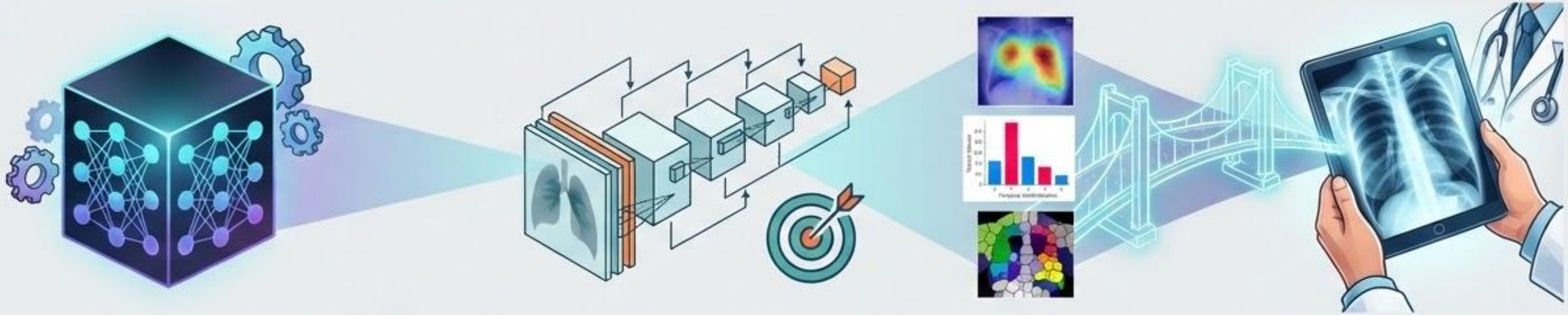
Pneumonia Detection & Explainable AI

Dheeraj Namargomala

Bridging the Gap Between AI and Clinical Trust



The Challenge:
Deep learning models achieve high accuracy in medical imaging but operate as "black boxes," making clinical adoption difficult.

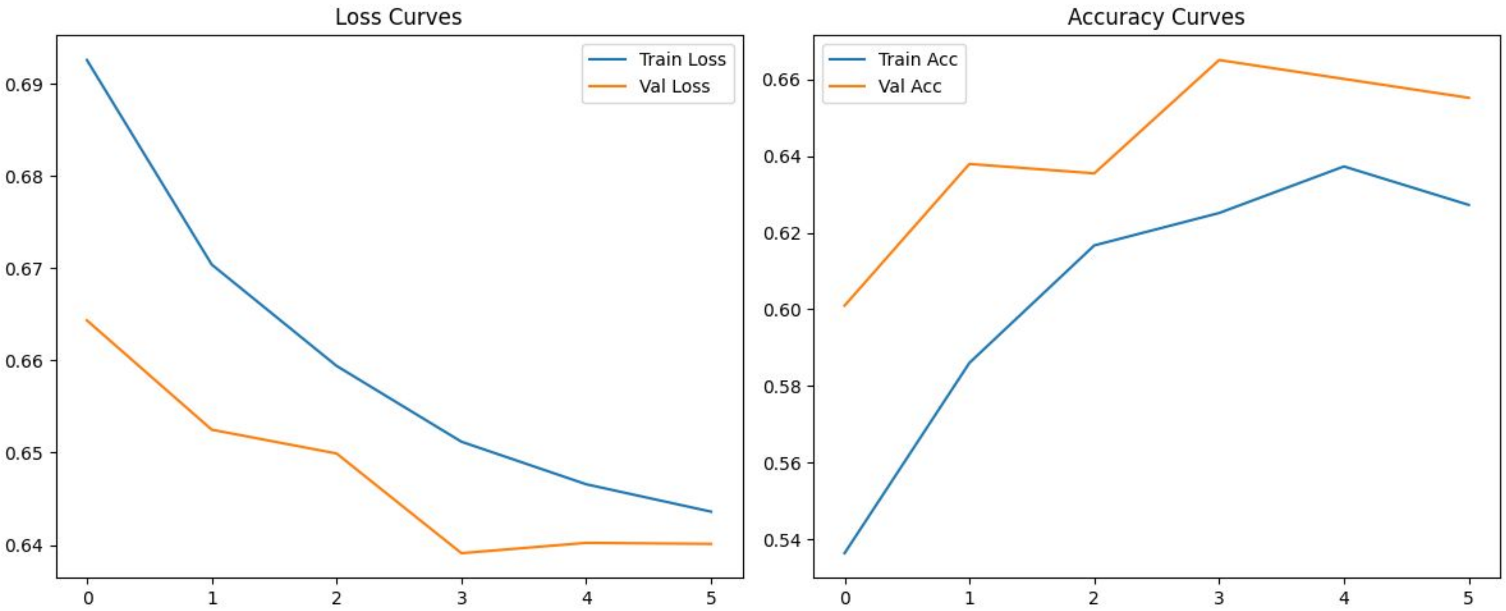
The Goal:
Train a robust Convolutional Neural Network (CNN) to detect pneumonia.

