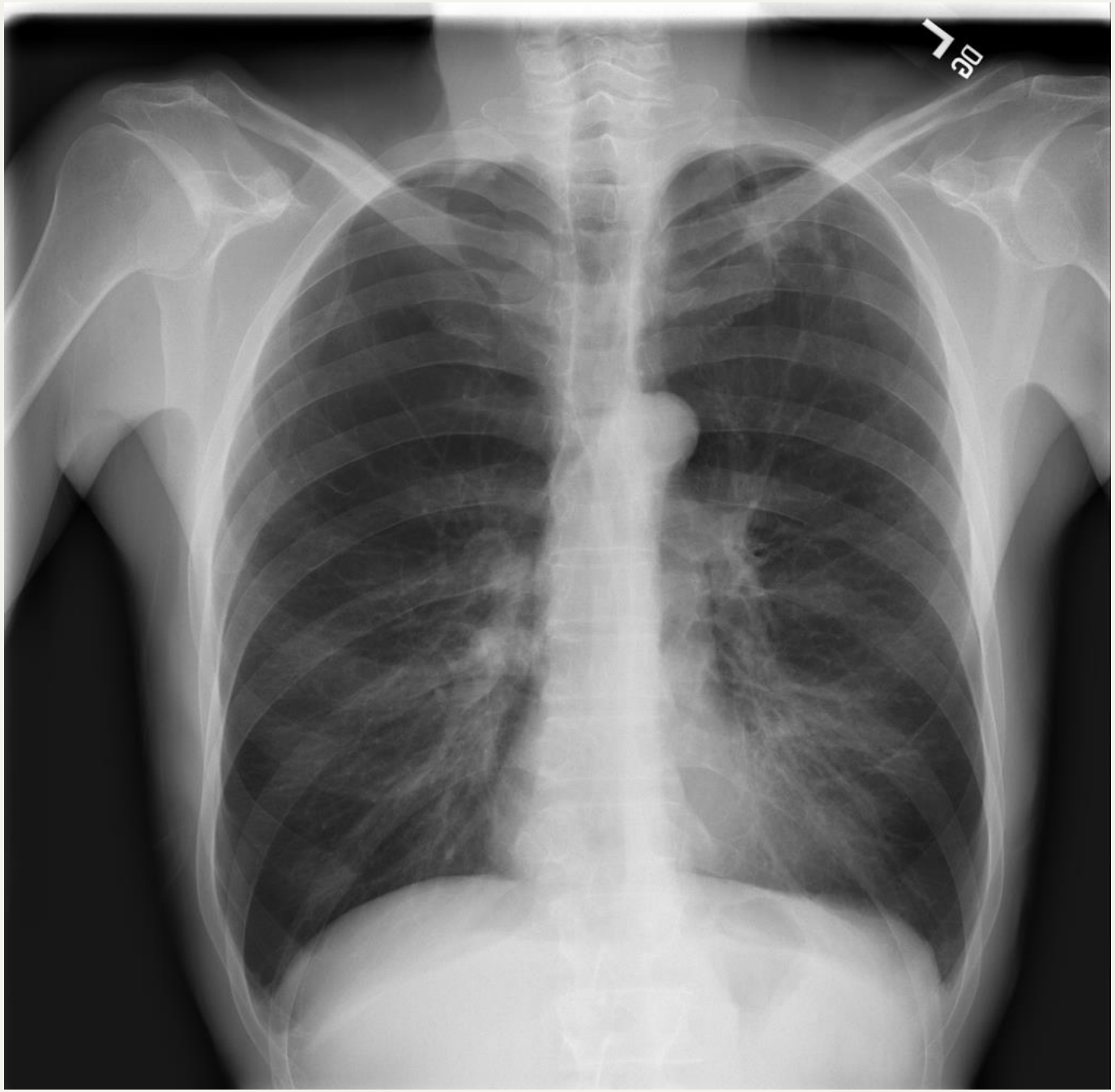
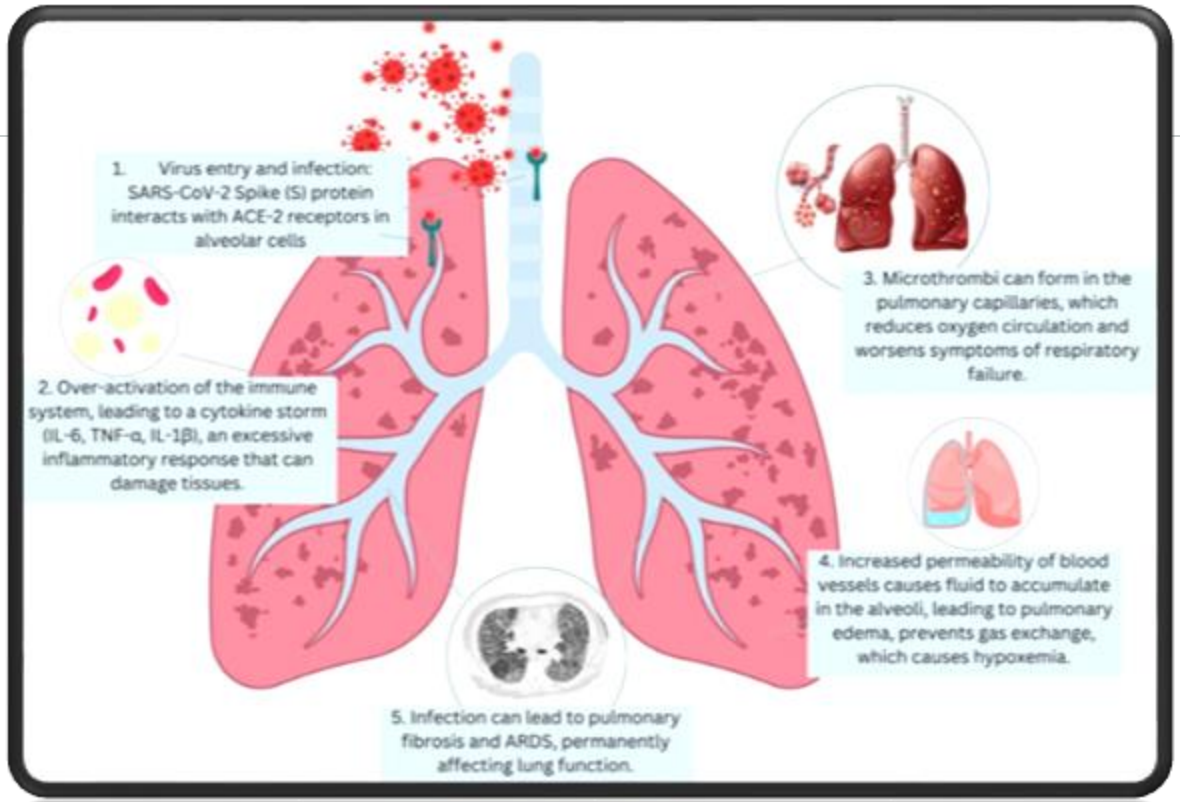


Machine Learning–Driven Pneumonia Screening from X-Ray Images: Performance Analysis of CNN and ViT Architectures

