Automated Detection of Chest X-Ray Abnormalities Using Deep Learning

Motivation & Problem Definition

Problem

- Chest X-ray interpretation is **time-critical** and **expert-dependent**, yet errors may occur due to:
- High workload
- Subtle abnormalities
- Visual similarity between normal & abnormal cases

Example

- A faint lung opacity may be overlooked → delayed diagnosis.

Goal

- Automatically classify chest X-rays as **Normal vs Abnormal** using deep learning.

Why Is This Problem Important?

- Chest X-rays are the **most common medical imaging exam**
- Early detection improves patient outcomes
- AI can assist radiologists**, not replace them**
- Real-world impact in low-resource healthcare settings

Dataset Description

- **NIH Chest X-ray Dataset (20K subset)**
- Binary classification:
    - **Normal**
    - **Abnormal**
- Pre-split into:
    - Training
    - Validation
    - Test sets
- **Preprocessing**
- Resized to **224×224**
- Normalized using ImageNet statistics
- Data augmentation applied during training

Baseline Model

- **CNN from Scratch**
- Simple convolutional layers
- Trained only on NIH data
- **Observation**
- Poor generalization
- Low validation accuracy
- Motivated the use of **transfer learning**

Transfer Learning Strategy

- **Why Transfer Learning?**
- ImageNet models learn:
  - edges
  - shapes
  - textures
- Faster convergence
- Better performance with limited data
- **Models Used**
- **ResNet18**
- **DenseNet121**
- **EfficientNet-B0**

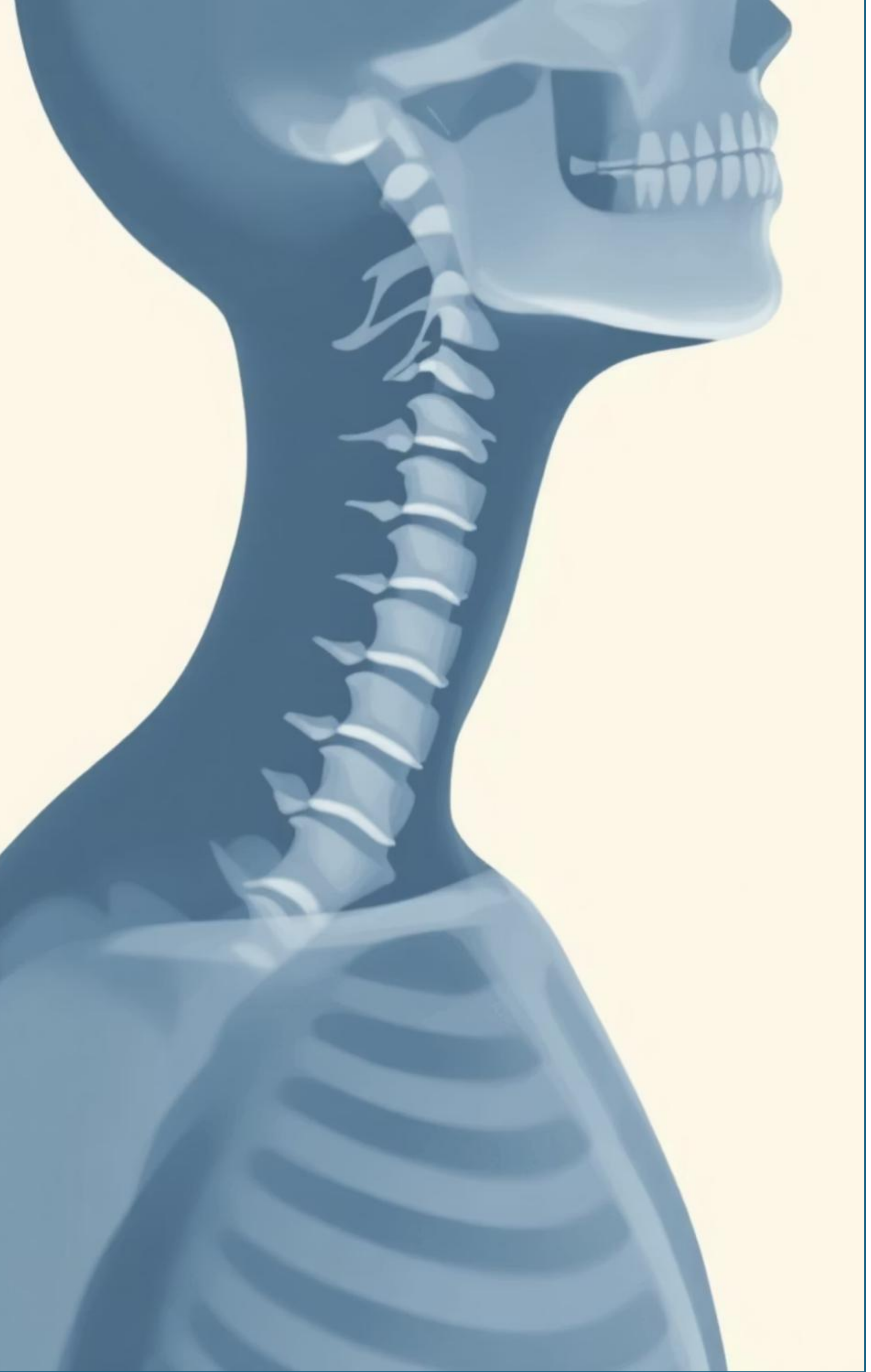
Two-Phase Fine-Tuning (Key Idea)

