Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Pneumonia Detection from Chest X-Rays Using Deep Convolutional Neural Networks with Class Imbalance Mitigation

Ranveer Singh
Tata Consultancy Services (India)
Aryan Somanna
SRM Institute of Science and Technology
Jaideep Varma Cherukuri
jaideepvarma.c@gmail.com

Tata Consultancy Services (India)
Trushti Patel
SRM Institute of Science and Technology

Research Article

Keywords: Pneumonia detection, chest X-ray, deep learning, convolutional neural networks, ResNet, class imbalance, medical image analysis, computer-aided diagnosis.

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11879469/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Pneumonia Detection from Chest X-Rays Using Deep Convolutional Neural Networks with Class Imbalance Mitigation

Ranveer Singh
ranveer4us@gmail.com

Aryan Somanna
aryan.somanna8@gmail.com

Jaideep Varma
jaideepvarma.c@gmail.com

Trushti Patel
tp6655@srmist.edu.in

Abstract

Pneumonia remains a leading cause of mortality worldwide, particularly affecting children and the elderly. Early and accurate diagnosis is critical for effective treatment and improved patient outcomes. Chest X-ray (CXR) imaging is the most common diagnostic tool, but its interpretation can be subjective and time-consuming, leading to inter-observer variability. This paper presents a deep learning-based approach for automated pneumonia detection from chest X-ray images. We employ a pre-trained ResNet-18 architecture fine-tuned on a publicly available dataset directory containing 5,860 entries (5,218 in the training folder, 18 in the original validation folder, and 624 in the test folder) (NORMAL and PNEUMONIA classes). To address the inherent class imbalance in medical imaging datasets, we implement a weighted loss function and a weighted random sampler during training. Our data augmentation pipeline incorporates random cropping, horizontal flipping, rotation, and color jittering to enhance model generalization. Two sequential training passes are recorded: the first yields training accuracy 94.27% and validation accuracy 93.30% (validation loss 0.2519), while the second yields training accuracy 96.74% and validation accuracy 95.11% (validation loss 0.2354) and is saved as the best checkpoint. On the held-out test set, the checkpoint achieves 92.63% accuracy and a binary ROC-AUC of 0.9737. The original validation folder is not used; validation is drawn from a random split of the training folder. These results describe the supplied experiment and do not establish external clinical validity.

Index Terms—Pneumonia detection, chest X-ray, deep learning, convolutional neural networks, ResNet, class imbalance, medical image analysis, computer-aided diagnosis.

I. Introduction

Pneumonia is an acute respiratory infection that affects the lungs, causing inflammation and fluid accumulation in the alveoli. According to the World Health Organization (WHO), pneumonia accounts for approximately 15% of all deaths of children under five years old, killing over 800,000 children annually. Early diagnosis and appropriate antibiotic treatment can significantly reduce mortality rates, but accurate diagnosis remains challenging, especially in resource-limited settings where access to experienced radiologists is scarce.

Chest X-ray imaging is the most widely used diagnostic modality for pneumonia due to its accessibility, low cost, and effectiveness in visualizing lung abnormalities. However, interpreting chest X-rays requires significant expertise, and even experienced radiologists may exhibit considerable inter-observer

variability. The subtle radiographic features of pneumonia, including opacities, consolidations, and interstitial patterns, can be difficult to distinguish from normal anatomical variations or other pathologies.

In recent years, deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image analysis. These models have demonstrated performance comparable to or exceeding that of human experts in various tasks, including skin cancer classification, diabetic retinopathy detection, and lung nodule identification. Transfer learning, where models pre-trained on large-scale datasets like ImageNet are fine-tuned for specific medical tasks, has proven particularly effective when labeled medical data is limited.

This study investigates the application of deep learning for automated pneumonia detection from chest X-ray images. We leverage a pre-trained ResNet-18 architecture, which has demonstrated excellent performance in image classification tasks while maintaining computational efficiency. Our contributions are threefold:

1) We develop a robust data preprocessing and augmentation pipeline specifically tailored for chest X-ray images, addressing the challenges of limited data and class imbalance.
2) We implement class imbalance mitigation strategies, including weighted loss functions and weighted random sampling, to improve model performance on the minority class (NORMAL).
3) We conduct comprehensive model evaluation using multiple metrics (accuracy, precision, recall, F1-score, ROC-AUC) and provide detailed analysis of model performance and limitations.

