

# **Deep Learning-Based Pneumonia Detection from Chest X-ray Images: A CNN Approach and Transfer Learning Approach**

Ayam Kattel

Electronics, Communication and Information Engineering

Pulchowk Campus

Email: 080bei011.ayam@pcampus.edu.np

December 29, 2025

## **Abstract**

Pneumonia is an inflammatory condition of the lung primarily affecting the alveoli. Each year, pneumonia affects about 450 million people globally (7% of the population) and results in about 4 million deaths. Chest X-ray scans are one of the common method for the diagnosis of pneumonia condition. This report presents an effective deep learning model to detect pneumonia based on the X-ray scans which could help radiologists in making better decisions. A supervised learning technique is proposed here which predicts the outcome depending on the dataset's image quality. Kaggle's dataset is used to train model first in own custom CNN model and then in pretrained model:ResNet18 via transfer learning. The comparision between both the models is briefly discussed later in this report. To improve the training in both the methods, the model was tuned with different hyperparameters. The results of both the models are compared using the accuracy and F1-score. The model's performance in detecting pneumonia demonstrates that the proposed ResNet18 pre-trained model can efficiently categorize normal and pneumoniachest X-rays images in practice. Therefore, the this model can be utilized to make a speedy diagnosis of pneumonia and can assist radiologists with their work.

## **1 Introduction**

Pneumonia, a severe respiratory infection characterized by inflammation of the lung's air sacs called as alveoli, is still one of the deadliest communicable diseases worldwide. In underdeveloped and developing countries like Nepal, pneumonia is a disastrous disease

that kills people like the elderly, and babies. Pneumonia can commonly be caused by bacteria or viruses in the air we breathe. Timely diagnosis of it is critical, as untreated cases can rapidly progress to respiratory failure. To solve this problem chest X-ray imaging serves as the primary diagnostic tool due to its affordability and widespread availability. Radiologists must distinguish subtle patterns such as interstitial opacities in the X-ray. Even among the experts , there is a high variation of ambiguity which leads to diagnostic inconsistency. This variability highlights the critical demand for systems that is unbiased and does automated evaluation.

This report presents a deep learning approach to solve this issue and predict pneumonia based on the image of the X-ray scans. Among five pre-trained models such as deep learning models such as ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3 ResNet18 was found to perform the best giving a good test accuracy of 90%. In addition to it our model aimed at maximizing the recall to ensure no sick patient is missed by penalizing by weight 1.2 more than that for the normal case. Preprocessing techniques like CLAHE is also highlighted in this report. Overall, this report is the comparision between Custom CNN and Resnet18 architechture with the constraints like performance,interpretability, robustness etc. to provide a robust diagnostic framework.

## 2 Dataset and Problem Statement

### 2.1 Problem Statement

The human based analysis of the medical images of patients to detect the Pneumonia may contain different subjective errors due to the expertise levels, fatigue and biasness etc. Also, if the number of patients are very high in number, the detection will be a very time consuming effort which might also cause mental fatigue. This stage is the most initial phase of the treatment of pneumonia. It would be great if this stage is automated and doctors can jump to next stage of the treatment so that they can save both the time and energy. This report presents an accurate and efficient deep-learning approach to automate the detection of Pneumonia by using medical image without subjective errors.

### 2.2 Dataset Description

The dataset is taken from kaggle [Chest X-Ray Images (Pneumonia)]. Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2

categories (Pneumonia/Normal). There is a class imbalance in the dataset in which pneumonia class is approximately 2.8 times more than that of normal class. The data distribution of normal and pneumonia class in training dataset is given below.