By: Eli Antoine, Jyothi Priya, Srikar Gowrishetty

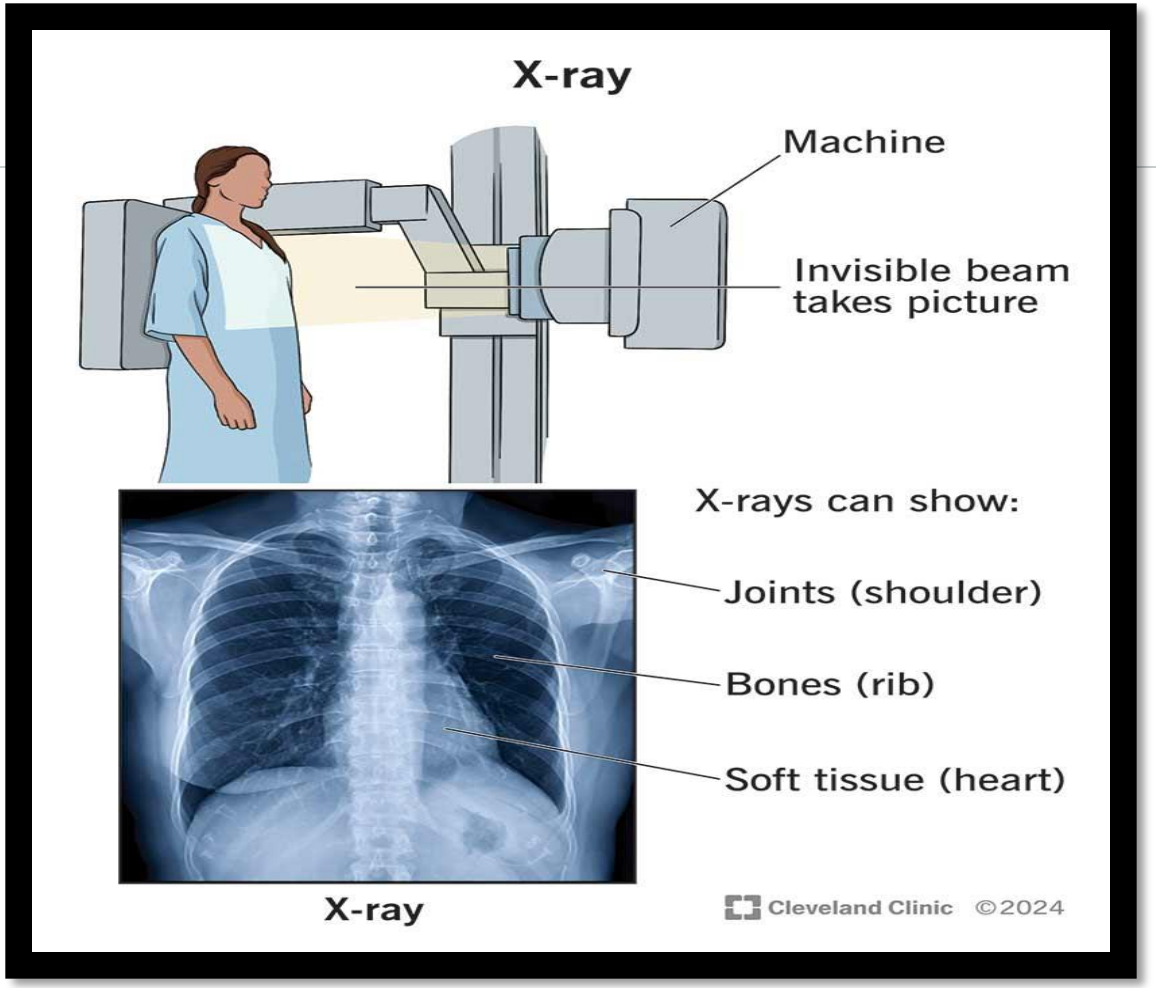
Clinical Problem: Pneumonia



Caliman-Sturdza, O.A.; Soldanescu, I.; Gheorghita, R.E. SARS-CoV-2 Pneumonia: Advances in Diagnosis and Treatment. Microorganisms ***2025****, 13, 1791.*

- Pneumonia is a lung infection affecting the alveoli and surrounding tissue, causing inflammation and fluid accumulation that reduces oxygen exchange.
- Caused by pathogens: Bacteria (e.g., Streptococcus pneumoniae) or viruses (e.g., influenza, COVID-19); can also occur from aspiration of foreign material.
- Main types: Bacterial pneumonia (often localized, rapid onset) and Viral pneumonia (usually diffuse, milder onset but can be severe).
- Symptoms: Fever, cough, chest pain, fatigue, shortness of breath, low oxygen levels; severe cases may progress to respiratory failure or sepsis.
- Treatments: Antibiotics for bacterial cases, antivirals for viral, plus oxygen therapy, hospitalization, and ventilation support if severe.

X-ray Imaging



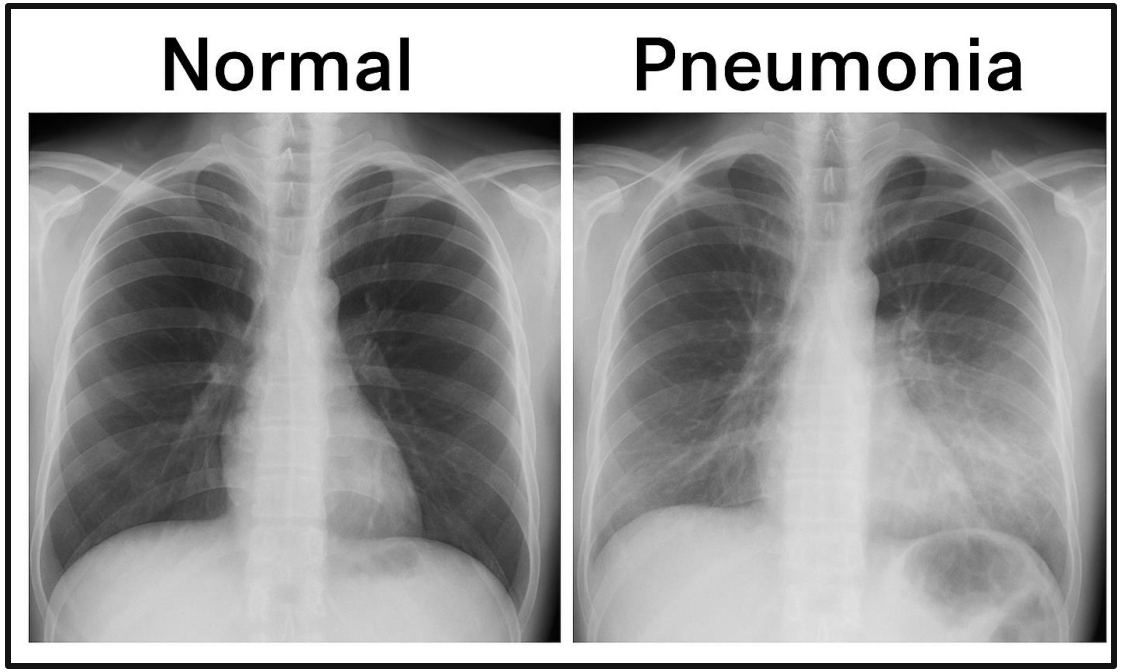
- X-rays use high-energy radiation to create two-dimensional images of the chest, where different tissues absorb radiation at varying levels: bones and fluid appear white, soft tissues appear gray, and air-filled spaces appear dark.
- Pneumonia typically manifests as bright, cloudy opacities in lung regions that should normally appear dark, reflecting inflammation or fluid-filled airspaces.
- Chest X-rays serve as a primary diagnostic tool because they are fast, inexpensive, widely available, and routinely used as the first imaging modality in emergency and hospital settings.
- X-ray imaging helps clinicians assess disease severity by revealing the extent and distribution of lung involvement, guiding decisions regarding hospital admission, oxygen therapy, and urgent interventions.