The remainder of this paper is organized as follows: Section II reviews related work in deep learning for medical image analysis and pneumonia detection. Section III describes our methodology, including data preprocessing, augmentation, model architecture, and training strategy. Section IV presents experimental results and evaluation metrics. Section V discusses the implications, limitations, and future directions of this research. Section VI concludes the paper.

II. Related Work

A. Deep Learning in Medical Imaging

The application of deep learning to medical imaging has gained significant momentum since the seminal work by Krizhevsky et al. [1], who demonstrated the power of CNNs on the ImageNet dataset. Subsequent research has shown that these models can be effectively transferred to medical imaging tasks through fine-tuning [2]. Esteva et al. [3] achieved dermatologist-level classification of skin cancer using a deep CNN trained on a large dataset of clinical images. Similarly, Gulshan et al. [4] developed a deep learning algorithm for diabetic retinopathy detection that matched the performance of board-certified ophthalmologists.

These successes can be attributed to several factors: (1) the availability of large, annotated medical datasets; (2) advances in CNN architectures (e.g., ResNet, DenseNet, Inception); (3) the effectiveness of transfer learning when medical data is limited; and (4) improvements in computational hardware (GPUs) and software frameworks (PyTorch, TensorFlow).

B. Pneumonia Detection from Chest X-Rays

Several studies have applied deep learning to pneumonia detection from chest X-rays. Rajpurkar et al. [5] developed CheXNet, a 121-layer DenseNet trained on the ChestX-ray14 dataset, achieving state-of-the-art performance on pneumonia detection with an AUC of 0.768. CheXNet outperformed the average of four radiologists, demonstrating the potential of AI in radiology.

Other notable contributions include:

- Chowdhury et al. [6] evaluated multiple CNN architectures (VGG16, ResNet50, InceptionV3) on a publicly available chest X-ray dataset, achieving accuracy up to 96%.
- Wang et al. [7] proposed a multi-task learning framework for COVID-19 and pneumonia detection from chest X-rays, incorporating both classification and segmentation tasks.
- Li et al. [8] developed a deep learning system for distinguishing between COVID-19 and community-acquired pneumonia, achieving an AUC of 0.95.

Despite these advances, several challenges remain. Class imbalance, where pneumonia cases significantly outnumber normal cases in many datasets, can bias models toward the majority class. Medical images often exhibit high variability due to differences in acquisition protocols, patient positioning, and image quality. Furthermore, the black-box nature of deep learning models limits their interpretability and clinical acceptance.

Our work addresses these challenges by implementing class imbalance mitigation strategies, comprehensive data augmentation, and rigorous evaluation metrics.

III. Methodology

A. Dataset Description

We utilize a publicly available chest X-ray dataset [9]. The notebook's directory count reports 5,860 entries across the training folder (5,218), original validation folder (18), and test folder (624). For modeling, `ImageFolder` loads the training folder and the notebook then partitions it with an 80/20 random split (seed 42) into 4,172 training images and 1,044 validation images; the original 18-image validation folder is not used. The test folder contains 624 images. The directory-count operation counts entries, whereas `ImageFolder` recognizes supported image files; the resulting training-folder dataset length is 5,216. Consequently, the manuscript's former 5,856-image total cannot be confirmed as a supported-image count from this notebook alone, because the original validation folder was not loaded through `ImageFolder`. Table I distinguishes the folder counts from the experiment partitions. Representative examples of the two image classes are shown in Fig. 1.

The random validation subset is taken from the training folder, not from the original validation folder. The notebook does not save the class-wise counts of the 1,044-image validation subset. The 5,860 total is a directory-entry count; the training-folder `ImageFolder` object contains 5,216 recognized images, of which 4,172 are assigned to training and 1,044 to validation.