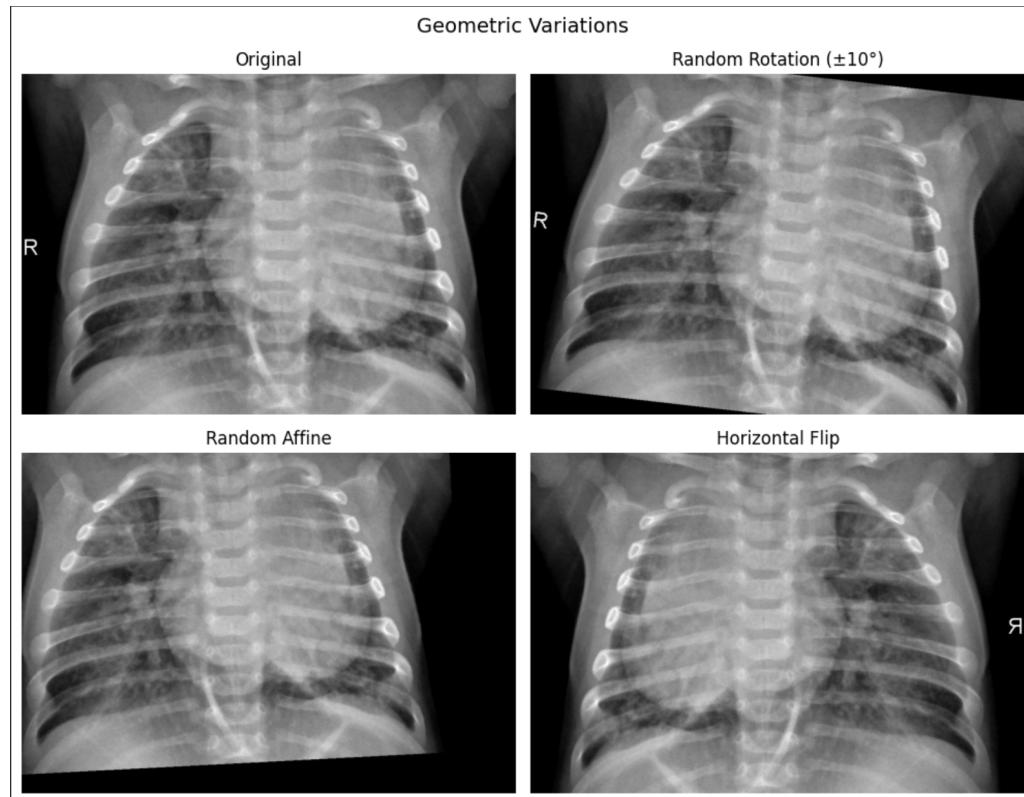
Normal	Pneumonia
1341	3875

Table 1: Class Distribution of Dataset

## 3 Data Preprocessing and Augmentation

### 3.1 Geometric Variations

- **Random Rotation and Affine Transforms** Images were randomly rotated by  $\pm 10^\circ$  and subjected to affine transformations (translation of 10% and scaling between 0.9 and 1.1). This ensures the model is invariant to slight tilts or variations in patient positioning during the X-ray procedure.
- **Horizontal Flipping:** This is applied to double the spatial variety of the dataset, helping the model learn features that are not dependent on the left-right orientation of the lungs. Internal covariate shift is mitigated during training.



## 3.2 Photometric and Contrast Enhancement

- **Color Jittering:** In this portion, the brightness and contrast was randomly adjusted by 20%. This causes variance in image quality so that it gives accurate result for different X-ray machine settings and exposure levels
- **Custom CLAHE Integration:** Contrast Limited Adaptive Histogram Equalization (CLAHE) is one of the most critical component in this pipeline with a clip limit of 2.0. Unlike regular histogram equalization, CLAHE operates on small local regions (boxes) to enhance local contrast without over-amplifying noise in homogeneous areas. This makes subtle pulmonary opacities much more distinct for the neural network.

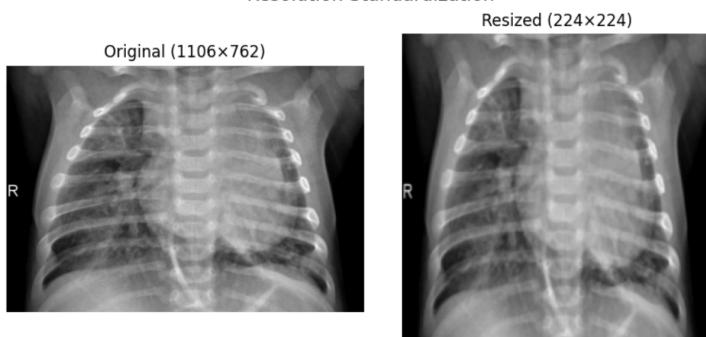
Photometric and Contrast Enhancement



## 3.3 Standardization, Normalization and Cropping

- **Resolution:** The images were then resized to 224x 224 pixels to maintain the balance between anatomical details and also to make computations more efficient.
- **Center Cropping:** A center crop of  $184 \times 184$  pixels was applied. This step ensures that the model focuses on the central lung part while discarding peripheral noise, uninformative borders, or artifacts often found at the edges of medical radiographs.
- **Input Normalization:** Pixel values were normalized using the standard ImageNet mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]). Doing this makes the input data look like ImageNet data to the model and also it improves the performance of model by ensuring similar scales for features.

Resolution Standardization



## 4 Model Selection

The selection of models is done by two approach,first by making model from scratch through custom architecture and second through transfer learning from ResNet's architecture. Both of the models are discussed below:

### 4.1 Custom Convolutional Neural Network

This architecture was designed as a light weight, modular architecture to determine the problem from scratch level. The main motive behind making this custom model is to observe the problem and inadequacy in this model and solve this inadequacy using various methods like transfer learning. This model is more flexible and transparent.

- **Convolution Layers:** The BaselineCNN consists of four convolutional blocks with increasing filter depths of 32, 64, 128, and 256, respectively. Each block employs  $3 \times 3$  convolutional kernels with appropriate padding to preserve spatial resolution. These layers progressively learn hierarchical spatial features, ranging from low-level edges to higher-level anatomical patterns relevant for pneumonia detection. ReLU activation is applied after each convolution to introduce non-linearity, enabling the model to capture complex feature relationships.
- **Batch Normalization:** Each convolutional layer is followed by batch normalization. This normalizes intermediate activations across the mini-batch. This also improves training stability and reduces internal covariate shift.
- **Max Pooling Layers:** A  $2 \times 2$  max pooling operation is applied after each block to downsample feature maps. This reduces spatial resolution from  $224 \times 224$  to  $14 \times 14$  while retaining dominant features and improving computational efficiency.
- **Global Average Pooling:** A Global Average Pooling layer replaces traditional flattening. This compresses the final  $14 \times 14 \times 256$  feature map into a 256-dimensional vector. This reduces parameter count and enhances spatial invariance.
- **Dropout and Output Layers:** A dropout rate of 0.7 is applied for regularization. This is followed by the output layer for binary classification (pneumonia vs. normal).