Phase 1: Feature Extraction

- Freeze backbone
- Train only classifier

Phase 2: Fine-Tuning

- Unfreeze last block + classifier
- Lower learning rate
- ✓ Prevents catastrophic forgetting
  ✓ Enables task-specific adaptation



Deep Learning Approaches: Model Evolution

We employed four distinct deep learning architectures, progressing from a custom CNN baseline to sophisticated transfer learning models.

1. **Custom CNN**

   Baseline model to understand complexity; failed to learn effectively.

2. **ResNet18**

   Foundation for transfer learning, leveraging pre-learned features.

3. **DenseNet121**

   Maximizing feature reuse with dense connectivity patterns.

4. **EfficientNet-B0**

   State-of-the-art scaling with compound scaling methodology.

Model 1: ResNet18

Why ResNet?

- Residual connections
- Solves vanishing gradient problem
- Efficient & stable

Training Highlights

- LR: 1e-3 → 1e-4
- Batch size: 32
- Early stopping (patience = 3)

Model 2: DenseNet121

Why DenseNet?

- Dense connectivity
- Feature reuse
- Parameter efficiency

Observations

- Faster training loss reduction
- Slight overfitting after early epochs
- Early stopping triggered

Model 3: EfficientNet-B0

Why EfficientNet-B0?

- Compound scaling (depth, width, resolution)
- MBConv + Squeeze-and-Excitation
- Best FLOPs / accuracy trade-off

Role

- Primary candidate for optimization

Hyperparameter Tuning Approaches

Method 1: Manual Tuning

- Used for ResNet18 & DenseNet121
- Trial-and-error based on learning curves

Method 2: Grid Search (EfficientNet-B0)

- Systematic exploration
- Validation accuracy–based selection

EfficientNet-B0 Grid Search

Search Space (16 configs)

- Unfreezing strategy: classifier / last block
- Learning rate: 1e-3, 3e-4
- Batch size: 16, 32
- Dropout: 0.2, 0.3

Best Configuration

- Last block unfreezing
- LR = 3e-4
- Batch size = 32
- Dropout = 0.3

Evaluation Metrics

Used on **held-out test set:**

- Accuracy
- Precision
- Recall
- F1-score
- **AUC-ROC**
- Why AUC?
- Robust to class imbalance
- Measures ranking quality

Quantitative Results (Table)

Table 1.Peformance Comparsion Table

Method	Model	Best Val Acc	Test Acc	Key Configuration
Manual	CNN(BASELINE)	0.49%	0.5%	Simple CNN, trained from scratch
Manual	ResNet18	66.37%	66.13%	LR=1e-4, BS=32, Phase2 unfreezing
Manual	DenseNet-121	68.67%	67.70%	LR = 1e-4, BS = 32, last dense block + classifier unfrozen
Manual	EfficientNet-B0	67.10%	67.67%	LR = 1e-3, BS = 32, classifier-only fine-tuning
Grid	EfficientNet-B0	68.47%	67.70%	LR = 3e-4, BS = 32, dropout = 0.3, last block unfrozen