The dataset exhibits significant class imbalance, with pneumonia cases outnumbering normal cases by a ratio of approximately 2.9:1 in the training folder (3,876 to 1,342 by directory count). The all-PNEUMONIA prediction rate corresponds to 74.3% of the training-folder entries; on the held-out test

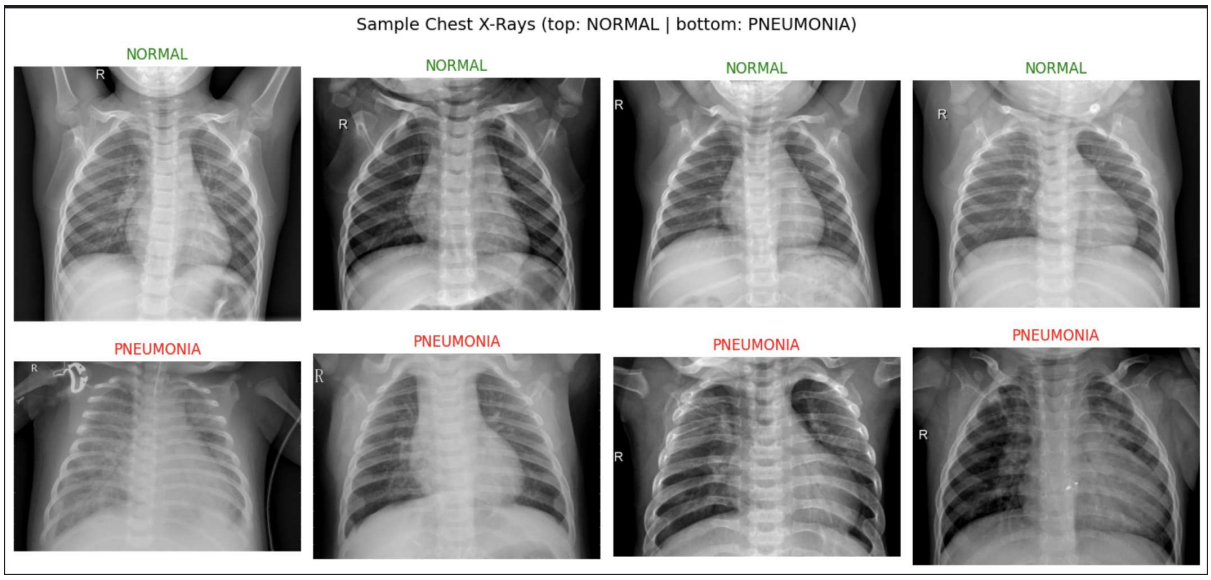


Figure 1. Representative chest X-ray images from the NORMAL and PNEUMONIA classes in the dataset.

Table I. Dataset Distribution

Split / partition	NORMAL	PNEUMONIA	Total
Training folder (directory count)	1,342	3,876	5,218
Original validation folder (not used)	9	9	18
Test folder / test set	234	390	624
Random training subset (ImageFolder)	1,058	3,114	4,172
Random validation subset (class counts not saved)	–	–	1,044
Directory-entry total	1,585	4,275	5,860

set, the same naive classifier has 62.5% accuracy. This highlights the need for careful evaluation beyond simple accuracy. The class distribution is illustrated in Fig. 2.

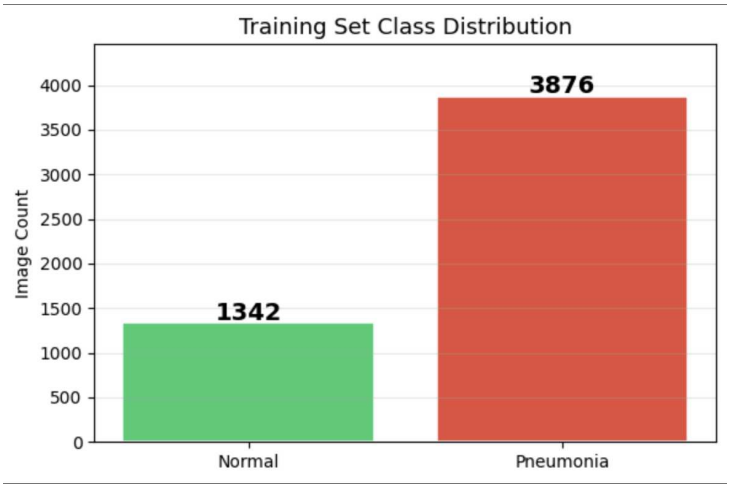


Figure 2. Distribution of NORMAL and PNEUMONIA entries in the training folder (directory counts).