Chest X-ray illustration. From *X-ray* (Cleveland Clinic, n.d.). https://my.clevelandclinic.org/health/diagnostics/21818-x-ray

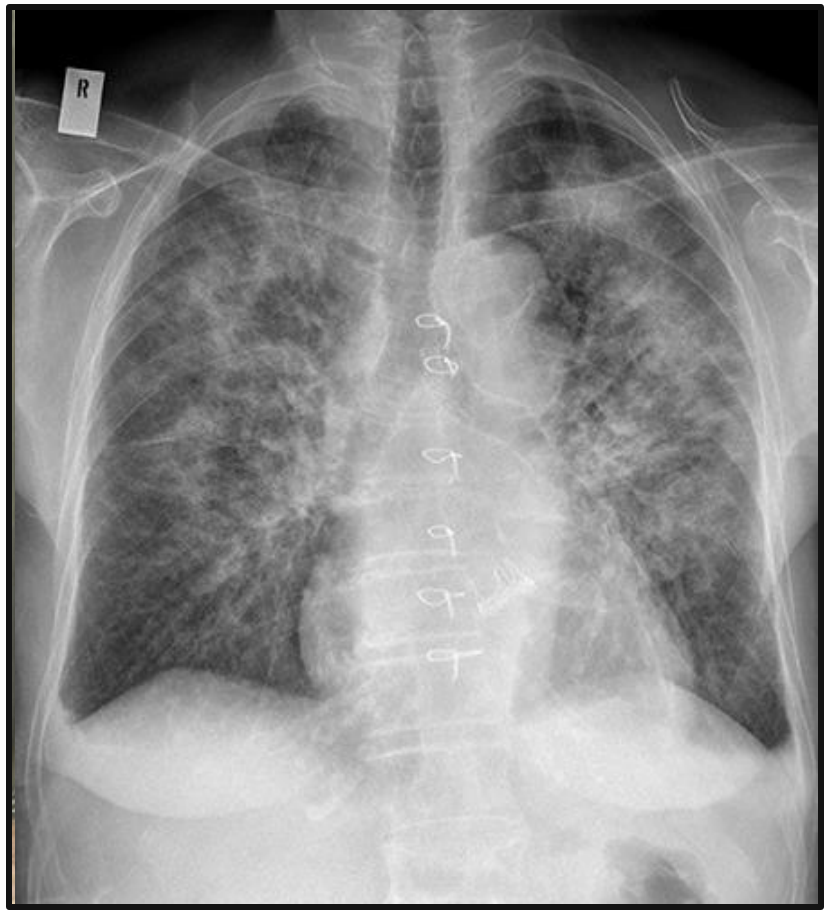
Differentiating Normal vs. Pneumonia X-Ray Images

- **Normal lungs:** mostly dark (air-filled) with visible rib outlines, heart, and diaphragm; no cloudy patches.
- **Pneumonia lungs:** show localized or diffuse white opacities (consolidation), sometimes with blurred lung margins.
- **Patterns:**
- **Lobar/consolidation:** large, dense area in one lobe (common bacterial pneumonia).
- **Patchy/multifocal:** scattered bright spots (viral or bronchopneumonia).

EXAMPLE



Gaps With Current Clinical Detection Methods



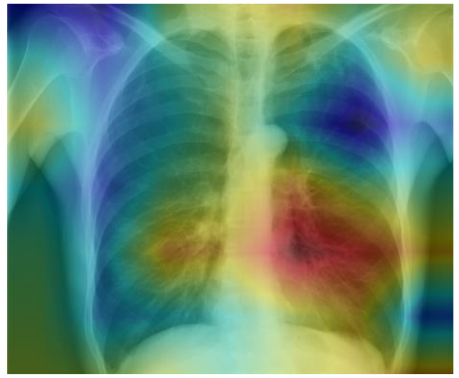
Radiology Masterclass. (n.d.). Chest X-ray — Cardiac disease: Pulmonary oedema

- Doctors face difficulties detecting faint opacities and distinguishing pneumonia from similar conditions, complicated by anatomical obstructions.
- Pneumonia can resemble a wide range of pulmonary and non-pulmonary conditions, including pulmonary edema, atelectasis, pulmonary hemorrhage, ARDS, pleural effusions, and certain lung cancers.
- Variations in imaging protocols, equipment quality, and acquisition standards across different hospitals and clinical settings result in highly inconsistent chest X-ray appearances, making reliable pneumonia detection more challenging

Input
Chest X-Ray Image

CheXNet
121-layer CNN

Output
Pneumonia Positive (85%)

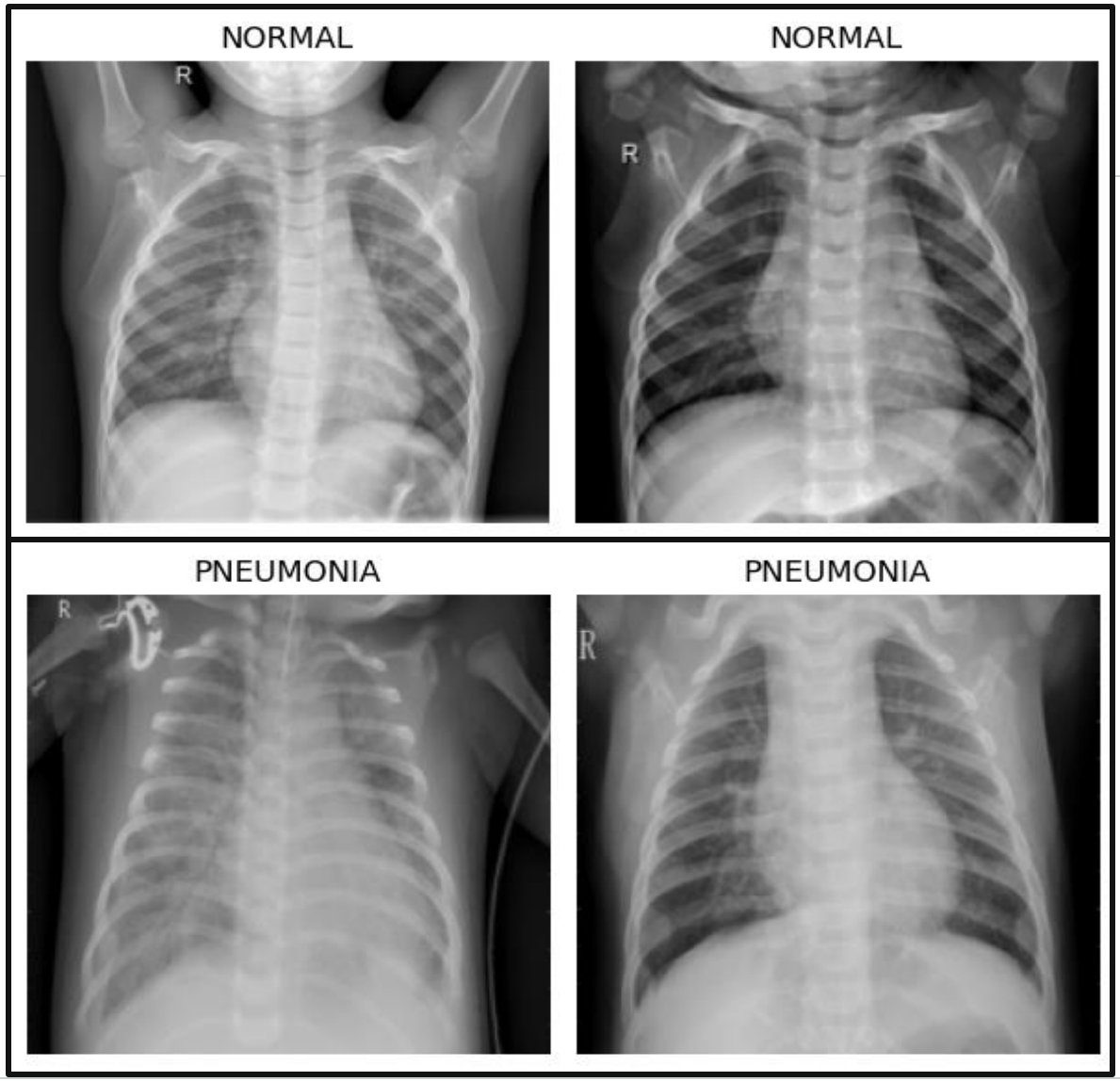


Clinical Importance of AI Medical Imaging

- AI models can detect subtle textural patterns and faint opacities that are often missed by clinicians, improving early and low-visibility pneumonia detection.
- Advanced feature extraction in CNN and transformer models helps distinguish pneumonia from visually similar conditions, even when patterns overlap or anatomy is complex.
- AI classifiers remain robust across variations in patient positioning, projection type, and image quality, helping normalize inconsistencies between hospitals and imaging equipment.