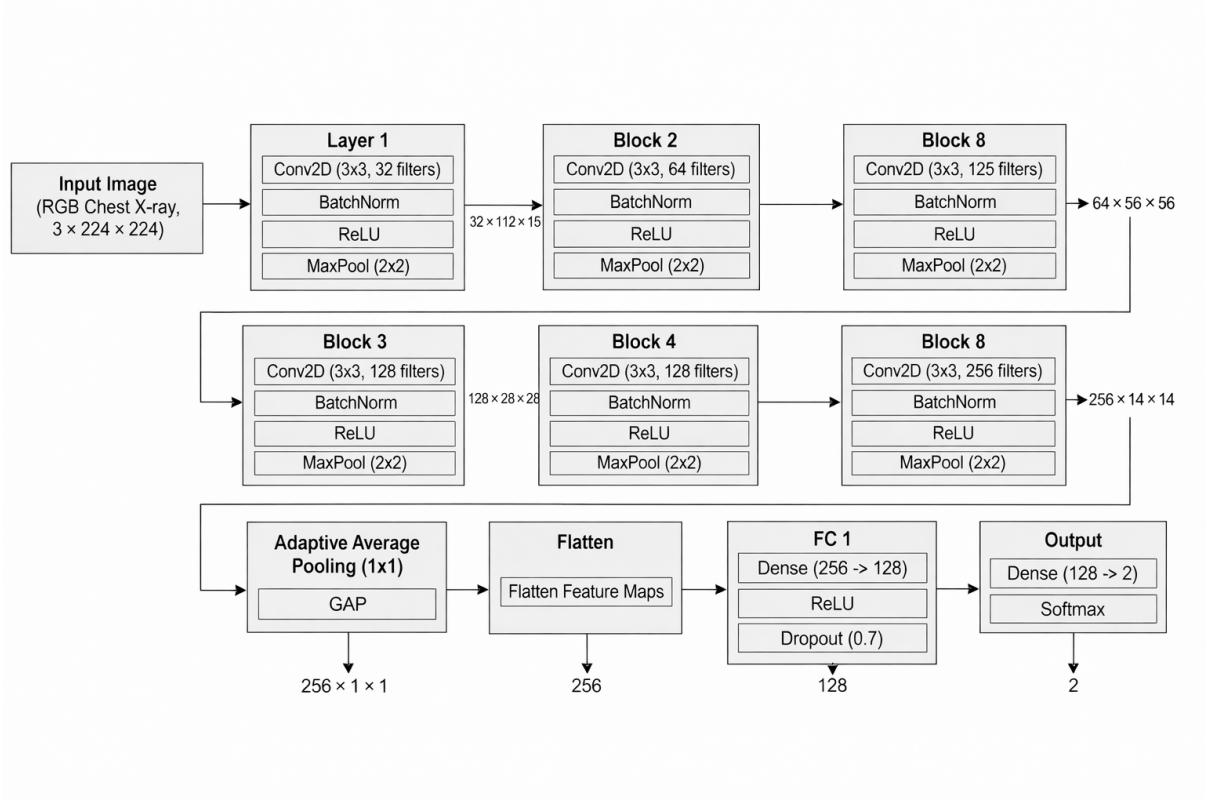


Figure 1: Fig: Architecture of Custom Baseline CNN

Table 2: BaselineCNN Architecture Description

Block	Layer Details	Output Size
Input	RGB Chest X-ray Image	$3 \times 224 \times 224$
Block 1	Conv2D ( $3 \times 3$ , 32 filters) + BatchNorm + ReLU + MaxPool (2 × 2)	$32 \times 112 \times 112$
Block 2	Conv2D ( $3 \times 3$ , 64 filters) + BatchNorm + ReLU + MaxPool (2 × 2)	$64 \times 56 \times 56$
Block 3	Conv2D ( $3 \times 3$ , 128 filters) + BatchNorm + ReLU + MaxPool (2 × 2)	$128 \times 28 \times 28$
Block 4	Conv2D ( $3 \times 3$ , 256 filters) + BatchNorm + ReLU + MaxPool (2 × 2)	$256 \times 14 \times 14$
GAP	Adaptive Average Pooling ( $1 \times 1$ )	$256 \times 1 \times 1$
Flatten	Flatten Feature Maps	256
FC 1	Dense ( $256 \rightarrow 128$ ) + ReLU + Dropout (0.7)	128
Output	Dense ( $128 \rightarrow 2$ ) + Softmax	2

## 4.2 Transfer Learning using ResNet18

Although other models (e.g.,VGG, DenseNet, ResNet50) also provide strong feature extraction, ResNet18 is lighter and faster to train, reducing computational cost and overfitting risk. This made me choose this model for performing transfer learning.ResNet18

also employs residual connections, enabling deep networks to be trained without suffering from vanishing gradients.

**Pre-trained Weights:** This model uses weights pre-trained on ImageNet-1K. Even though this dataset differs from medical images, the early layers have learned fundamental visual features such as edges and textures which are transferable

**Partial Freezing and Fine-Tuning:** To perform the transfer learning, earlier layers were freezed(made the parameters untrainable) and latter layers were added manually and trained for the classification.

- **Frozen Layers:** Layers 1–3 are frozen, preventing weight updates. This preserves low-level feature detectors and speeds up training.
- **Trainable Layer:** The final residual block (layer4) remains trainable, allowing the network to learn high-level, task-specific features.

**Custom Classification Head:** The original 1000 class output layer is replaced with a binary classification . It consists of a 512 node dense layer, Batch Normalization, and Dropout (0.5), connecting ResNet’s feature maps to the binary output for pneumonia detection.