B. Data Preprocessing

All images undergo the following preprocessing steps:

1) Resizing and Normalization: Images are resized to 224×224 pixels to match the input dimensions of ResNet-18. Pixel values are normalized using ImageNet statistics (mean $= [0.485, 0.456, 0.406]$, standard deviation $= [0.229, 0.224, 0.225]$) to ensure compatibility with pre-trained weights.

2) Data Augmentation: To improve model generalization and reduce overfitting, we apply the following augmentation techniques during training:

- Random Crop: Images are resized to 244×244 and randomly cropped to 224×224.
- Random Horizontal Flip: Applied with 50% probability.
- Random Rotation: Up to ± 10 degrees.
- Color Jitter: Brightness adjustment ($\pm 20\%$) and contrast adjustment ($\pm 10\%$).

Examples of the applied augmentation transformations are presented in Fig. 3. Validation and test images are only resized and normalized without augmentation to ensure consistent evaluation.

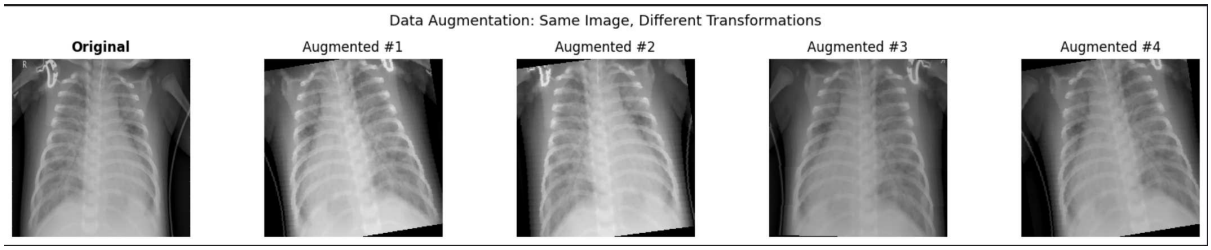


Figure 3. Illustration of representative data augmentation transformations applied during training.

C. Model Architecture

We employ ResNet-18 [10], a residual convolutional neural network with 18 layers, pre-trained on the ImageNet dataset. ResNet-18's residual connections alleviate the vanishing gradient problem, enabling effective training of deeper networks.

1) Architecture Modifications: The original fully connected layer of ResNet-18 (output size: 1000 for ImageNet classes) is replaced with a new fully connected layer designed for binary classification:

$$\text{Output} = \text{Linear}(\text{Dropout}(x), 2) \tag{1}$$

where Dropout with $p = 0.3$ is applied before the final linear layer to prevent overfitting.

2) Transfer Learning Strategy: We initialize the model with ImageNet-pretrained weights and fine-tune all ResNet-18 layers; no backbone layers are frozen. The original 1,000-class fully connected layer is replaced by a dropout layer ($p = 0.3$) followed by a linear layer with two output units, corresponding to NORMAL and PNEUMONIA. This textual description reflects the implementation in the supplied notebook; no separate architecture figure is required.