Stanford ML Group. (n.d.). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays

Dataset Overview



Dataset: Chest X-Ray Pneumonia

Source: Kaggle Chest X-Ray Pneumonia dataset

Task: Binary classification

- 0 = Normal, 1 = Pneumonia

Total images: ~5,000+ posterior–anterior

Train:

- Normal: 1341
- Pneumonia: 3875

Validation:

- 8 Normal, 8 Pneumonia (very small set)

Test:

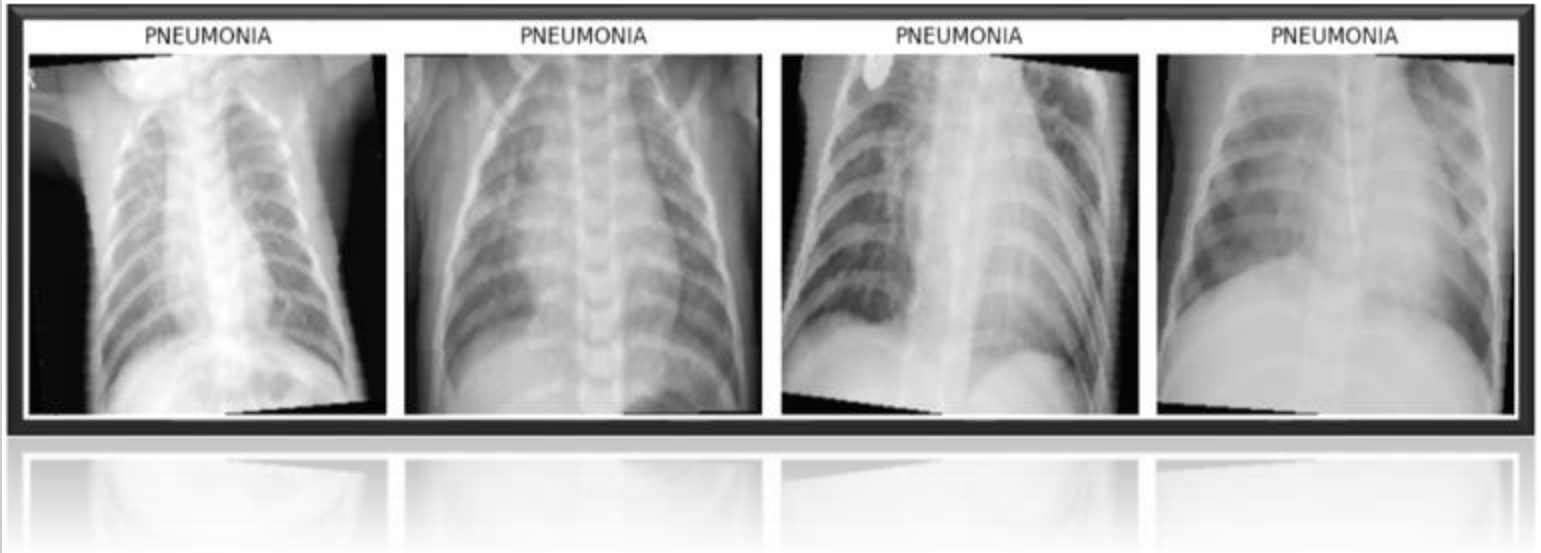
- 234 Normal, 390 Pneumonia

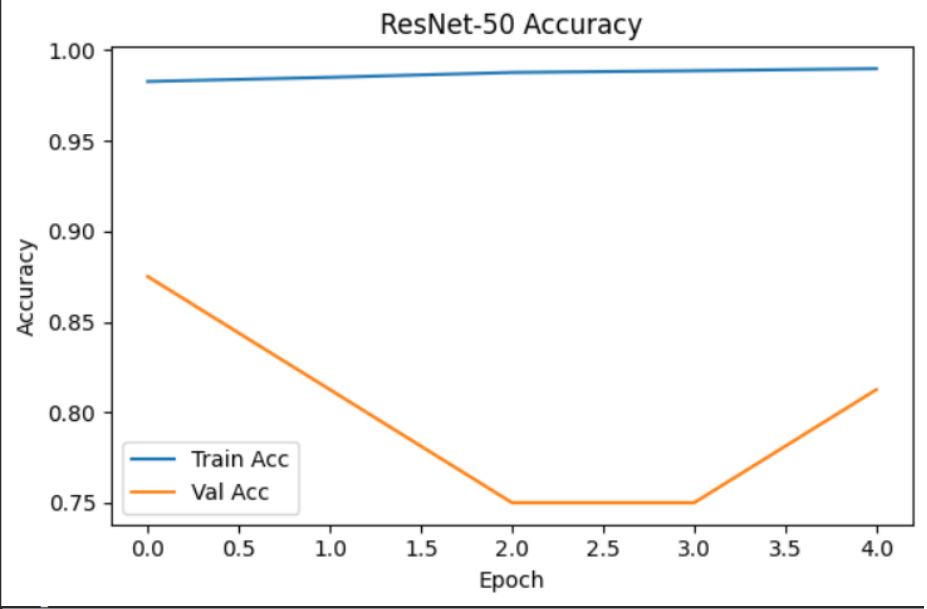
Initial Challenges:

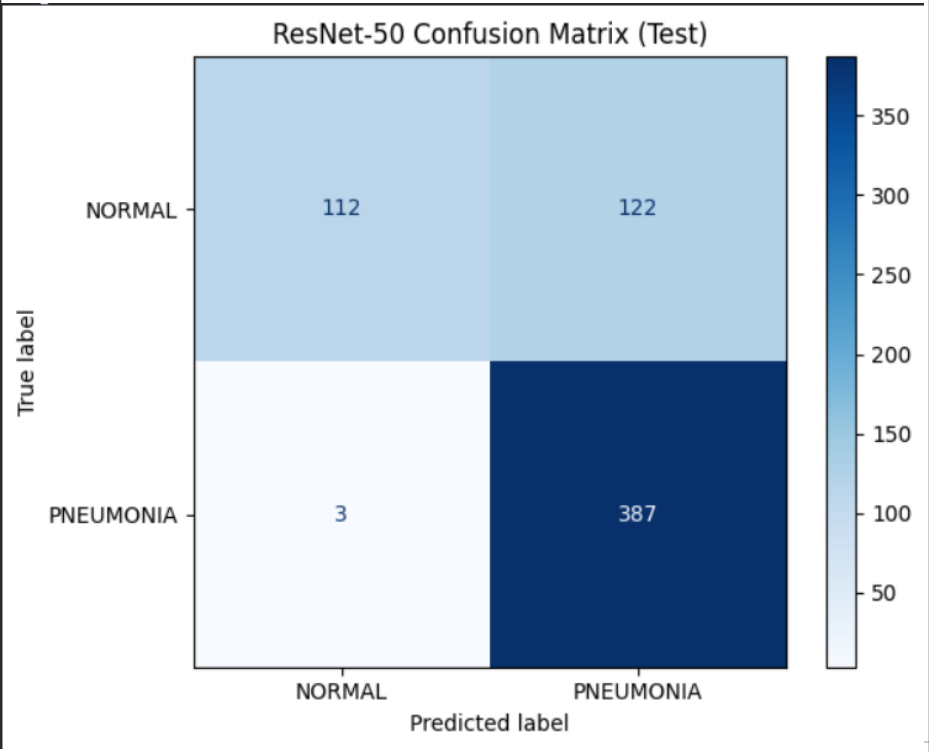
- Strong class imbalance in train set
  - Pneumonia images ≫ Normal images

Preprocessing & Augmentation

- **Training preprocessing:**
  - Resize to 224×224
  - Random horizontal flip
  - Random rotation (±10°)
  - Normalize with ImageNet mean/std
- **Validation & test preprocessing:**
  - Resize to 224×224
  - Normalize with ImageNet mean/std
  - (No augmentations applied)
- **Data loading:**
  - Batch size = 32
  - Training data shuffled each epoch
- **Dataset summary:**
  - Report number of train / val / test images
  - Two classes used: NORMAL and PNEUMONIA