ROC

- **Interpretation**
- AUC ≈ **0.72**
- Indicates fair discriminative ability
- Better than random (0.5)

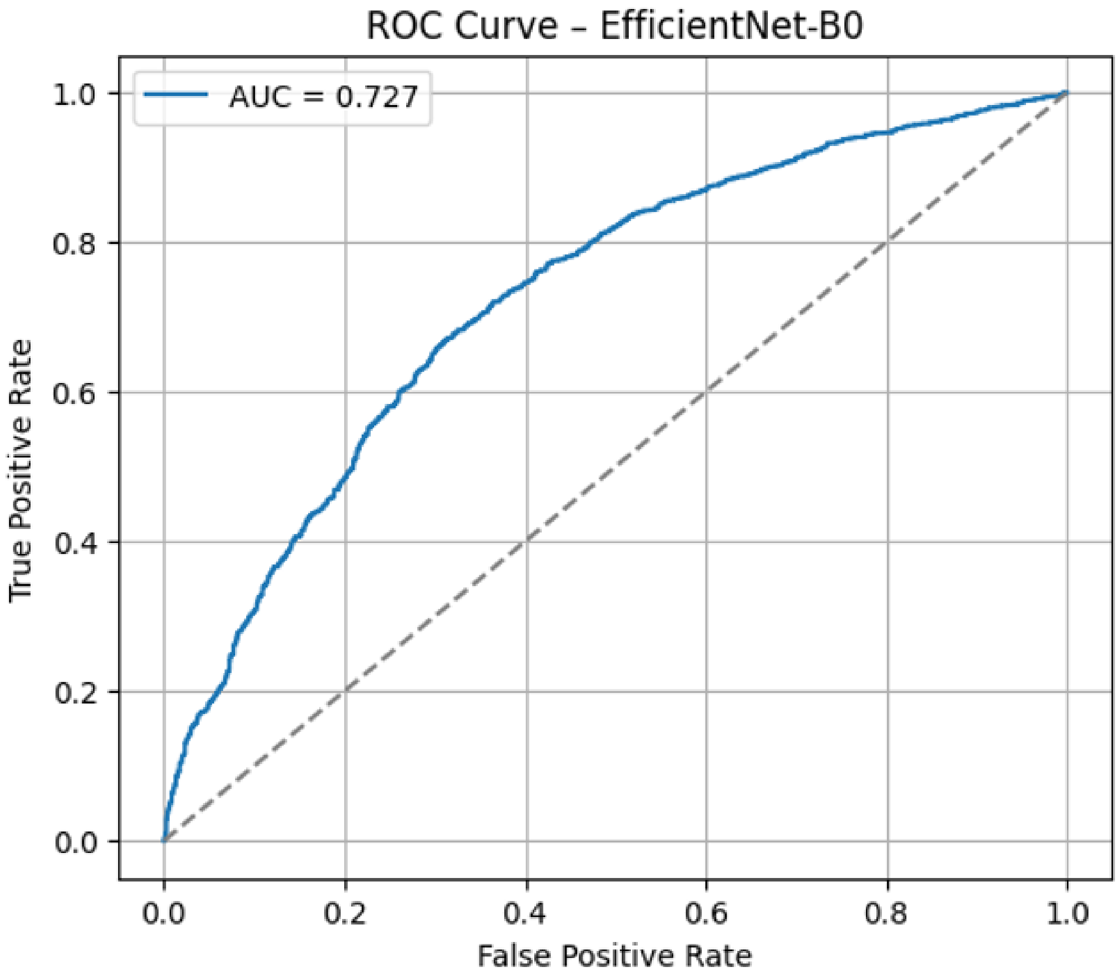


Figure 2.ROC Curve Comparison of Best Models.