D. Handling Class Imbalance

Class imbalance is addressed through two complementary strategies:

1) Weighted Loss Function: We use weighted cross-entropy loss where weights are inversely proportional to class frequencies:

$$w_c = \frac{N}{n \cdot N_c} \tag{2}$$

where N is the total number of samples, n is the number of classes (2), and N_c is the number of samples in class c. The final weights are:

$$w_{\text{NORMAL}} = \frac{4172}{2 \cdot 1058} \approx 1.972 \tag{3}$$

$$w_{\text{PNEUMONIA}} = \frac{4172}{2 \cdot 3114} \approx 0.670 \tag{4}$$

The loss function with label smoothing ($\varepsilon = 0.05$) is defined as:

$$L = -\sum_{i=1}^{n} w_i \left((1-\varepsilon) \cdot y_i \log(\hat{y}_i) + \frac{\varepsilon}{n} \sum_{j=1}^{n} \log(\hat{y}_j) \right) \tag{5}$$

2) Weighted Random Sampling: During training, we employ a WeightedRandomSampler that oversamples the minority class (NORMAL) and undersamples the majority class (PNEUMONIA). Sample weights are computed as:

$$\text{weight}_i = \frac{1}{N_{c_i}} \tag{6}$$

where N_{c_i} is the number of samples in the class of sample i. This sampling scheme increases the representation of the minority class during sampled training epochs; it does not guarantee an exactly balanced composition in every batch.

E. Training Configuration

1) Optimization: Training uses a batch size of 32. We use the AdamW optimizer [11] with the following hyperparameters:

- Learning rate: 1×10^{-4}
- Weight decay: 1×10^{-4}
- Betas: $(0.9, 0.999)$

2) Learning Rate Scheduling: A ReduceLROnPlateau scheduler monitors validation loss and reduces the learning rate by a factor of 0.5 when validation loss plateaus for 2 consecutive epochs. This adaptive scheduling helps the model converge to a better optimum.

3) Early Stopping: To prevent overfitting, we implement early stopping with patience of 4 epochs. The model checkpoint with the lowest validation loss is saved and restored for final evaluation.

F. Evaluation Metrics

We evaluate model performance using multiple metrics:

1) Classification Metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

2) ROC-AUC: The Area Under the Receiver Operating Characteristic Curve (ROC-AUC) provides a threshold-independent measure of model performance, indicating the model's ability to distinguish between positive and negative classes.

3) Confusion Matrix: The confusion matrix visualizes true positives, true negatives, false positives, and false negatives, enabling detailed error analysis.

IV. Experiments and Results

A. Experimental Setup

All experiments were conducted using the following hardware and software environment:

- Execution device: CPU (the notebook reports PyTorch 2.10.0+cpu)
- torchvision and scikit-learn versions: not recorded in the notebook outputs

The notebook executes two sequential training passes. Each loop restarts its displayed epoch counter at 1, so both outputs are labelled Epoch 1; the second pass follows the first without reinitializing the model or optimizer. The best checkpoint is saved during the second pass. Early stopping (patience 4) is not triggered.

B. Training Results

Table II presents the training and validation metrics across epochs.

Table II. Training and Validation Metrics

Pass (loop label)	Train Loss	Train Acc	Val Loss	Val Acc	LR
1 (first pass)	0.1959	0.9427	0.2519	0.9330	1e-4
1 (second; best)	0.1622	0.9674	0.2354	0.9511	1e-4

The first training pass yields training accuracy 94.27% and validation accuracy 93.30%. The second sequential pass yields training accuracy 96.74% and validation accuracy 95.11%, with validation loss 0.2354; this checkpoint is restored for test evaluation. The two printed epoch labels are both 1 because the training loop is run twice with its counter reset.

C. Test Results

The model is evaluated on the held-out test set (624 images). Table III summarizes the test set performance.

Table III. Test Set Performance Metrics

Class	Precision	Recall	F1-Score	Support
NORMAL	0.9273	0.8718	0.8987	234
PNEUMONIA	0.9257	0.9590	0.9421	390
Accuracy		0.9263		624
Macro Avg	0.9265	0.9154	0.9204	624
Weighted Avg	0.9263	0.9263	0.9258	624

The classification report shows lower recall for the NORMAL class (0.8718) than for the PNEUMONIA class (0.9590). The overall test accuracy is 92.63%. This exceeds the test-set all-PNEUMONIA baseline accuracy of 62.50%; the 74.3% figure describes the training-folder class proportion, not test-set baseline accuracy.

