet
**Classifier Modification:**

- Linear → 512 → ReLU → Dropout(0.5) → Linear → 14 classes (sigmoid activation).

**Loss Function:** BCEWithLogitsLoss (with pos_weight for imbalance handling)
**Fine-tuning:** After epoch 5, feature extractor unfrozen, LR reduced to 0.0001
**Optimizer:** Adam, LR=0.0005, Weight Decay=1e-4
**Early Stopping:** patience=10

## 4.2 VGG19 Model

**Pretrained Weights:** ImageNet
**Classifier Modification:**

- Linear → 4096 → ReLU → Dropout(0.5) → Linear → 1024 → ReLU → Dropout(0.5) → Linear → 14 classes.

**Loss Function:** Focal Loss ($\gamma$=2.0) with class-wise $\alpha$ balancing (computed from training set).
**Optimizer:** Adam, LR=0.0005
**Gradient Accumulation:** 16 steps
**Early Stopping:** patience=5

# 5 Evaluation Metrics

- ROC-AUC (Per-class, Macro, Micro)
- Subset Accuracy (Exact Match)
- Macro Class-wise Accuracy
- Precision, Recall, F1-Score (per class)

# 6 Results

## 6.1 DenseNet121 Performance:

- **Subset Accuracy:** 0.1231
- **Macro Class-wise Accuracy:** 0.7856
- **Macro AUC:** 0.7591
- **Micro AUC:** 0.7708

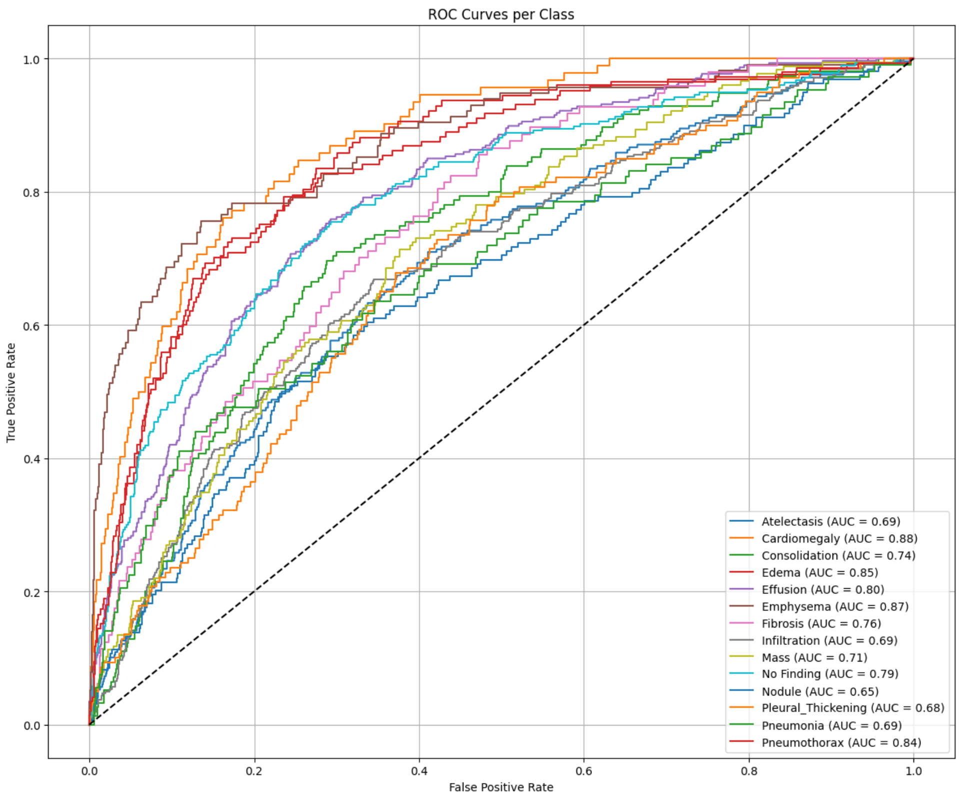


Figure 3: DenseNet121 ROC Curves

## 6.2 VGG19 Performance:

- **Subset Accuracy:** 0.0809
- **Macro Class-wise Accuracy:** 0.7562
- **Macro AUC:** 0.7054
- **Micro AUC:** 0.7375

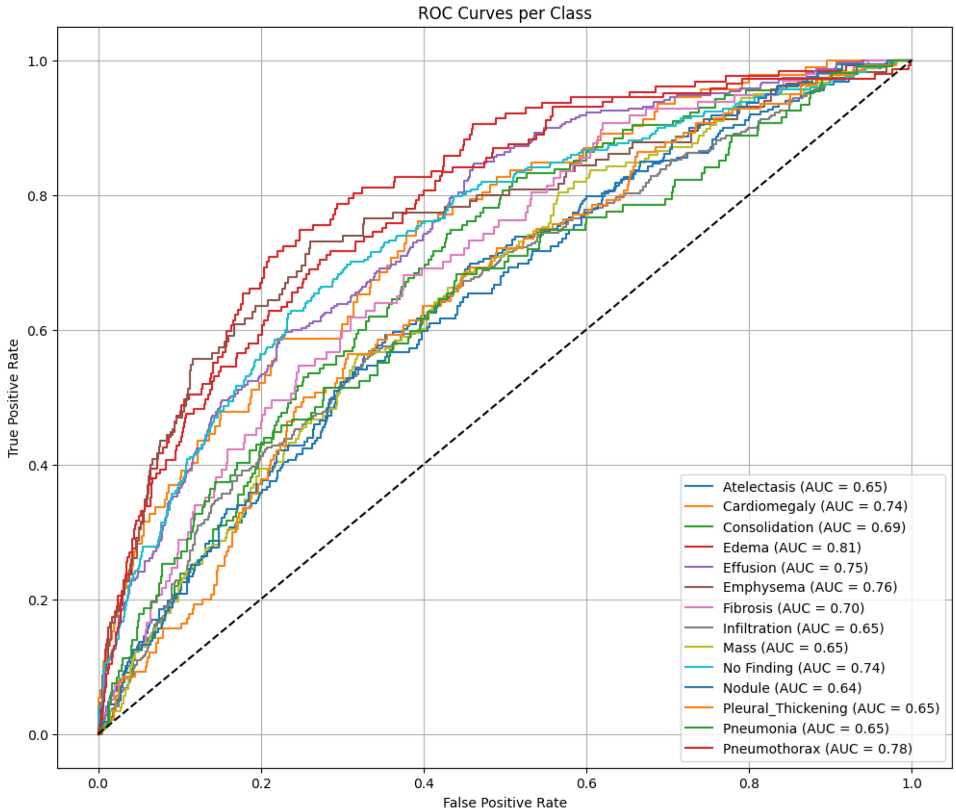


Figure 4: VGG19 ROC Curves

# 7 Conclusion

A multi-label classification pipeline for chest X-ray images was successfully implemented. DenseNet121 and VGG19 architectures were evaluated, showing potential for automatic disease detection. However, additional fine-tuning and more data for rare classes would improve real-world clinical usability.