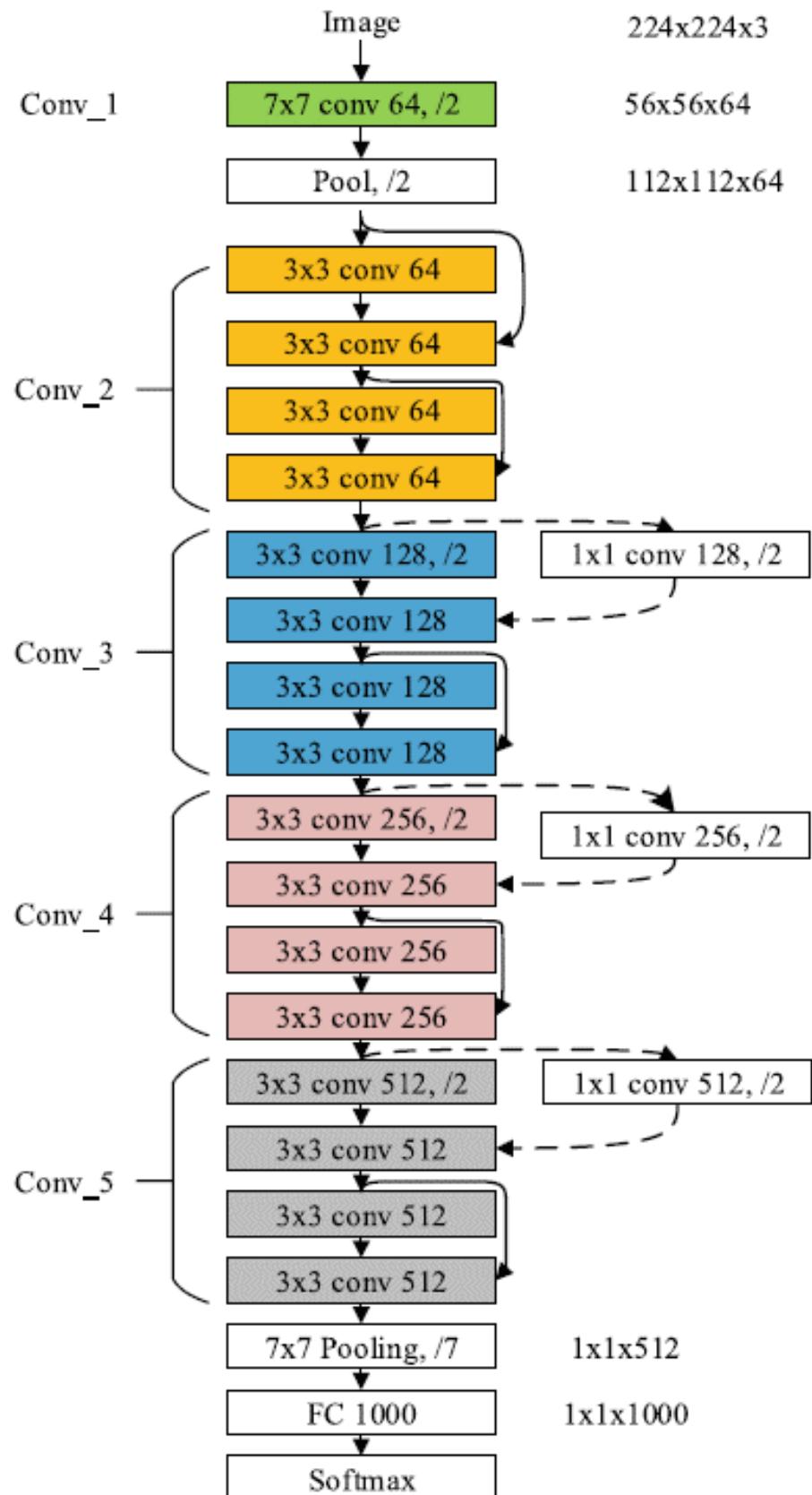


Figure 2: Fig: Architecture of ResNet18

Table 2: ResNet18 Transfer Learning Architecture for Pneumonia Detection

Stage	Layer Details	Output Size
ResNet18 Base	Pre-trained ResNet18 up to GAP (Conv1 + Residual Blocks 1–4)	512
FC 1	Dense ( $512 \rightarrow 512$ ) + BatchNorm + ReLU + Dropout (0.5)	512
Output	Dense ( $512 \rightarrow 2$ )	2

## 5 Training Information

### 5.1 Training Setup

The model was developed and trained using the PyTorch deep learning framework. The training environment was hosted on Google Colab, utilizing its cloud-based infrastructure for execution. To accelerate the training process, GPU acceleration (CUDA) was integrated via the `torch.device` configuration, allowing for efficient processing of large image batches when a compatible NVIDIA GPU was available.

### 5.2 Hyperparameters

The training process utilized a set of fine-tuned hyperparameters to ensure model stability and convergence:

Table 3: Hyperparameter Comparison: ResNet18 (Transfer Learning) vs Custom CNN (Baseline)

Hyperparameter	ResNet-18 (Transfer Learning)	Custom CNN (Baseline)
Optimizer	Adam (Weight Decay: $1 \times 10^{-3}$ )	Adam (Weight Decay: $1 \times 10^{-5}$ )
Learning Rate	Differential: $1 \times 10^{-5}$ (Backbone), $1 \times 10^{-4}$ (FC)	0.001 (Static initial rate)
Batch Size	32	32
Epochs	10 (with Early Stopping)	10 (with Early Stopping)
Loss Function	BCEWithLogitsLoss (weight=2.0)	BCEWithLogitsLoss (weight=1.2)
LR Scheduler	ReduceLROnPlateau (Patience: 3)	ReduceLROnPlateau (Patience: 3)
Dropout Rate	0.5	0.7

### 5.3 Training Strategy

The training strategy was centered on two models :Custom CNN model and pretrained ResNet-18 architecture and both the models are compared.

**Custom CNN Architecture:** The baseline model was built using four convolutional blocks, each containing a  $3 \times 3$  convolution, Batch Normalization, ReLU activation, and Max-Pooling. This was followed by Global Average Pooling and a dense classification head.

**ResNet-18 (Transfer Learning):** This model leveraged weights pre-trained on ImageNet. The strategy involved freezing the initial layers and fine-tuning only the final residual block (layer4) and the custom classification head to adapt to the specific nuances of X-ray imagery.

**Data Augmentation:** Both models shared a common augmentation pipeline to ensure a fair comparison. This included CLAHE (Contrast Limited Adaptive Histogram Equalization) for contrast enhancement, alongside random rotations ( $10^\circ$ ), horizontal flips, and affine transformations.

**Class Imbalance Handling:** To address the higher frequency of "Pneumonia" cases, both models used weighted loss functions. ResNet-18 used a higher weight (2.0) to prioritize recall, while the Custom CNN used a weight of 1.2.

**Regularization:** Dropout and Weight Decay were implemented in both models to prevent overfitting, with the Custom CNN utilizing a higher dropout rate (0.7) to compensate for training from scratch on a medium-sized dataset.