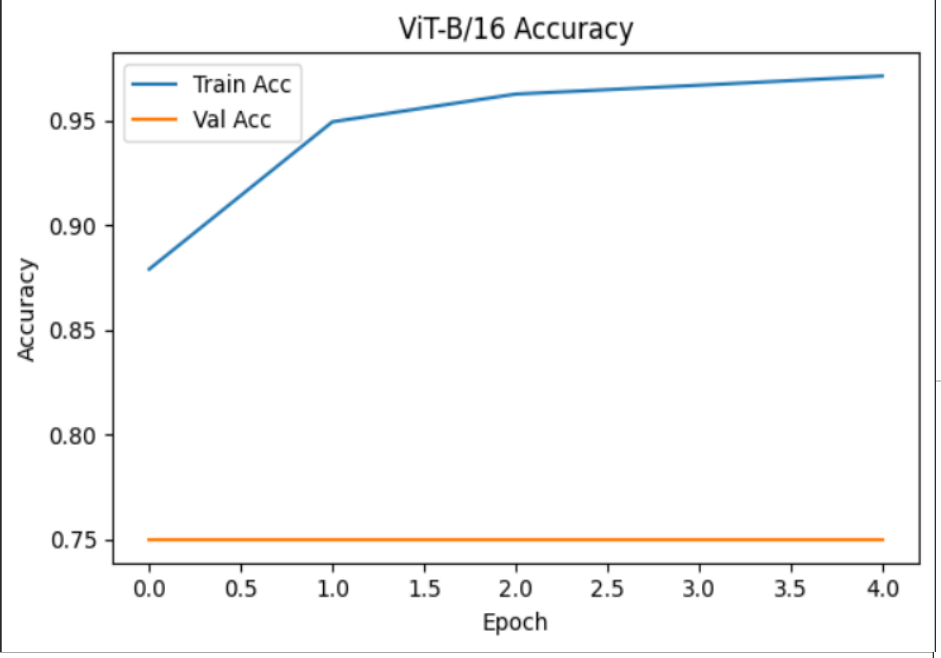


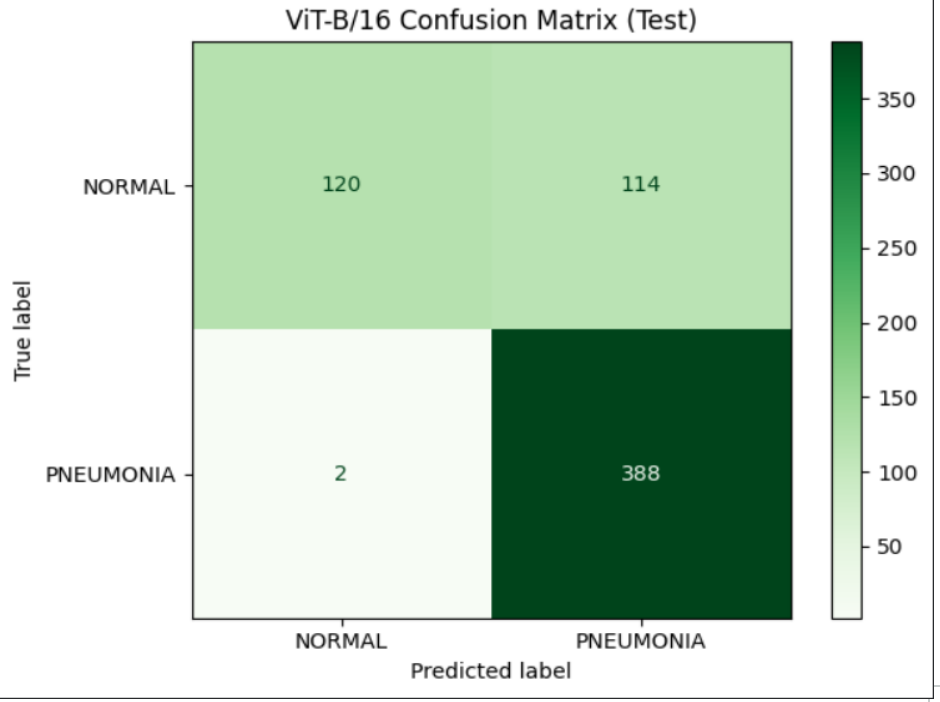




Proposed Model 1 : ResNet-50

- Model trained for multiple epochs on the training set
- Performance evaluated after each epoch on validation set
- Tracked both training and validation accuracy to monitor learning
- Training accuracy stays very high (~0.98+), showing the model learns patterns well
- Validation accuracy fluctuates (0.75–0.87), indicating limited validation data and slight overfitting
- Overall pattern shows the model generalizes reasonably but benefits from more balanced training and more epochs

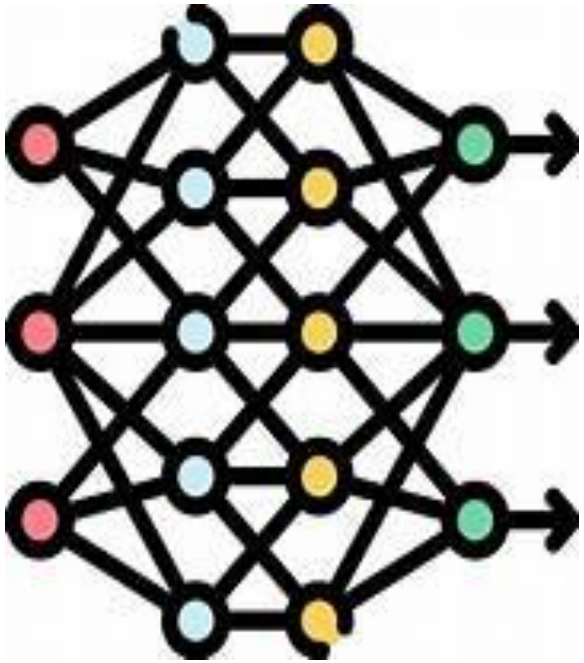




Proposed Model 2 : Vision Transformer (ViT-B/16)

- Trained ViT-B/16 for multiple epochs and saved the checkpoint with the **best validation accuracy**
- Training accuracy increased steadily (~97%), while validation accuracy stabilized around **75%**, showing mild overfitting
- Test confusion matrix shows:
  - **Excellent pneumonia detection:** 388/390 correctly classified
  - **Moderate normal detection:** 120/234 correctly classified
- Overall pattern: very high sensitivity for pneumonia, lower specificity for normal cases

Training Setup & How we used the Models



- Loss function: **Cross-Entropy Loss**
- Optimizer: **Adam**
- Batch size: **32**
- Different learning rates:
  - Smaller LR for pretrained backbone
  - Larger LR for new classification head
- Early experiments:
  - ~5 epochs, **original imbalanced training set**
- Later experiments:
  - Up to 30 epochs
  - **Balanced sampling** to correct class imbalance

Model Comparisons (ResNet50 Vs ViT-B/16)

Normal Class Performance

- **Precision:** Both models are extremely high (ViT: 0.98, ResNet: 0.97)
- **Recall:** ViT performs slightly better (0.51 vs 0.48)

  Both struggle with Normal class recall, but **ViT is slightly better**.

Pneumonia Class Performance

- **Precision:** Nearly identical (ViT: 0.77, ResNet: 0.76)
- **Recall:** Both achieve excellent sensitivity (0.99 for both)

  Both models are **very strong at detecting pneumonia**.

ResNet50 Test Performance

=== ResNet-50 Test Performance ===

	precision	recall	f1-score	support
NORMAL	0.97	0.48	0.64	234
PNEUMONIA	0.76	0.99	0.86	390
accuracy			0.80	624
macro avg	0.87	0.74	0.75	624
weighted avg	0.84	0.80	0.78	624

ViT-B/16 Test Performance

=== ViT-B/16 Test Performance ===

	precision	recall	f1-score	support
NORMAL	0.98	0.51	0.67	234
PNEUMONIA	0.77	0.99	0.87	390
accuracy			0.81	624
macro avg	0.88	0.75	0.77	624
weighted avg	0.85	0.81	0.80	624

Overall metrics and the Class Imbalances

- Overall Metrics (Best Model: ViT-B/16)

Accuracy: 0.81

Normal Class: Precision 0.98, Recall 0.51

Pneumonia Class: Precision 0.77, Recall 0.99

Weighted F1 Score: 0.80

ViT achieves the highest overall performance among both models

- Observed Problem: Class Imbalance

Dataset has **~3× more Pneumonia images** than Normal

Both models learn a strong bias toward **predicting Pneumonia**

High Pneumonia recall (0.99) but **poor Normal recall** (0.48–0.51)

