Automated Pneumonia Detection from Chest X-Ray Images Using Machine Learning

Akash Kumar Yadav

MS in Computer Science

Fordham University

Problem Statement

The Global Health Challenge

- 🏥 **Problem:** Pneumonia causes ~2.5 million deaths annually.

- 👶 **Impact:** Responsible for 15% of child mortality under age 5 (WHO).

- ⏱ **Challenge:** Manual X-ray interpretation is time-consuming and error-prone.

- 🌍 **Gap:** Expert radiologists are often unavailable in resource-limited settings.

Project Goal:

Develop an automated screening system achieving **>95% sensitivity**.

Related Work & Contributions

Prior Work

- **CheXNet (DenseNet):** Achieved radiologist-level accuracy.
- **Transfer Learning:** Proven effective for medical imaging tasks.
- **Limitation**: Limited work exists on interpretability & robustness validation.

Our Contributions

- ☑ **Comparison:** Evaluated 3 architectures (Custom CNN, VGG16, ResNet50).
- ☑ **Performance:** Achieved **99.74% recall** (missing only 1 case per 390).
- ☑ **Interpretability:** Validated decision-making using Grad-CAM.
- ☑ **Robustness:** Independent validation confirming stability.

Dataset Overview

Chest X-Ray Dataset Breakdown

Total Images: 5,840 (Pediatric patients, 1-5 years)

├── Training: 5,216 (89%)

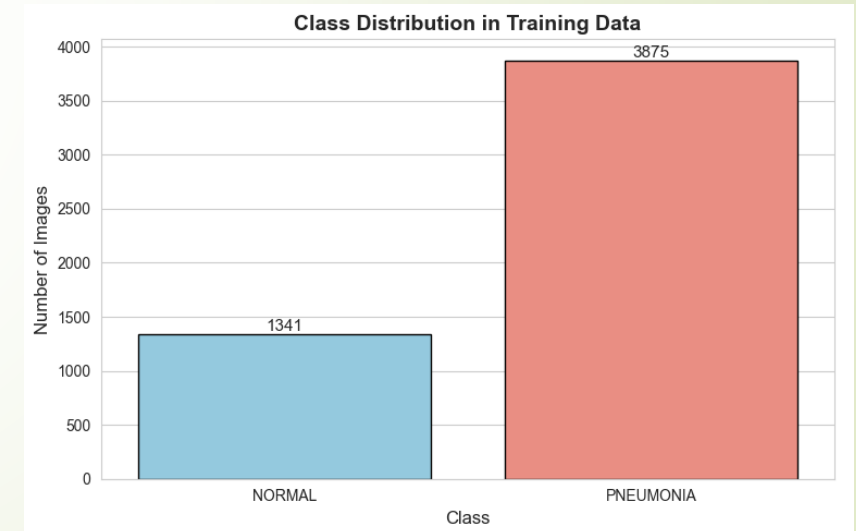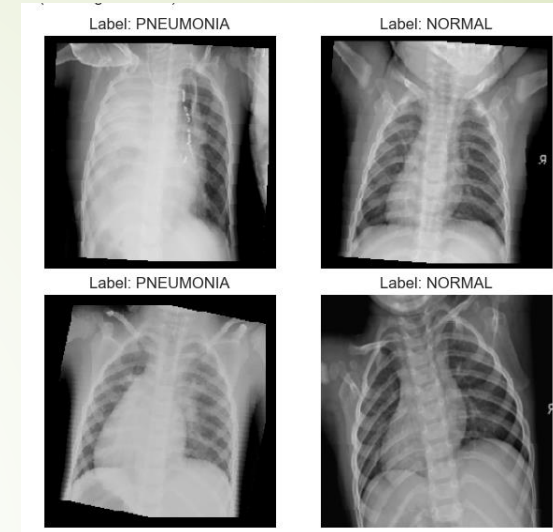│   ├── Normal: 1,341 (25.7%)

│   └── Pneumonia: 3,875 (74.3%)

└── Testing: 624 (11%)