## 6 Evaluation Metrics

To objectively compare the performance of the ResNet-18 and Custom CNN models, the following evaluation metrics were utilized:

- **Accuracy:** Measures the overall percentage of correct predictions. While it is a standard evaluation metric, it is interpreted with caution due to the class imbalance present in the pneumonia dataset.
- **Recall (Sensitivity):** This is the primary metric for this medical application. It quantifies the model's ability to correctly identify all actual pneumonia cases. A high recall ensures that very few pneumonia patients are misclassified as *Normal*, which is critical for patient safety.

- **F1-Score:** The harmonic mean of Precision and Recall. This metric evaluates the balance between the two, ensuring that high recall is not achieved merely by over-predicting the *Pneumonia* class.
- **Area Under the ROC Curve (AUC):** Measures the model's ability to discriminate between classes across all possible probability thresholds. A higher AUC indicates a more robust model capable of effectively separating healthy lungs from pneumonia-affected lungs regardless of the decision boundary.

## 7 Model Comparison

Table 4: Performance Comparison Between Custom CNN and ResNet18

Metric	Custom CNN (Baseline)	ResNet-18 (Transfer)	Improvement
Training Accuracy	92.60%	97.49%	+4.89%
Test Accuracy	85.26%	91.19%	+5.93%
Training AUC	0.9737	0.9975	+0.0238
Convergence	Early Stop at Epoch 18	Early Stop at Epoch 8	ResNet more stable

## 8 Model Performance Interpretation

This section interprets our model and ensures that the prediction is done by the specific property of lungs in the lungs rather than noise or any other unwanted parameter in the X-ray.

### 8.1 Visualization Techniques (Grad-CAM)

To understand the underlying reason behind the prediction of model, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. Grad-CAM generates heat maps that highlight the spatial regions of the chest X-ray images that most strongly influence the model's classification decisions

**ResNet-18 Interpretation:** The activation maps produced by the ResNet-18 model predominantly exhibit high intensity over the pulmonary parenchyma (lung tissue). In pneumonia positive cases, the highlighted regions correspond closely with areas of increased opacity. This indicates that the model has learned clinically relevant visual features associated with pneumonia, enhancing its interpretability and clinical reliability.

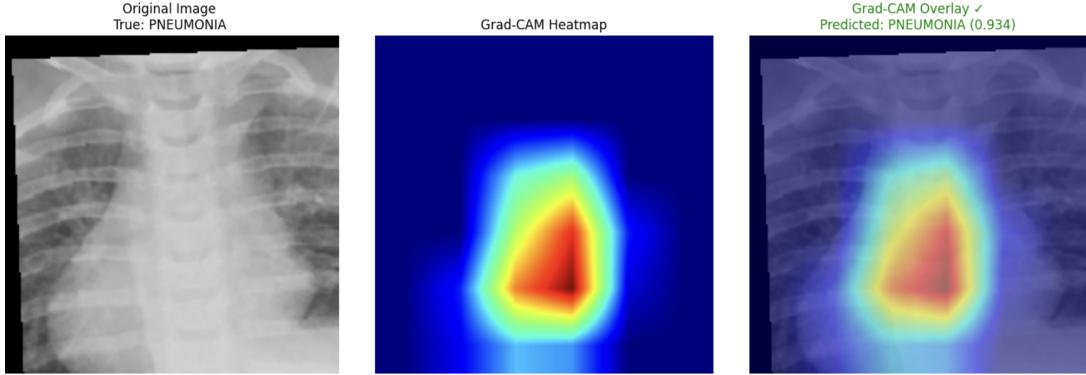


Figure 3: ResNet-18 Grad-CAM visualization for a pneumonia-positive case .

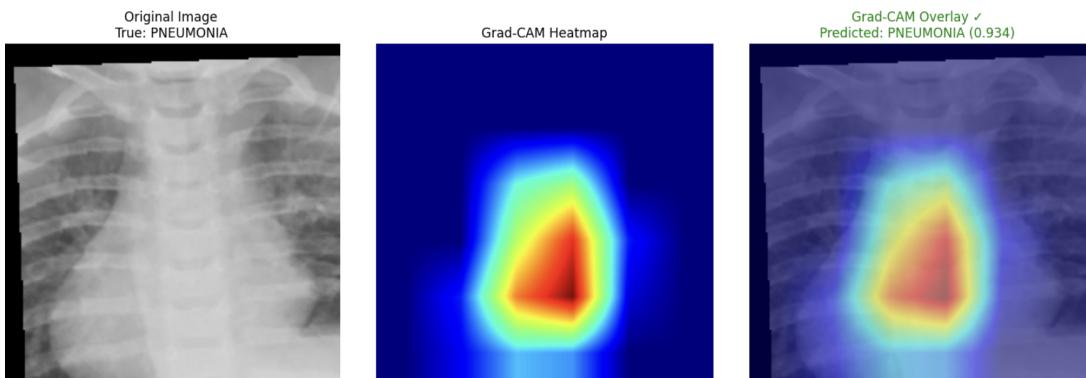


Figure 4: ResNet-18 Grad-CAM visualization for a pneumonia-positive case .

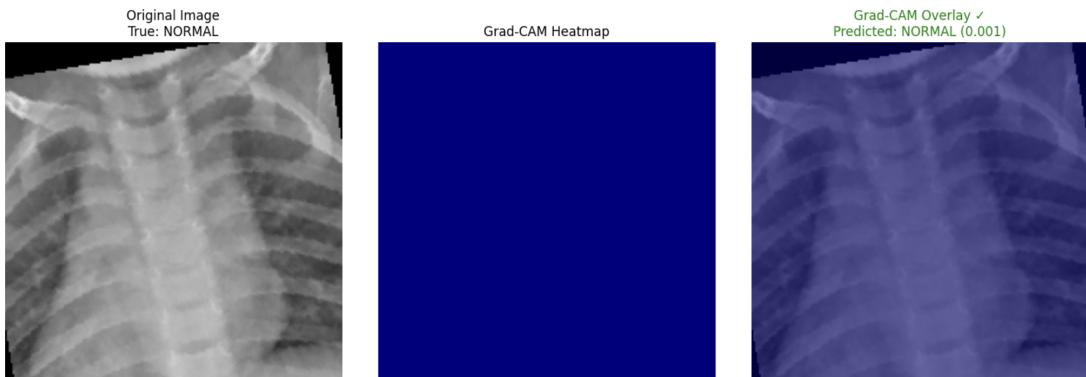


Figure 5: ResNet-18 Grad-CAM visualization for a normal chest X-ray.