Leads to many Normal images being misclassified as Pneumonia

Indicates the model is not seeing enough Normal examples during training

Fixing Class Imbalances in Training

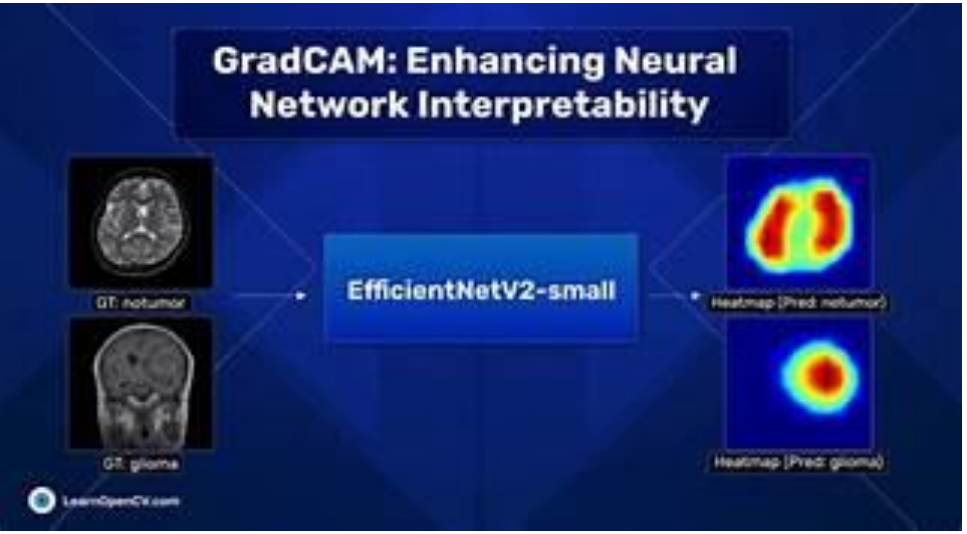
- **Problem:** training set has ~3 times more Pneumonia than Normal

- **Approach: WeightedRandomSampler**
- Computing inverse-frequency weights per class
- Normal samples get higher sampling probability
- Each batch becomes **approximately balanced**

Effect:

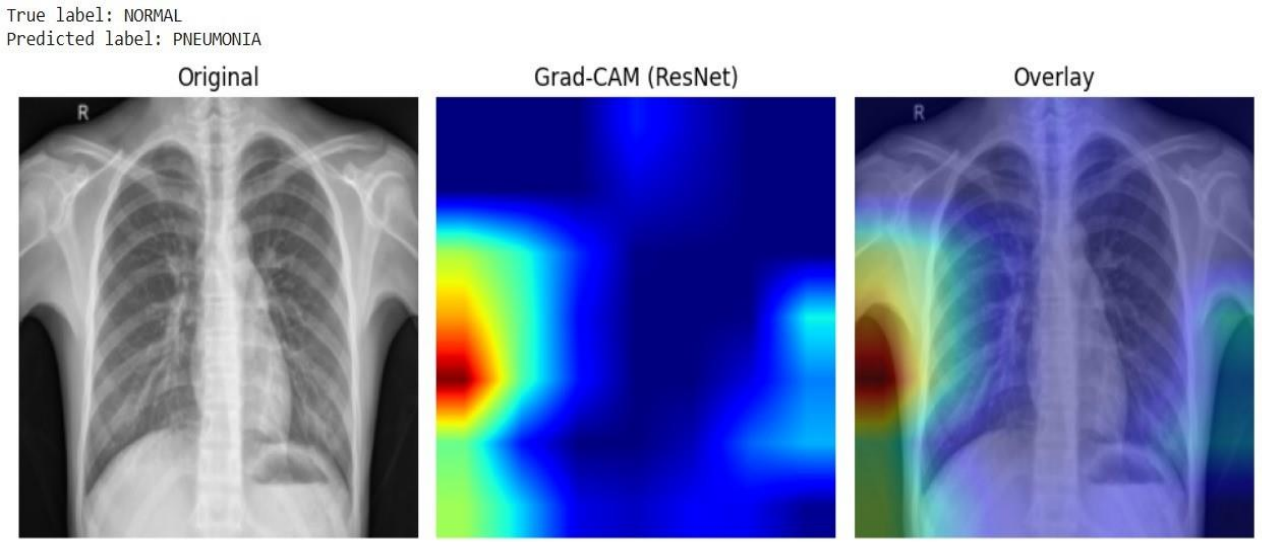
- Forces model to see Normal and Pneumonia equally often
- Reduces bias toward predicting “Pneumonia”

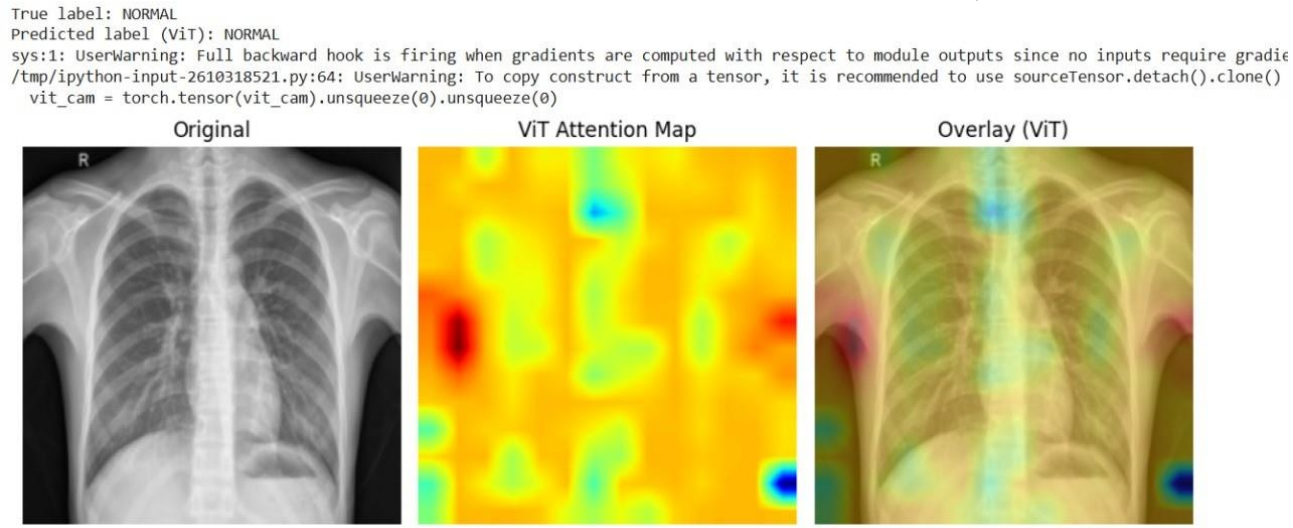
GRAD-CAM



- What is Grad-CAM?
- Grad-CAM (Gradient-weighted Class Activation Mapping) generates **heatmaps** showing which regions of the image influenced the model's prediction
- Uses gradients from the **last convolution layer** to highlight important areas
- Produces a visual map of **model attention** over the X-ray image
- Grad-CAM for Pneumonia Detection
- Highlights abnormal lung regions (e.g., opacities, infiltrates) that are consistent with pneumonia
- Shows whether a model prediction is **based on meaningful pathology**
- Builds clinician trust by explaining *why* the model predicted "Pneumonia"