Confusion Matrix

- Balanced false positives & false negatives
- No strong class bias
- Stable decision boundaries

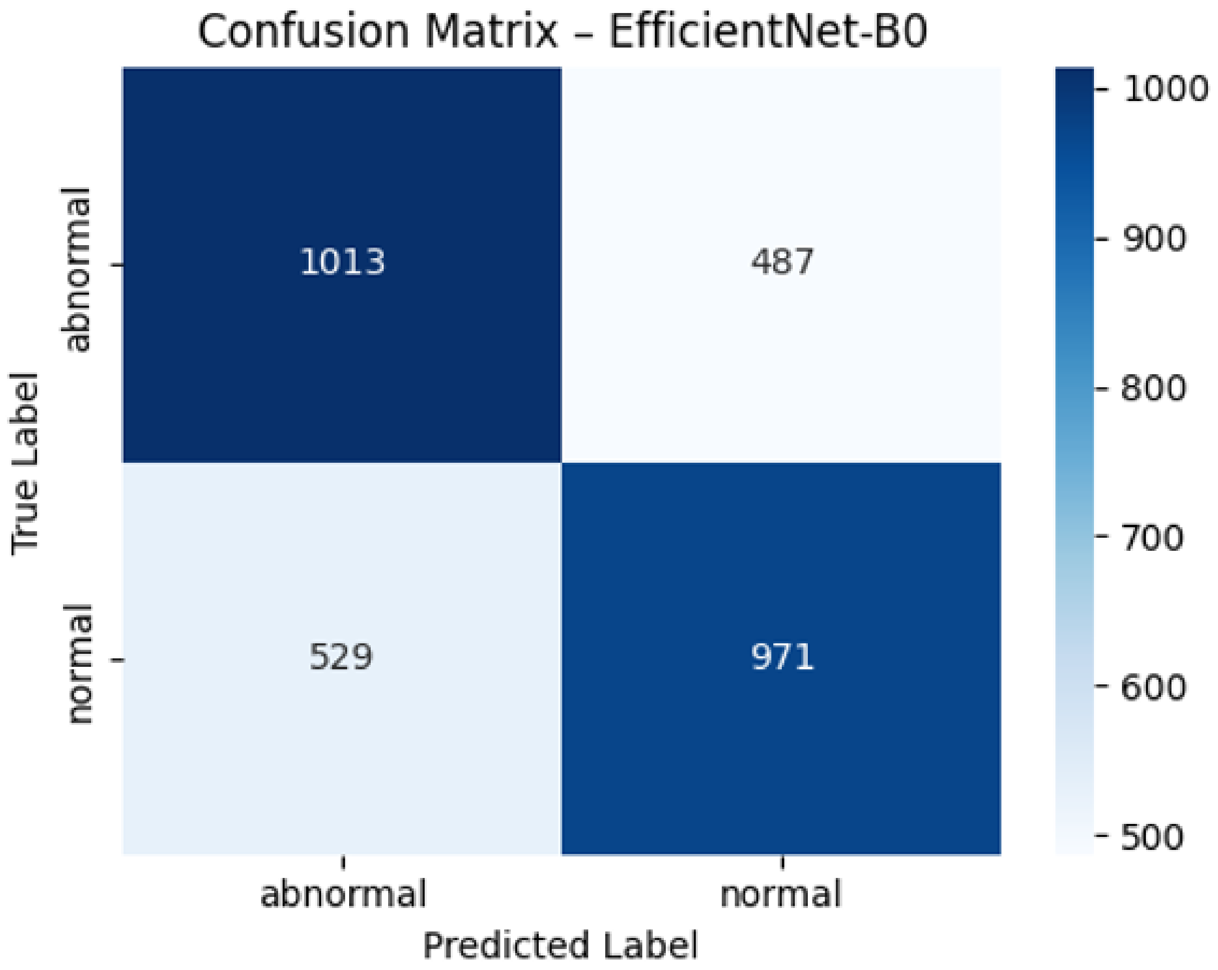


Figure 1.Confusion Matrix of the Best Model.