**Custom CNN Interpretation:** In contrast, the activation maps of the Custom CNN appear more spatially dispersed. Although the model generally identifies the lung regions, it occasionally focuses on peripheral anatomical structures such as the rib cage or diaphragm. This behavior suggests that, despite achieving reasonable numerical performance, the Custom CNN lacks the architectural depth and specialization of ResNet-18, making its learned representations less medically focused and slightly less reliable for diagnostic purposes.

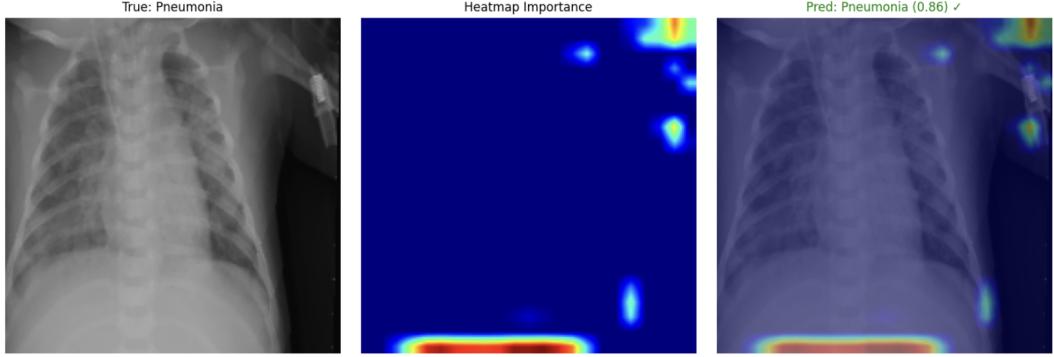


Figure 6: ResNet-18 Grad-CAM visualization for a pneumonia-positive case .

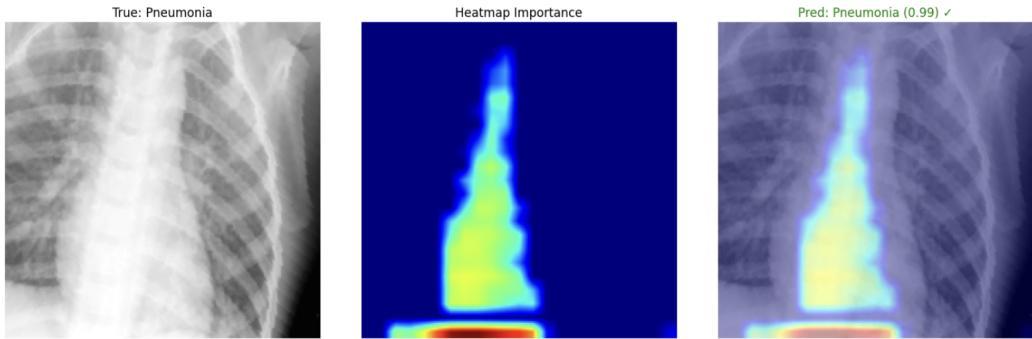


Figure 7: Custom model Grad-CAM visualization for a pneumonia-positive case.

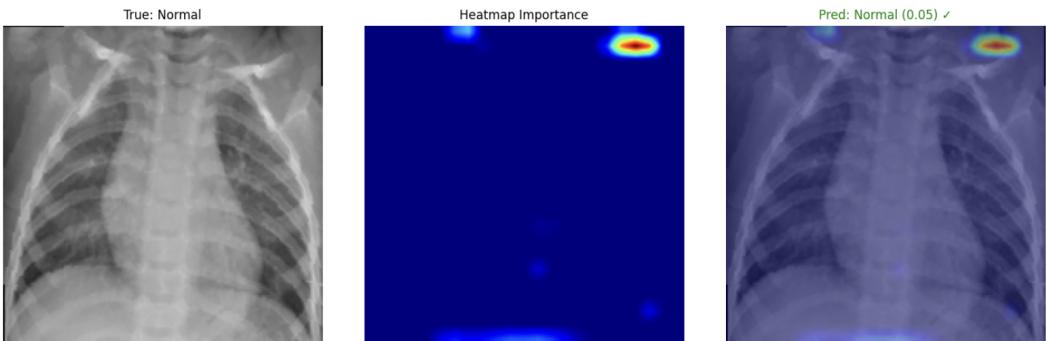


Figure 8: Custom Grad-CAM visualization for a normal chest X-ray.

## 8.2 Loss Curves and Convergence

Loss curves were analyzed to visualize the learning dynamics of both models and to detect potential issues such as overfitting or underfitting during training.

**Loss Curve Analysis:** For both architectures, the training loss demonstrated a consistent downward trend, indicating effective learning. However, the validation loss of the ResNet-18 model was notably more stable and converged more rapidly. In contrast, the

Custom CNN exhibited noticeable fluctuations in validation loss, a common characteristic of models trained from scratch on relatively small medical datasets. Early stopping was applied to mitigate overfitting, triggering at Epoch 8 for the ResNet-18 model and Epoch 6 for the Custom CNN, preserving generalization to unseen patient data.