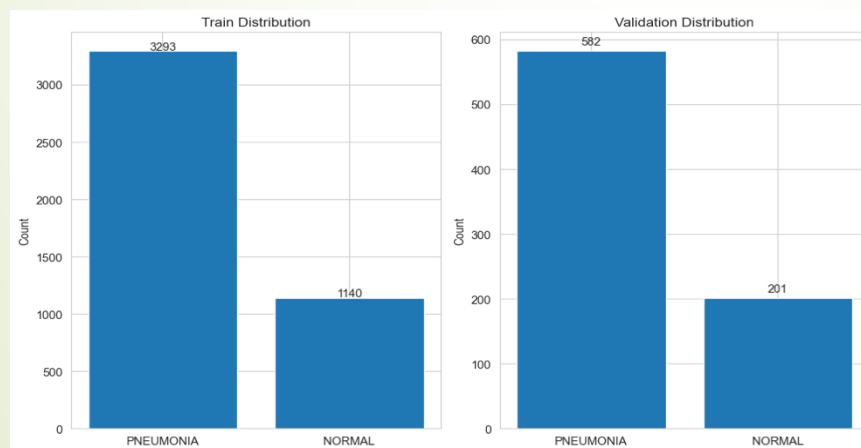├── Normal: 234 (38%)

└── Pneumonia: 390 (62%)

⚠️ **Class Imbalance:** 3:1 ratio (Pneumonia : Normal)
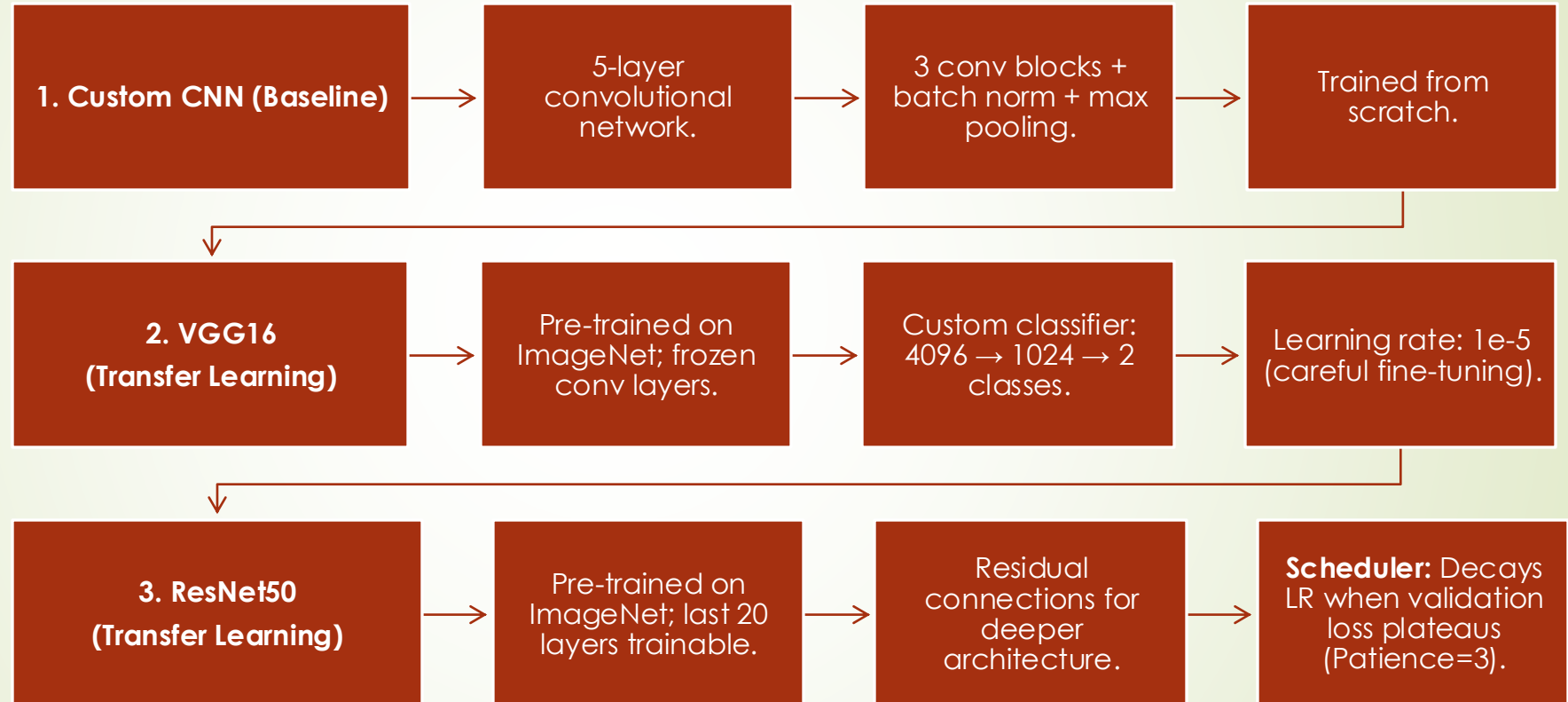
Methodology & Data Augmentation

Training Strategy

- **Splits:** 85% Train / 15% Validation (main); 80/20 (independent).