The Innovation:
Apply and compare three distinct Explainable AI (XAI) frameworks (Grad-CAM, SHAP, and LIME) to interpret the model's predictions.

The NIH Chest X-Ray14 Dataset

- **Source Data:** Utilized the publicly available NIH dataset (resized to 224x224 for computational efficiency).
- **Class Balancing:** Pneumonia cases are historically underrepresented. We created a balanced, stratified subset to prevent the model from biasing toward the majority "Normal" class.
- **Augmentation:** Applied random horizontal flips and rotations during training to improve model generalization.
- **Splitting:** 70% Training / 15% Validation / 15% Testing.

Loss & Accuracy

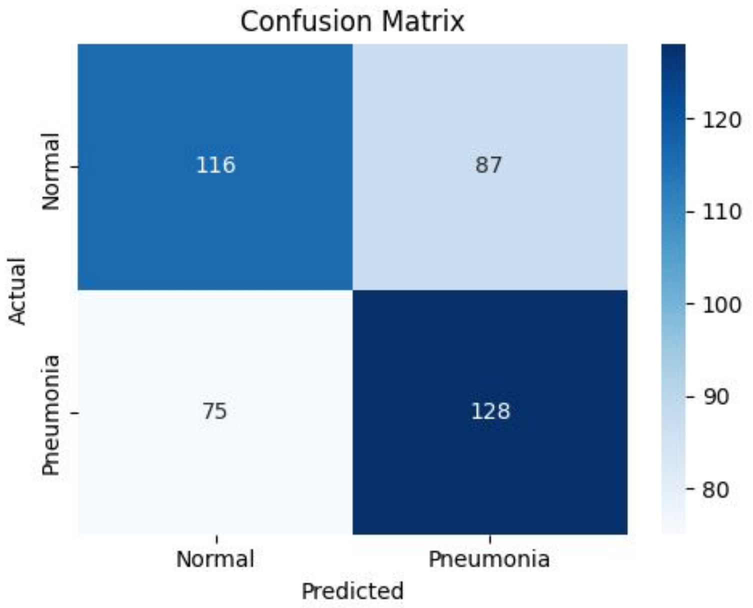


Deep Learning Core: DenseNet-121

- **Architecture:** Dense Convolutional Network (DenseNet-121).
- **Why DenseNet?:** Excellent feature reuse, reduced parameter count, and highly effective for medical image classification.
- **Transfer Learning:** Initialized with ImageNet weights to leverage pre-learned edge and texture detection.
- **Modification:** Final classification layer modified to output a simple binary answer (Pneumonia vs. Normal).