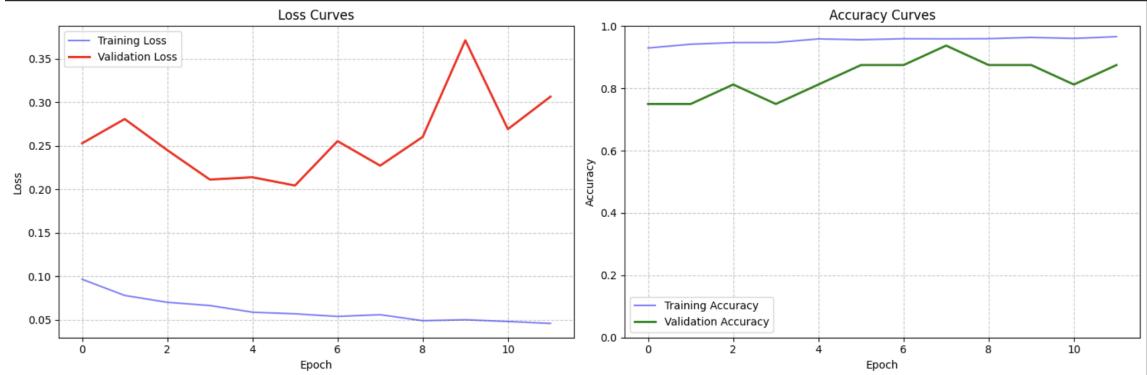


Figure 9: Loss Curve of ResNet model

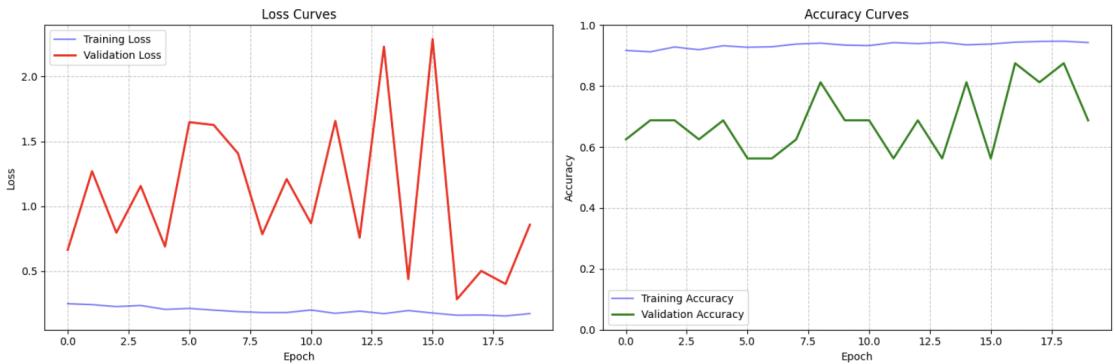


Figure 10: Loss Curve of Custom Model

### 8.3 Confusion Matrices and Error Analysis

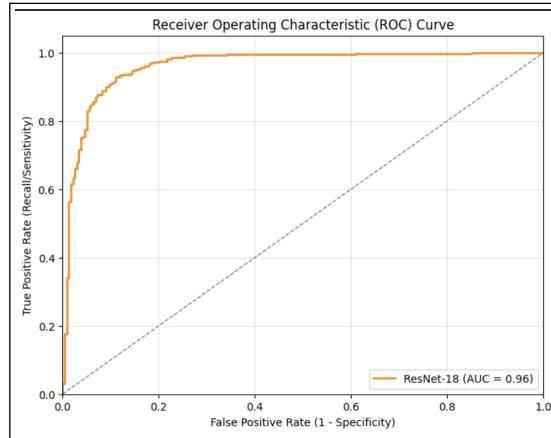
Confusion matrix analysis was conducted to obtain a detailed understanding of the classification performance and error patterns of both models.

**True Positives (Recall):** The ResNet-18 model achieved a very high true positive rate, demonstrating superior sensitivity in identifying pneumonia cases. In a clinical setting, this is particularly critical, as it minimizes false negatives and reduces the risk of missed diagnoses.

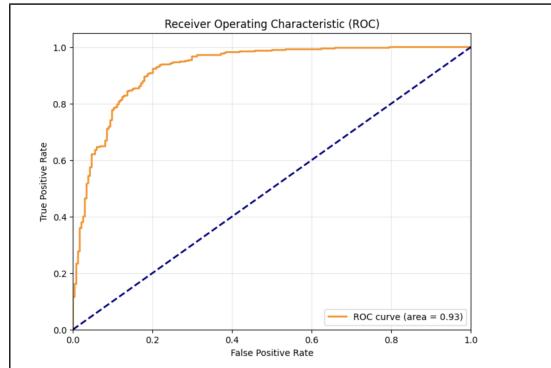
**False Positives:** The Custom CNN produced a higher number of false positives, incorrectly classifying some normal cases as pneumonia. While this may lead to unnecessary follow-up examinations, it is generally considered less severe than false negatives in medical diagnostics, though it reflects lower precision.

**Reliability:** The higher Area Under the Curve (AUC) score achieved by the ResNet-18 model (0.99) compared to the Custom CNN (0.97) further confirms its robustness

across different decision thresholds. This superior discriminative capability makes ResNet-18 a more suitable candidate for deployment in a clinical decision support system.



(a) ROC Curve of ResNet-18



(b) ROC Curve of Custom CNN Model

Figure 11: Comparison of ROC curves for ResNet-18 and the proposed Custom CNN model.