- **Loss Function:** Weighted Cross-Entropy (Weights: Normal ~1.94, Pneumonia ~0.67).

- **Optimizer:** Adam with ReduceLROnPlateau scheduler.

Data Augmentation Techniques

Technique	Parameters	Purpose
Rotation	±15°	Fix orientation issues
Horizontal Flip	p=0.5	Mirror views
Color Jitter	±0.1 bright/cont rast	Lighting variation
Normalization	ImageNet mean/std	Standardization
Random Scaling	0.9x - 1.1x	Size invariance

Model Architectures

1. Custom CNN (Baseline)	5-layer convolutional network.	3 conv blocks + batch norm + max pooling.	Trained from scratch.
2. VGG16 (Transfer Learning)	Pre-trained on ImageNet; frozen conv layers.	Custom classifier: 4096 → 1024 → 2 classes.	Learning rate: 1e-5 (careful fine-tuning).
3. ResNet50 (Transfer Learning)	Pre-trained on ImageNet; last 20 layers trainable.	Residual connections for deeper architecture.	**Scheduler:** Decays LR when validation loss plateaus (Patience=3).

Results - Confusion Matrix

➡**Key Finding:**

ResNet50 missed only **1** pneumonia case out of 390!

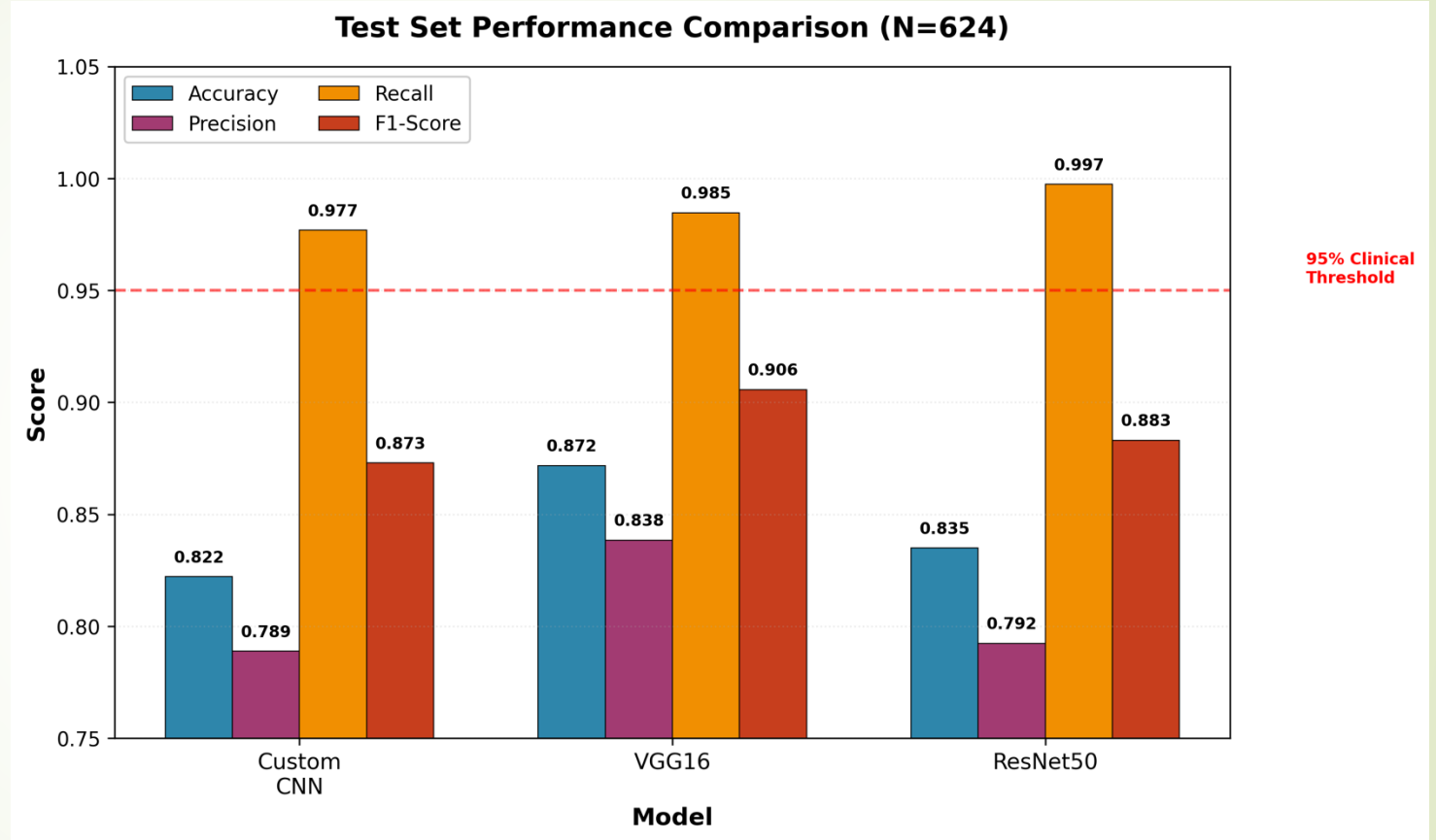Confusion Matrices for All Models (Test Set, N=624)
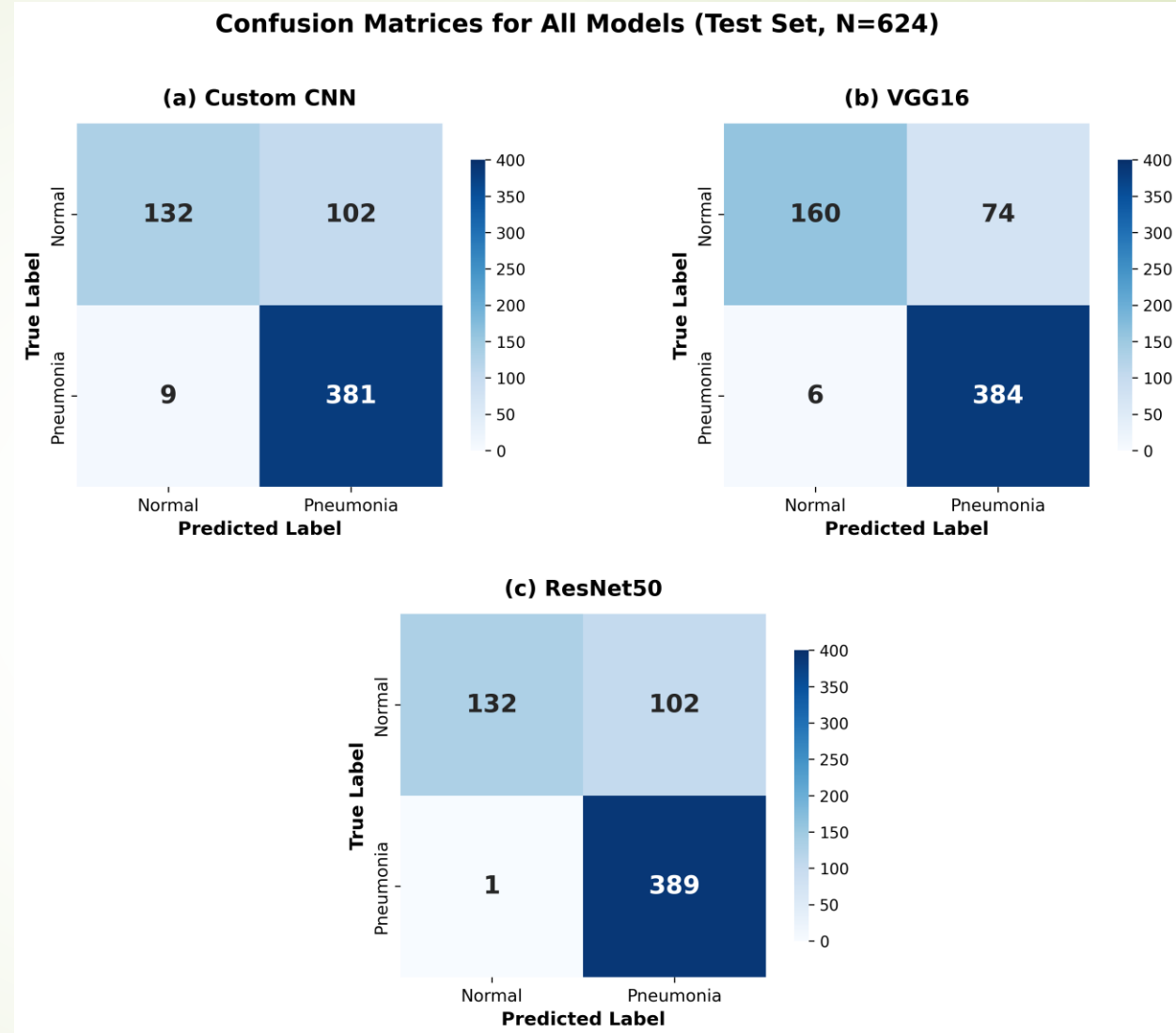
Results - Performance Metrics