Grad-CAM Visualizations

What We See

Focus on lung regions

Attention on pathological areas

No reliance on image borders or artifacts

➡️ Confirms meaningful feature learning

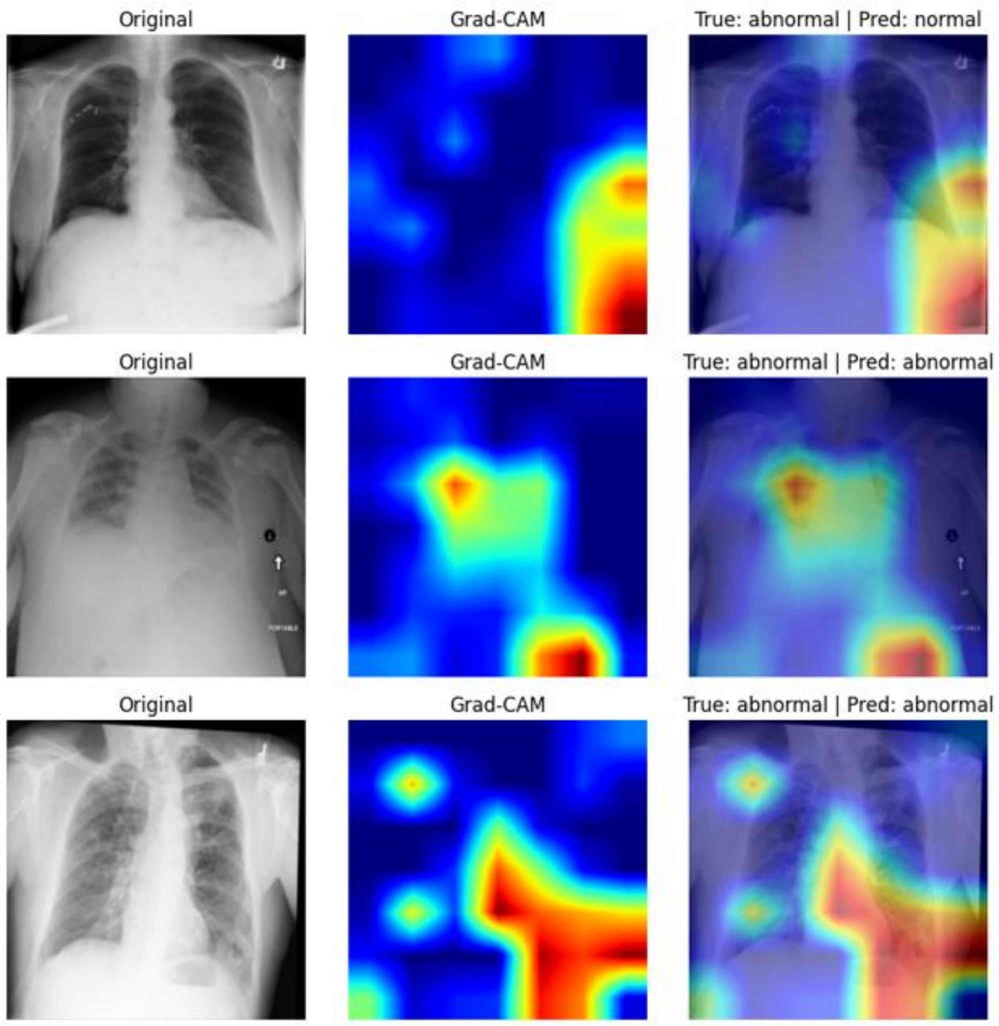


Figure 3.Grad-CAM Visualization of Model Attention Regions.

Discussion & Challenges

- **Key Findings**
- Transfer learning vastly outperforms CNN baseline
- EfficientNet-B0 + grid search performs best
- Grad-CAM improves interpretability
- **Challenges**
- Subtle abnormalities
- Label noise
- Limited dataset size

Conclusion

✓ Transfer learning is effective for chest X-ray analysis

✓ EfficientNet-B0 achieved the best overall performance

✓ Explainability is crucial for medical AI trust

Future Work

- Multi-class disease classification
- Clinical expert validation
- Larger datasets
- Ensemble models
- Deployment-oriented optimization