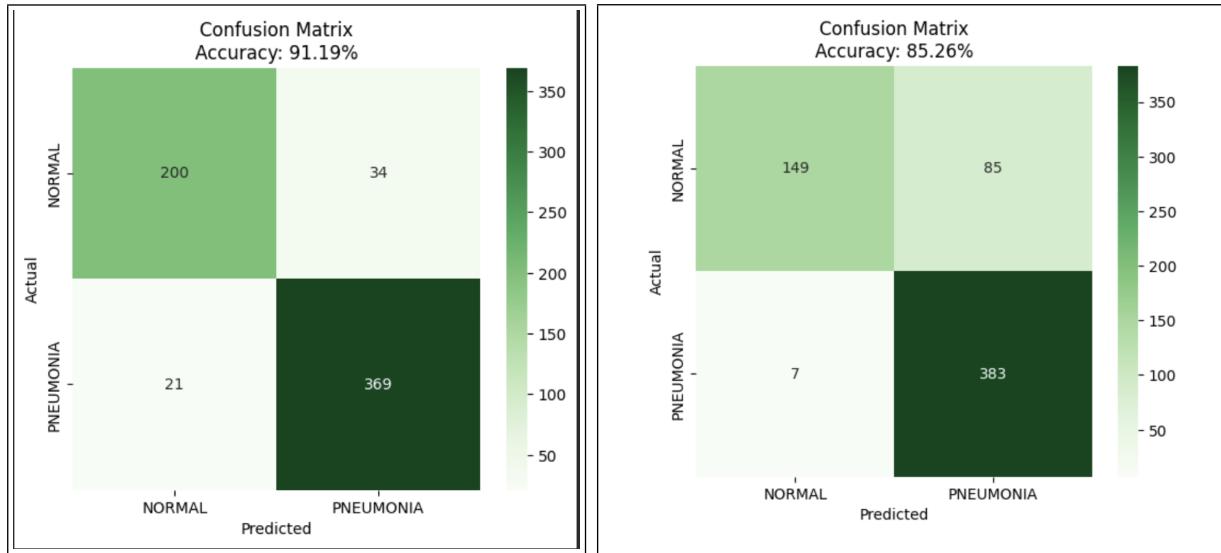


Figure 12: Confusion matrix comparison between ResNet-18 and Custom CNN models

## 9 Additional Observations

### 9.1 Impact of Class Imbalance

The dataset faced a noticeable class imbalance, with a higher number of pneumonia cases compared to normal cases. This imbalance biased the models toward predicting the pneumonia class. This lead to high recall but reduced precision for the normal class. Consequently, several healthy chest X-ray samples were incorrectly classified as pneumonia. This highlighted the influence of skewed class distributions on model behavior.

### 9.2 Transfer Learning vs. Training from Scratch

A clear performance advantage was observed for the ResNet-18 model initialized with ImageNet pre-trained weights when compared to the Custom CNN trained from scratch. The transfer learning approach converged more rapidly and achieved higher overall accuracy. This indicates that the low-level visual features, such as edges, textures, and basic anatomical structures, learned from natural images are effectively transferable to chest X-ray classification tasks.

### 9.3 Overfitting in Later Epochs

After approximately 10–15 training epochs, the training loss continued to decrease while the validation loss either stabilized or slight increase sometimes. This divergence between training and validation performance showed a bit overfitting. This suggested that the

models began to learn dataset-specific noise rather than generalizable features related to lung pathology.

## 9.4 Model Interpretability (Grad-CAM)

Grad-CAM analysis revealed that correct predictions focused on clinically relevant regions, including lung opacities. However, in certain misclassified cases, the attention maps highlighted non-pathological regions such as image borders or radiographic markers (e.g., “L” or “R”). This behavior suggests that the models occasionally relied on noise instead of meaningful pathological features, underscoring the importance of interpretability analysis in medical imaging applications.

## 10 Conclusion

This study highlights the effectiveness of transfer learning in medical imaging like detecting pneumonia from chest X-rays. Also the drastic change in the accuracy was observed in two different models. The ResNet18’s model performed very good with a test accuracy of 91.19% but the custom CNN model lacked behind in the test accuracy of 85.26%. The proposed model achieved a test accuracy of 91.19% and an AUC score of 0.95, indicating its robust capability to distinguish between pneumonia and normal cases. The model’s high recall rate of 96% for pneumonia detection underscores its strength in minimizing false negatives, which is crucial in clinical environments where early diagnosis directly correlates with improved patient outcomes. By automating the detection process, this AI-driven approach alleviates the diagnostic burden on radiologists, enabling faster analysis and more consistent results. Despite the promising results, the study also identifies areas for improvement. The model exhibited a slight tendency to misclassify normal cases as pneumonia, emphasizing the need for further refinement. Nevertheless, project highlights the power of transfer learning models in medical diagnostics and sets the stage for future advancements that could lead to the widespread adoption of AI-assisted diagnostic tools, ultimately improving healthcare delivery and patient outcomes on a global scale.

## References

- [1] R. Mahum et al., “A Novel Real-Time Framework for Automated Detection and Classification of Pneumonia from Chest X-Ray Images Using Deep Learning Techniques,” *Diagnostics*, vol. 11, no. 10, p. 1891, Oct. 2021. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8423280/>
- [2] N. Gautam and S. Maharjan, “Pneumonia Detection using Convolutional Neural Network,” in *Proc. 12th IOE Graduate Conference*, vol. 12, pp. 1222-1230, 2022.

- [Online]. Available: <http://conference.ioe.edu.np/publications/ioegc12/IOEGC-12-179-12260.pdf>
- [3] “Deep Learning for Medical Imaging Analysis,” *arXiv preprint arXiv:2510.00035*, 2025. [Online]. Available: <https://arxiv.org/abs/2510.00035>
- [4] P. Mooney, “Chest X-Ray Images (Pneumonia),” *Kaggle Datasets*, 2018. [Online]. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [5] M. S. Ahmed, M. R. Ahmed, M. M. Rahman, M. N. Sakib, and M. Al-Amin, “Pneumonia Detection in Chest X-Ray Images Based on Convolutional Neural Networks and Transfer Learning,” *Journal of Software Engineering and Applications*, vol. 18, no. 1, pp. 1-15, Jan. 2025. [Online]. Available: <https://www.scirp.org/journal/paperinformation?paperid=140592>