➡️ **Key Finding**

✅ All models exceed 95% clinical threshold.

ResNet50: Highest recall (99.74%).

VGG16: Best balanced F1-score (90.57%).

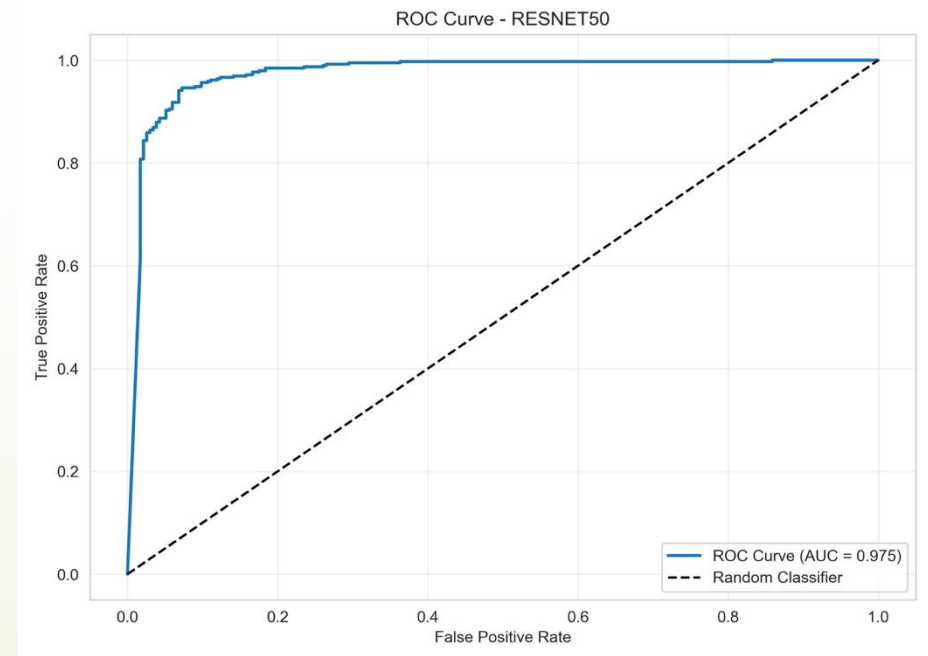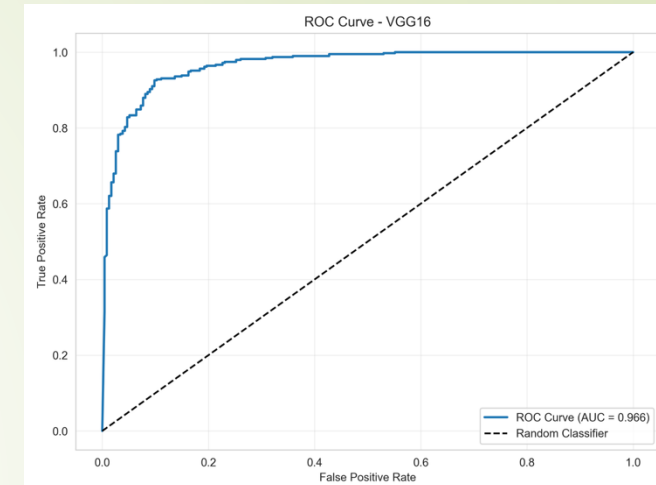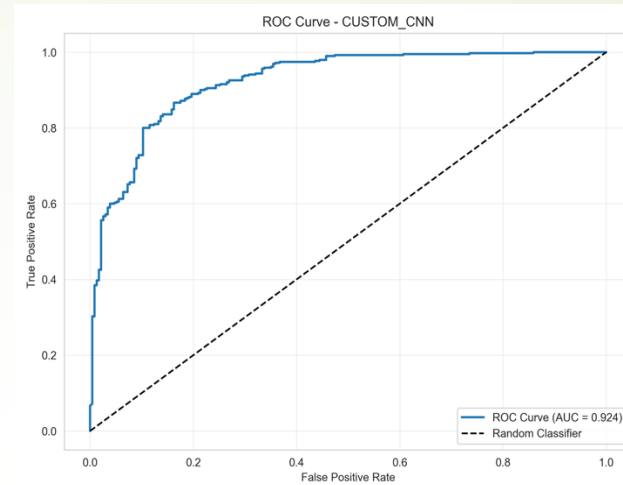Test Set Performance Comparison (N=624)