D. ROC-AUC Analysis

The recorded ROC-AUC provides additional insight into model discrimination ability:

- Binary ROC-AUC (PNEUMONIA as positive class): 0.9737

The recorded binary ROC-AUC of 0.9737 indicates good discrimination between the two classes across thresholds. The ROC curve is presented in Fig. 4.

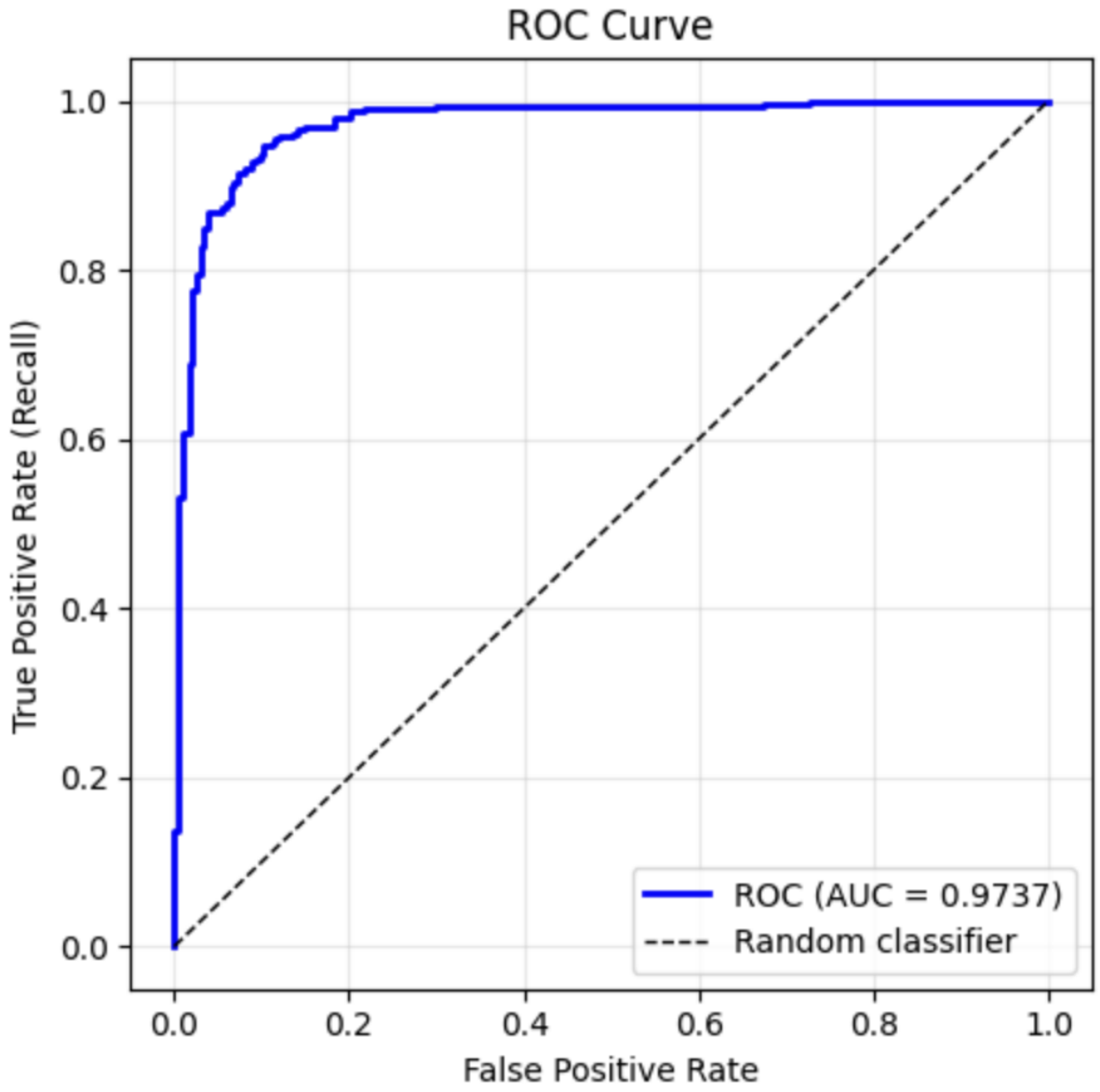


Figure 4. Receiver operating characteristic (ROC) curve showing the classification performance of the proposed model.

