Automated Pneumonia Diagnosis Using CNNs on Radiographic Images

Sameera Aluri

Problem to Investigate

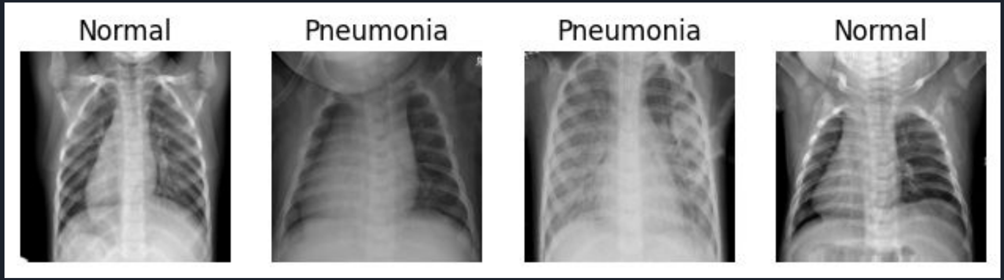
- **GOAL:** Automatically detect pneumonia in patients using chest X-ray images using a deep learning approach
- **PROBLEM STATEMENT:** Can we build a deep learning model that accurately identifies pneumonia from chest X-ray images, distinguishing it from normal lungs?
- **OBJECTIVE:** Train a CNN-based binary classifier to detect:
  - Class 0: Normal
  - Class 1: Pneumonia
- **MOTIVATION:** Pneumonia is a major global health burden
  - Chest X-rays are diagnostic standard, but error-prone
  - AI can support faster, scalable, and more consistent diagnoses

Project Scope

- Included
  - Binary classification (Pneumonia vs. Normal)
  - Deep learning models (custom CNN + pretrained model)
  - Image preprocessing & augmentation
  - Evaluation with standard ML metrics
- Not Included
  - Multiclass disease classification
  - Clinical deployment or integration
  - Real-time image pipeline or API

Dataset Overview

- **Source:** Kaggle – Chest X-Ray Images (Pneumonia)
- **Image Type:** Grayscale X-Ray JPEGs
- **Structure:**
  - ~5,200 training images
  - ~600 test images
  - Labels: *NORMAL* or *PNEUMONIA*
- **Challenges:**
  - Slight class imbalance
  - Variable image resolution
  - No patient metadata (pure image-based)



Measurable Objectives

1. Build a custom CNN model from scratch
2. Fine-tune a pretrained model
3. Evaluate with metrics: accuracy, precision, recall, F1 score, AUC
4. Use Grad-CAM to visualize decision regions

Key Approaches & Findings

- **CNN-Based Models**
  - CheXNet (Rajpurkar et al., 2017)
    - 121-layer DenseNet trained on ChestXray14 (adult)
    - AUC ~0.915 (matches radiologist-level accuracy)
- **Transfer Learning**
  - Pretrained models (e.g., ResNet, VGG, EfficientNet)
  - Kermany et al., (2018) fine-tuned models on pediatric data
    - Achieved high accuracy with limited samples
- **Custom Architectures & Ensembles**
  - PneumoniaNet (Alsharif et al., 2021)
    - Pediatric-specific CNN with ~99.7% accuracy
  - Qiuyu et al. 2024
    - Ensemble: DenseNet + EfficientNet + Attention
    - AUC > 0.95
- **Explainability**
  - Grad-CAM (Selvaraju et al., 2017)
    - Visual heatmaps highlight regions influencing predictions
    - Essential for clinical interpretability

Comparing Methodologies

Study	Model	Dataset	AUC / Accuracy	Notes
Rajpurkar	DenseNet-121	ChestXray14 (Adult)	AUC ~0.915	Strong baseline
Kermany	VGG/ResNet (transfer)	Pediatric	~93-95%	Few images, expert-level perf.
Alsharif	PneumoniaNet	Pediatric	~99.7%	Custom CNN
Qiuyu	Ensemble (DenseNet + EfficientNet)	Pediatric	AUC > 0.95	Uses attention modules
Cohen	ResNet	External Pediatric	AUC ~0.54	Shows generalization issue