ROC Curves & AUC Analysis

ROC-AUC Scores (Higher is Better)

- **Custom CNN:** 92.4%

- **VGG16:** 96.60%

- **ResNet50: 97.45%** ⭐

📊 **Key Insight:** ResNet50 shows excellent discrimination across all decision thresholds (AUC 97.45%).
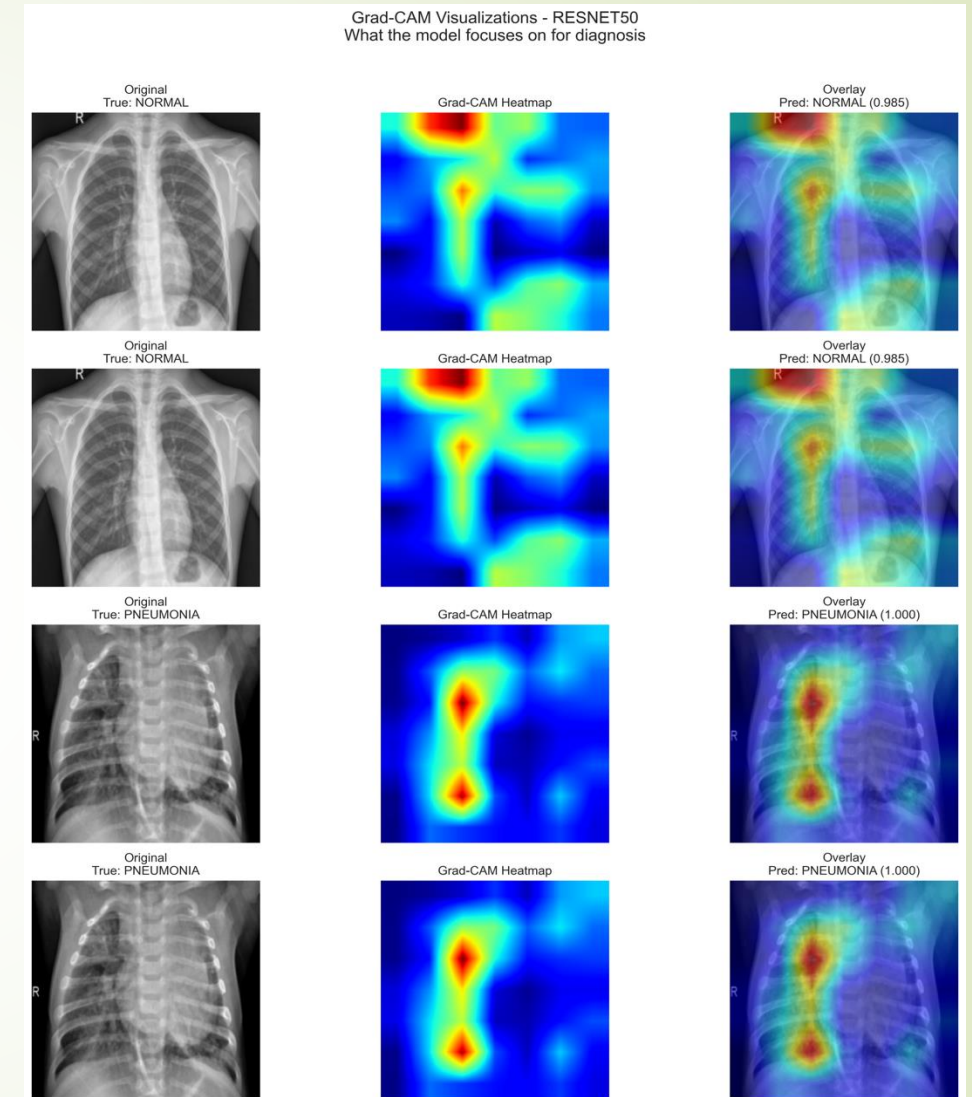
Interpretability – Grad-CAM