GRAD-CAM Outputs for ResNet50 and Vision Transformer





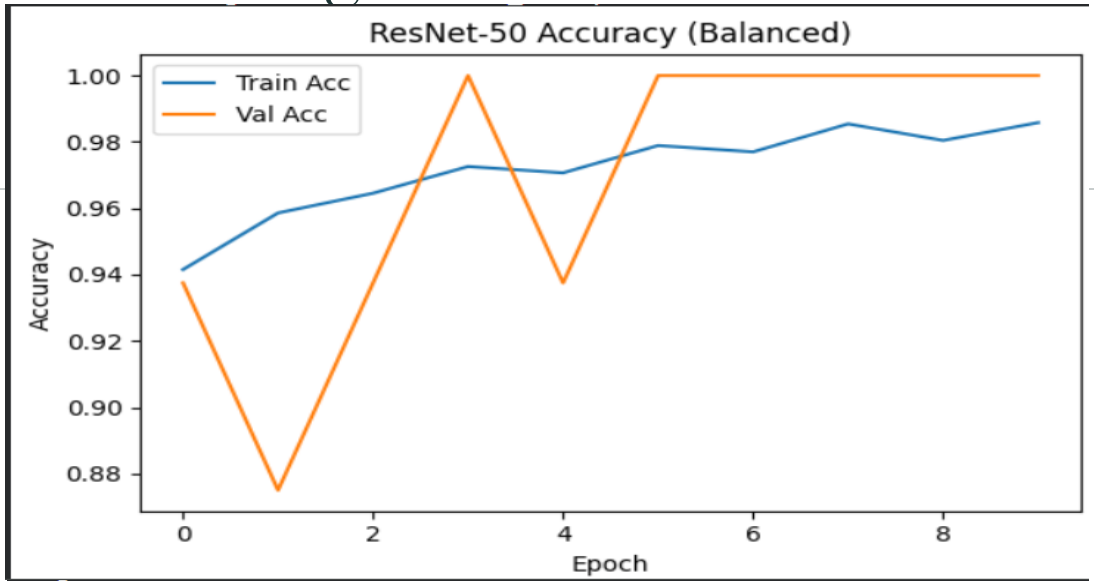
1. ResNet Grad-CAM Output

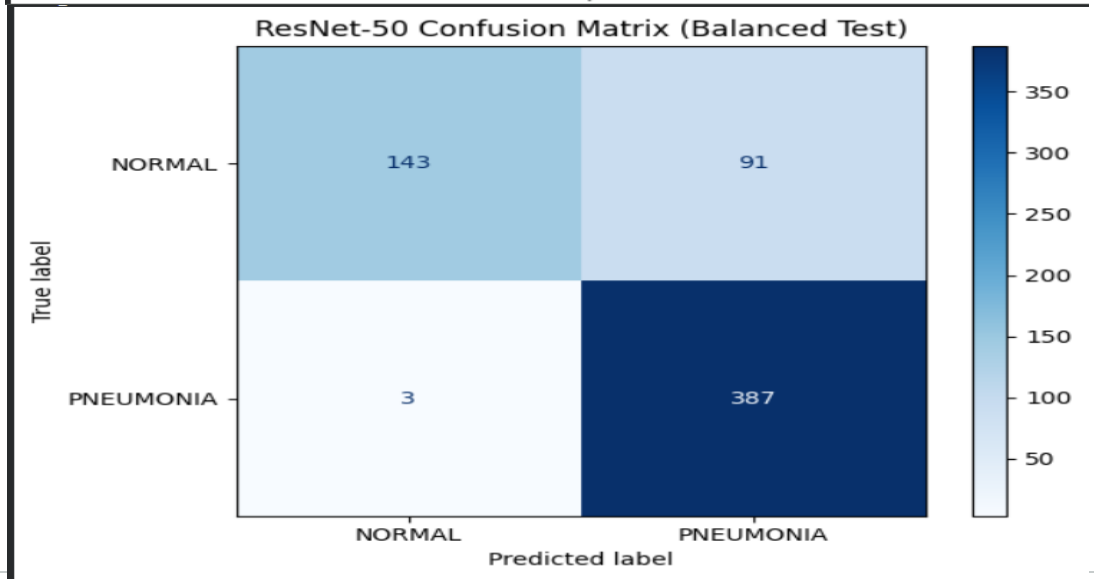
- True label: **NORMAL**, but model predicted **PNEUMONIA**
- Grad-CAM heatmap highlights a **strong red/yellow region on the left lung area**
- This indicates ResNet focused on a local shadow/texture that it interpreted as abnormal
- The model incorrectly treats this normal anatomical variation as a pneumonia pattern

2. ViT Attention Map Output

- True label: **NORMAL** and ViT correctly predicted: **NORMAL**
- Attention map shows **distributed, patch-based activation** across the lungs.
- The model highlights central lung fields, upper lobes, and mild edges consistent with **global feature extraction**.
- Suggests ViT is more robust and less influenced by noise or border effects compared to ResNet.

ResNet50 Performance after balancing the classes





1. Accuracy Curve (Balanced Training)

- Training and validation accuracy both improve and reach **~98–100%**, showing strong learning
- Validation accuracy becomes **much more stable** after balancing
- Indicates that balancing the dataset helped the model generalize better and reduced the earlier bias toward pneumonia

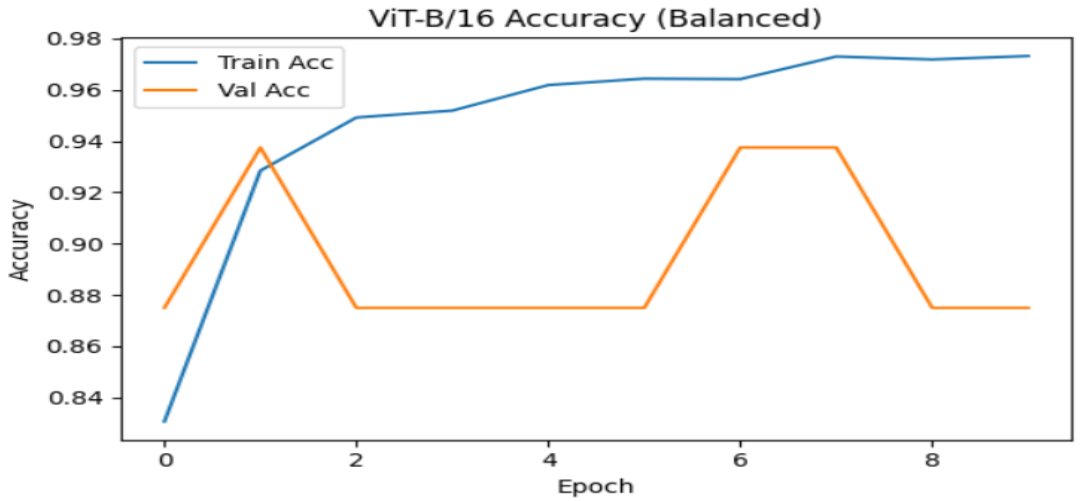
2. Confusion Matrix (Balanced Test Set)

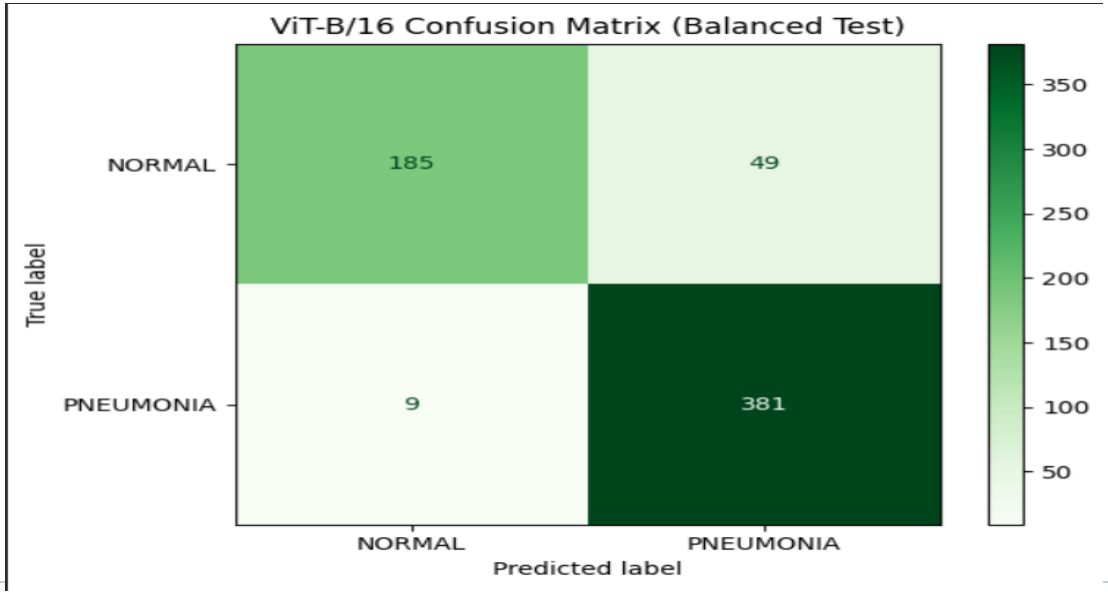
- **NORMAL:** 143 correctly classified, 91 misclassified
  - Normal recall improved significantly compared to imbalanced training
- **PNEUMONIA:** 387 correctly classified, only **3 errors**
  - Pneumonia detection remains extremely strong
- Overall: model now performs more **balanced** between the two classes, with fewer false pneumonia predictions

Key Takeaway

Balancing the dataset greatly reduced model bias and improved the model's ability to correctly recognize **NORMAL** chest X-rays while maintaining high pneumonia sensitivity.