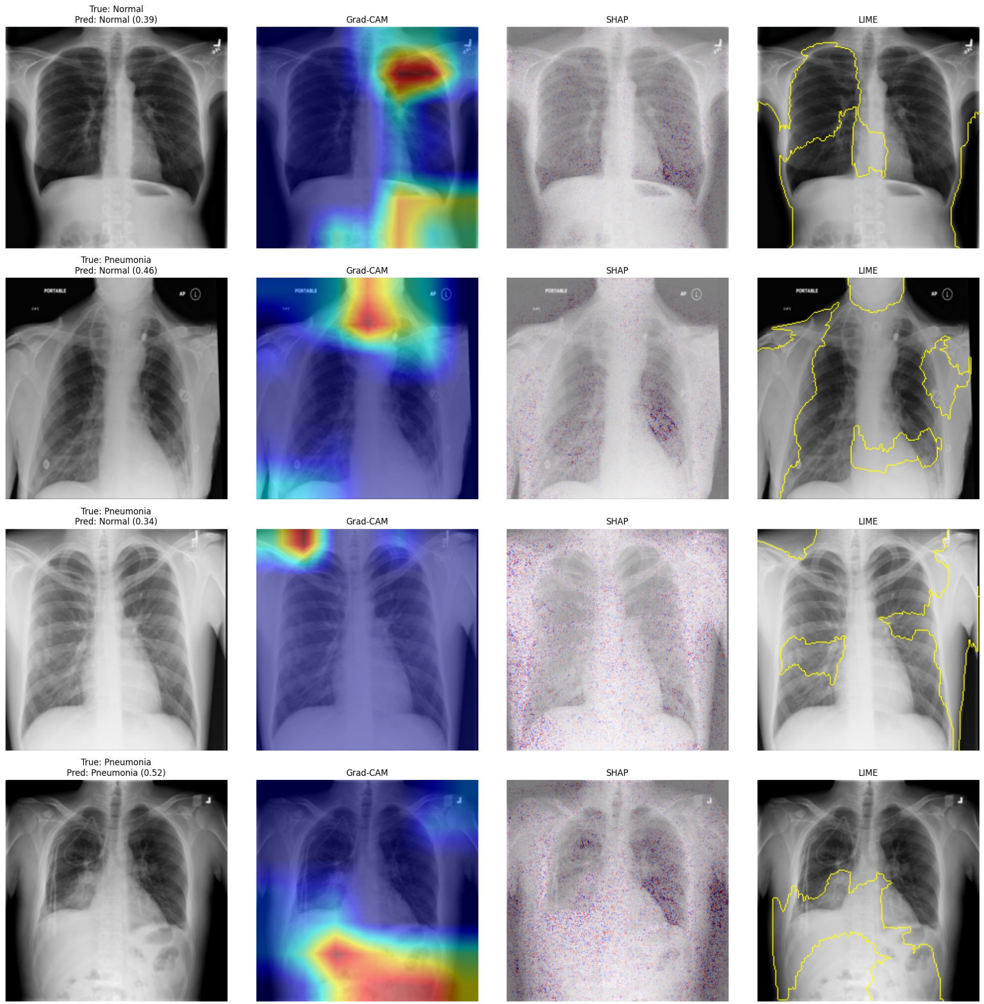
Evaluating Model Accuracy

- **Optimization:** Trained using BCEWithLogitsLoss and an Adam Optimizer with an adaptive learning rate scheduler.
- **Performance:** Achieved strong convergence across 8 epochs. (Stopping early to prevent overfitting)
- **Key Metrics:** Evaluated using ROC-AUC, Precision, Recall, and F1-Score.
- **Confusion Matrix:** Monitored False Positives (over-diagnosis) vs. False Negatives (missed cases).



Why Explainability (XAI) Matters

- **High Stakes:** In healthcare, a wrong decision impacts patient lives.
- **Artifact Overfitting:** AI can 'cheat' by memorizing radiographic markers, text on the image, or bone structures instead of lung pathology.
- **Regulatory Compliance:** Future medical AI requires traceability and transparency.
- **Our Approach:** Post-hoc interpretability using Grad-CAM, SHAP, and LIME.


True: Normal
Pred: Normal (0.39)
Grad-CAM
SHAP
LIME
True: Pneumonia
Pred: Normal (0.46)
Grad-CAM
SHAP
LIME
True: Pneumonia
Pred: Normal (0.34)
Grad-CAM
SHAP
LIME
True: Pneumonia
Pred: Pneumonia (0.52)
Grad-CAM
SHAP
LIME

Gradient-weighted Class Activation Mapping

- **How it Works:** Uses the gradients of the final convolutional layer to produce a coarse localization map.
- **The Result:** Highlights the regions of the image most responsible for the "Pneumonia" prediction in red/yellow.
- **Pros:** Computationally fast (near real-time), highly intuitive for clinicians.

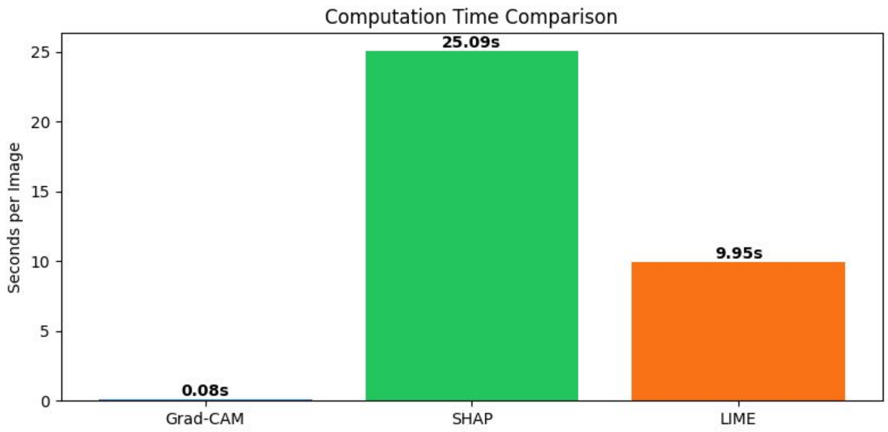
Source: https://stanfordmlgroup.github.io/projects/chexnet/

XAI Methods 2 & 3 - SHAP & LIME

- **SHAP (SHapley Additive exPlanations):** Game-theory approach showing pixel-level positive and negative contributions.
- **LIME (Local Interpretable Model-agnostic Explanations):** Segments the image into "superpixels" to identify specific pathological regions.
- **Pros:** Highly rigorous and robust.
- **Cons:** Computationally expensive; best for offline auditing.

Grad-CAM vs. SHAP vs. LIME

- **Grad-CAM:** Best for rapid, real-time clinical deployment.
- **LIME:** Excellent for fine-grained boundary identification.
- **SHAP:** Best for deep, comprehensive model auditing.
- **Time:** Grad-CAM (~0.1s/image) vs LIME/SHAP (several seconds/image).



Clinical Validation & Future Work

ROI Validation: Our XAI methods successfully localized haziness and consolidation in the lower lobes, aligning with actual clinical pneumonia symptoms.

Limitations: LIME and SHAP computation times limit live use; model still occasionally flags clavicle bones.

Next Steps:

- Integrate model into a lightweight web dashboard.
- Implement attention-guided training to force the model to ignore non-lung areas.

Questions?