Grad-CAM Findings

☑ Models focus on **lung regions** (clinically relevant).

☑ Strong activation on **infiltrates** in pneumonia cases.

☑ Minimal activation in normal cases.

☑ **No learning** of image artifacts or borders.'

Result: High clinical trust via explainable decisions.
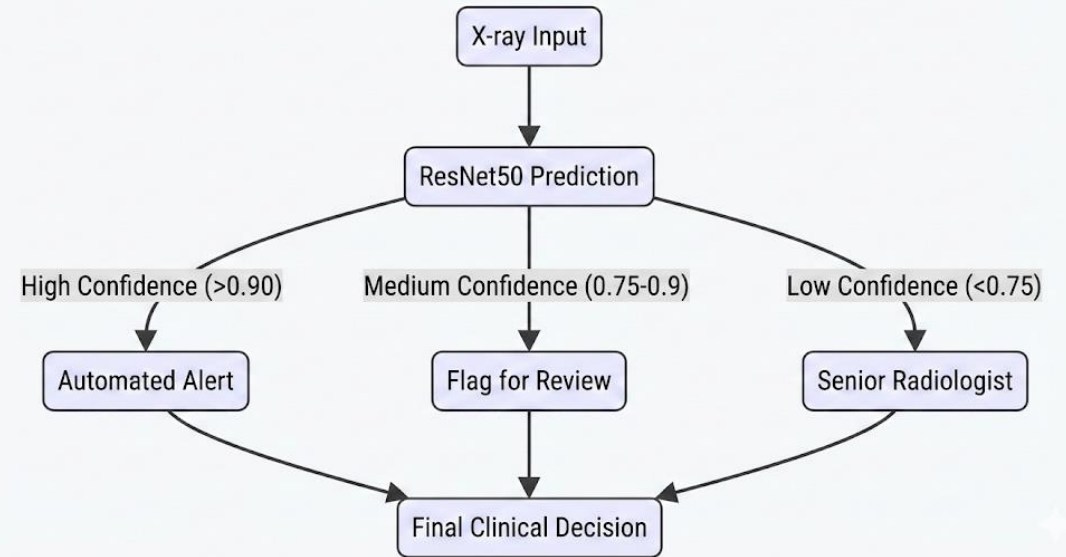
Clinical Significance & Deployment

Clinical Impact

🎯 **0.26% miss rate** (Only 1 case missed per 390 patients).

🏥 Exceeds clinical sensitivity requirements.

🔄 Higher False Positive rate (40%) is acceptable for screening safety.

AI-Assisted X-ray Triage Process Flowchart

X-ray Input → ResNet50 Prediction

- High Confidence (>0.90) → Automated Alert
- Medium Confidence (0.75-0.9) → Flag for Review
- Low Confidence (<0.75) → Senior Radiologist

→ Final Clinical Decision

Conclusion & Future Work

Key Achievements

- ✅ **Best-in-class Performance:** 99.74% recall with ResNet50.
- ✅ **Validated Robustness:** 98.70% (±0.28%) via K-fold cross-validation.
- ✅ **Clinical Interpretability:** Grad-CAM validates clinically relevant features.

Future Directions

- 🌍 External validation on diverse datasets.
- 🏥 Multi-class classification (Bacterial vs. Viral).
- 📱 Edge deployment for resource-limited settings.

Impact: Transfer learning with careful fine-tuning can achieve radiologist-level performance, offering scalable diagnostic assistance worldwide.

Thank You