Model Architecture

- **Backbone:** Pretrained ResNet50 (include_top=False)
- **Fine-tuning:** Last 50 layers unfrozen
- **Input:** 224×224 cropped X-rays (3-channel)
- **Head:**
  - GlobalAveragePooling → Dropout (0.5)
  - Dense (128, ReLU) → Dropout (0.3)
  - Output: Dense (1, Sigmoid)
- **Loss:** BinaryFocalCrossentropy (γ=2.0)
- **Optimizer:** Adam (LR=1e-5)
- **Metrics:** Accuracy, AUC, Recall
- **Imbalance Handling:** class_weight + focal loss
- **Callbacks:** EarlyStopping, ReduceLROnPlateau

```
Epoch 5/15
131/131 ━━━━━━━━━━━━━━━━━━━━ 192s 1s/step - accuracy: 0.8480 - auc: 0.9299 - loss: 0.0860 - recall: 0.8243 - val_accuracy: 0.8554 - val_auc:
0.9561 - val_loss: 0.0840 - val_recall: 0.8129 - learning_rate: 2.0000e-06
Epoch 6/15
131/131 ━━━━━━━━━━━━━━━━━━━━ 212s 2s/step - accuracy: 0.8487 - auc: 0.9284 - loss: 0.0884 - recall: 0.8375 - val_accuracy: 0.7586 - val_auc:
0.9591 - val_loss: 0.1232 - val_recall: 0.6787 - learning_rate: 2.0000e-06
Epoch 7/15
131/131 ━━━━━━━━━━━━━━━━━━━━ 0s 1s/step - accuracy: 0.8518 - auc: 0.9309 - loss: 0.0819 - recall: 0.8321
Epoch 7: ReduceLROnPlateau reducing learning rate to 3.999999989900971e-07.
131/131 ━━━━━━━━━━━━━━━━━━━━ 204s 2s/step - accuracy: 0.8519 - auc: 0.9309 - loss: 0.0819 - recall: 0.8322 - val_accuracy: 0.7193 - val_auc:
0.9613 - val_loss: 0.1464 - val_recall: 0.6258 - learning_rate: 2.0000e-06
Epoch 8/15
131/131 ━━━━━━━━━━━━━━━━━━━━ 213s 2s/step - accuracy: 0.8449 - auc: 0.9353 - loss: 0.0810 - recall: 0.8193 - val_accuracy: 0.7615 - val_auc:
0.9613 - val_loss: 0.1246 - val_recall: 0.6826 - learning_rate: 4.0000e-07
Epoch 8: early stopping
Restoring model weights from the end of the best epoch: 5.
```

Cropping Issue

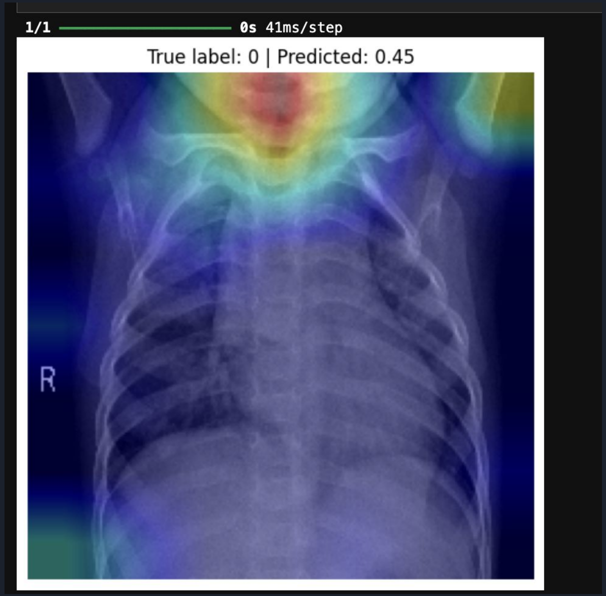
Problem:

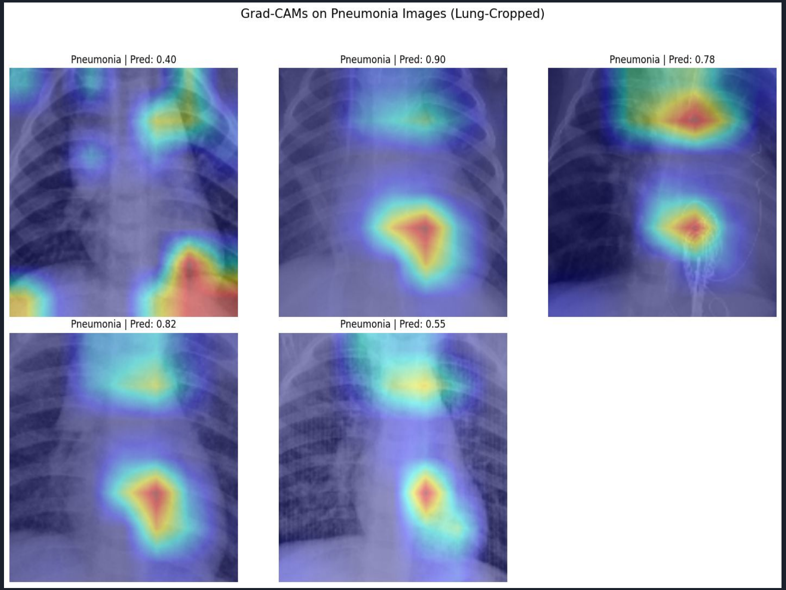
- Original X-rays include non-lung regions (neck, shoulders)
- Model sometimes focused on irrelevant areas in training
- Led to overfitting and poor generalization especially for NORMAL cases

Solution:

- Applied fixed lung-region cropping before training
- Removed distractions and guided model toward relevant anatomy

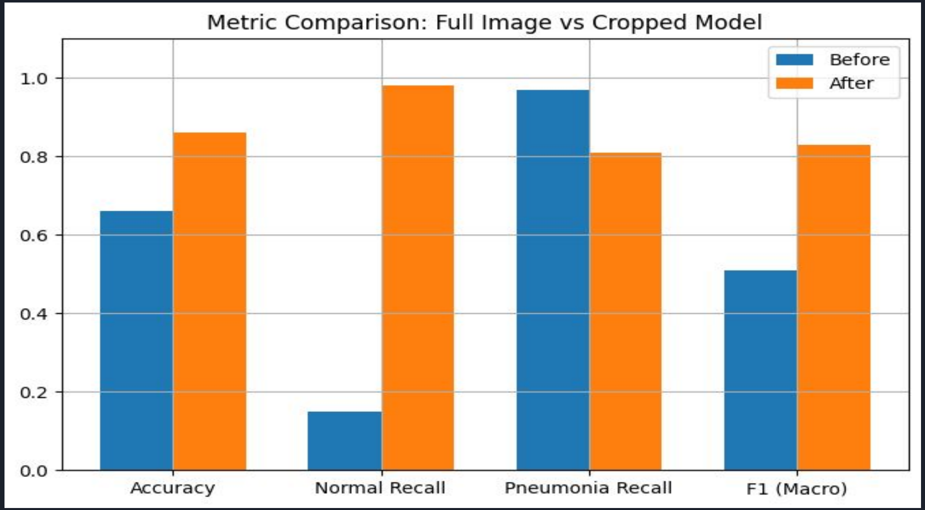
Before vs. After Cropping





Before vs. After Cropping

	Full Image	Cropped Image
Grad-CAM Focus	Neck, corners	Lungs, lower lobes
Normal Recall	15%	98%
Accuracy	66%	86%



Evaluation Metrics & Results

BEFORE:

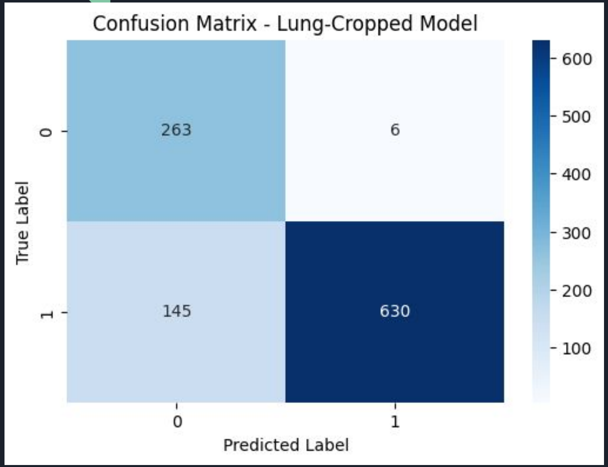
Classification Report:

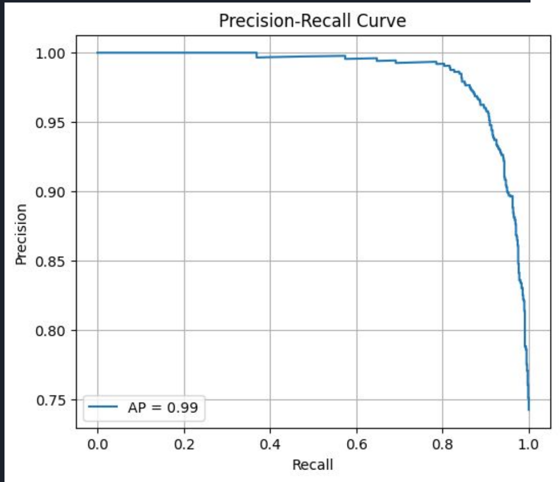
	precision	recall	f1-score	support
Normal	0.72	0.15	0.24	234
Pneumonia	0.65	0.97	0.78	390
accuracy			0.66	624
macro avg	0.69	0.56	0.51	624
weighted avg	0.68	0.66	0.58	624

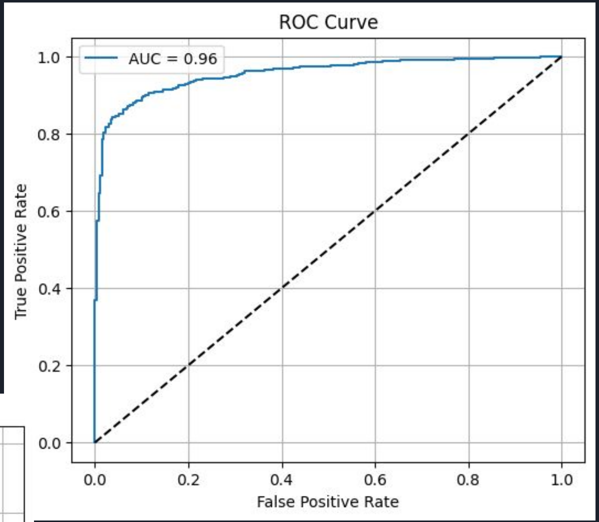
AFTER:

	precision	recall	f1-score	support
Normal	0.64	0.98	0.78	269
Pneumonia	0.99	0.81	0.89	775
accuracy			0.86	1044
macro avg	0.82	0.90	0.83	1044
weighted avg	0.90	0.86	0.86	1044

Evaluation Metrics & Results







Takeaways

- Lung cropping significantly improved model focus and reduced overfitting
- Class imbalance handled with focal loss + class weights
- Normal class recall improved from 15% to 98%
- Grad-CAM visualizations confirmed anatomically meaningful attention
- Final model achieved 86% accuracy and 0.96 AUC
- Outperformed baseline and matched or exceeded literature benchmarks
- Demonstrated that simple preprocessing can drive large performance gains

Future Work

- Integrate lung segmentation (U-Net) for dynamic, patient-specific cropping
- Implement uncertainty estimation for clinical decision support
- Expand to multi-class classification (e.g., bacterial vs viral pneumonia)
- Fine-tune on larger or more diverse datasets for better generalization

Questions?