ViT Performance after balancing the classes





1. Accuracy Curve (Balanced Training)

- Training accuracy steadily increases and stabilizes around **96–97%**
- Validation accuracy fluctuates but remains in the **88–94%** range
- Shows good learning but slightly more variance than ResNet after balancing
- Balancing improves stability and reduces earlier overfitting toward pneumonia

2. Confusion Matrix (Balanced Test Set)

- **NORMAL:** 185 correctly classified, 49 misclassified
  - Normal recall improved significantly compared to imbalanced ViT
- **PNEUMONIA:** 381 correctly classified, 9 misclassified
  - Pneumonia detection remains very strong despite balancing
- Much better balance between both classes compared to before balancing

Key Takeaway

Balancing the dataset helps ViT become **less biased** and improves its recognition of **NORMAL** images while keeping pneumonia performance high.

Model Comparisons (ResNet50 Vs ViT-B/16) Balanced

- **Normal Class Performance (Balanced)**
- **Precision:** ResNet = 0.98, ViT = 0.95
- **Recall:** ViT performs much better (0.79 vs 0.61)

ViT correctly identifies more NORMAL cases and is far less biased.

Pneumonia Class Performance (Balanced)

- **Precision:** ViT = 0.89, ResNet = 0.81
- **Recall:** Both remain very high (ViT = 0.98, ResNet = 0.99)

  Both models are still excellent at detecting pneumonia, with ViT giving fewer false alarms.

- Overall, ViT-B/16 provides **more balanced and reliable performance**, especially on NORMAL images.

ResNet50 on Balanced Dataset

```
=== ResNet-50 (Balanced) Test Performance ===
              precision    recall  f1-score   support

      NORMAL       0.98      0.61      0.75       234
   PNEUMONIA       0.81      0.99      0.89       390

    accuracy                           0.85       624
   macro avg       0.89      0.80      0.82       624
weighted avg       0.87      0.85      0.84       624
```

ViT on Balanced Dataset

```
=== ViT-B/16 (Balanced) Test Performance ===
              precision    recall  f1-score   support

      NORMAL       0.95      0.79      0.86       234
   PNEUMONIA       0.89      0.98      0.93       390

    accuracy                           0.91       624
   macro avg       0.92      0.88      0.90       624
weighted avg       0.91      0.91      0.90       624
```

Overall metrics and Drawbacks

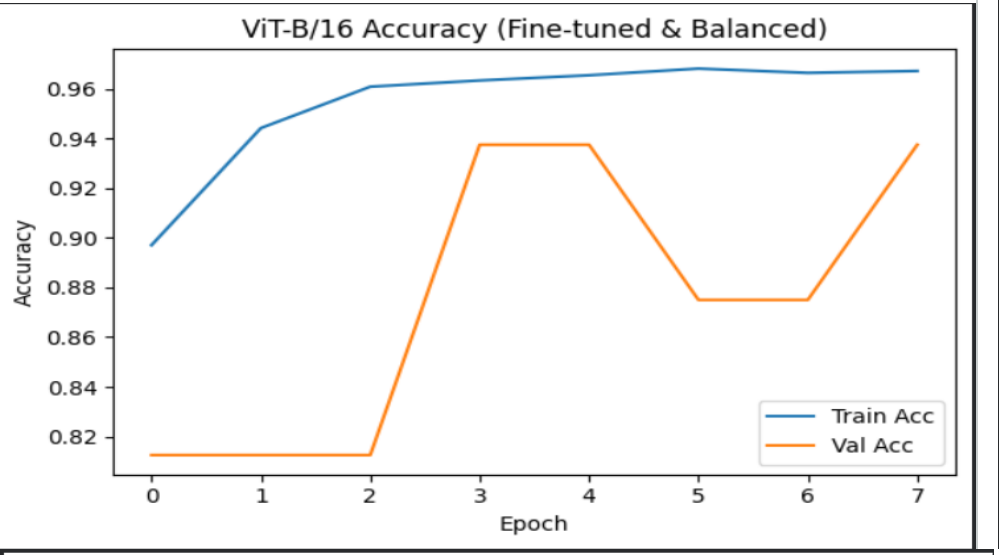
Overall Metrics (Balanced Models)

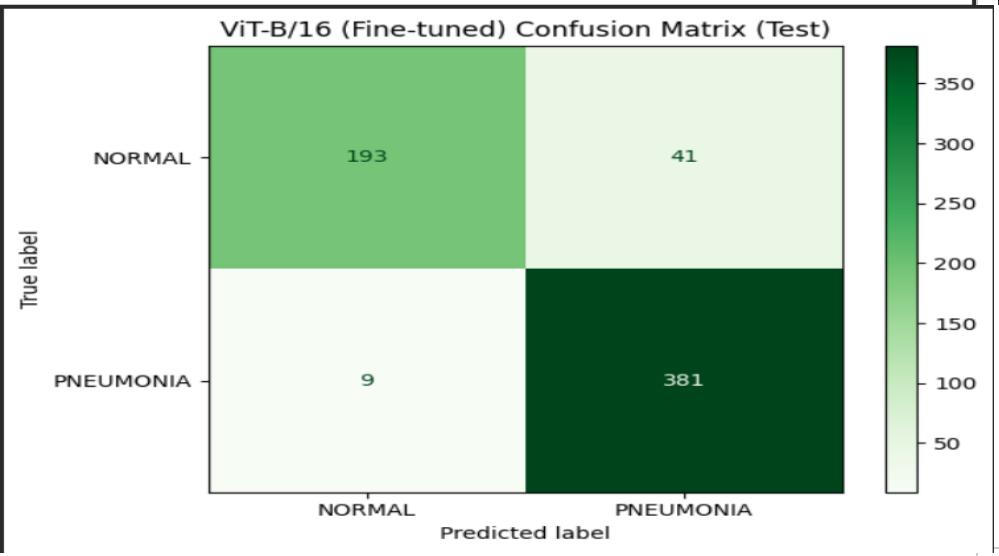
- Accuracy:
  - ViT-B/16: **0.90**
  - ResNet-50: **0.84**
- Macro Average (Precision / Recall / F1):
  - ViT-B/16: **0.92 / 0.88 / 0.90**
  - ResNet-50: **0.89 / 0.80 / 0.82**
- Weighted Average (Precision / Recall / F1):
  - ViT-B/16: **0.91 / 0.91 / 0.90**
  - ResNet-50: **0.87 / 0.85 / 0.84**

ViT achieves the highest overall performance among both models

- Observed Problem: Overfitting
- ViT is slightly overfitting when we applied the class balancing.

Finetuning the ViT Model






```
=== ViT-B/16 (Fine-tuned) Test Performance ===
              precision    recall  f1-score   support

      NORMAL       0.96      0.82      0.89       234
   PNEUMONIA       0.90      0.98      0.94       390

    accuracy                           0.92       624
   macro avg       0.93      0.90      0.91       624
weighted avg       0.92      0.92      0.92       624
```

ViT-B/16 Fine-Tuned Results (Final)

- **Normal:** Prec 0.96 | Rec 0.82 | F1 0.89
- **Pneumonia:** Prec 0.90 | Rec 0.98 | F1 0.94
- **Overall accuracy: 0.92**
- Best balance across both classes with highest F1-scores
- Most reliable and best-performing model in the pipeline
- **Before Finetuning, the accuracy was 0.907, but after it changed to 0.9199.**

THANK YOU.

SCAN ME

GitHub Link