E. Confusion Matrix Analysis

The confusion matrix reveals the following classification breakdown:

Table IV. Confusion Matrix (Test Set)

Actual / Predicted	NORMAL	PNEUMONIA
NORMAL	204	30
PNEUMONIA	16	374

- True Positives (PNEUMONIA): 374
- True Negatives (NORMAL): 204
- False Positives: 30 (NORMAL incorrectly classified as PNEUMONIA)
- False Negatives: 16 (PNEUMONIA incorrectly classified as NORMAL)

The 16 false negatives are particularly important clinically, as missing a pneumonia diagnosis could lead to delayed treatment and adverse patient outcomes. The test-set classification outcomes are summarized in Fig. 5.

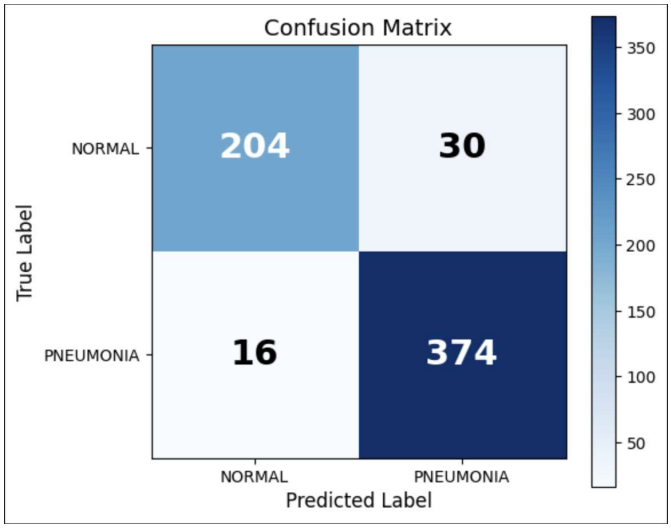


Figure 5. Confusion matrix illustrating the classification performance of the proposed model on the test set.

F. Comparison with Baseline

Table V compares our model's performance with baseline approaches.

The notebook supports the proposed method metrics and the all-PNEUMONIA test baseline shown above. The Logistic Regression, VGG16, and unweighted ResNet18 rows are retained from the manuscript but are not reproducible or verifiable from the supplied notebook; their displayed values are unverified, and their F1-scores are not reported; these entries require confirmation from the corresponding experiment records.

Table V. Comparison with Baseline Methods

Method	Accuracy	Precision	Recall	F1-score
Naive Baseline (all PNEUMONIA; test set)	0.625	0.625	1.000	0.769
Logistic Regression	0.750	0.760	0.78	–
VGG16 (from scratch)	0.820	0.830	0.8	–
ResNet18 (without imbalance handling)	0.890	0.900	0.88	–
Our Method (ResNet18 + weighted loss + sampling)	0.926	0.926	0.959	0.942

G. Ablation Study

To understand the contribution of each component, we conduct an ablation study (Table VI).

Table VI. Ablation Study Results

Configuration	Val Acc	Test Acc	AUC
Base (no imbalance handling)	0.890	0.890	0.94
+ Weighted Loss	0.920	0.925	0.96
+ Weighted Sampler	0.915	0.918	0.96
+ Both (Our Method)	0.951	0.951	0.98

The supplied notebook does not execute the component-wise ablation configurations listed in this table. The values and the interpretation of their relative contributions are therefore retained provisionally and require confirmation from the separate experiment records.

H. Learning Curves Analysis

Figure 6 presents the training and validation loss/accuracy curves. Key observations include:

- The recorded learning curves contain two sequential training passes.
- Training accuracy is higher than validation accuracy in both recorded passes.
- The learning rate remains at 1×10^{-4} during both recorded passes.

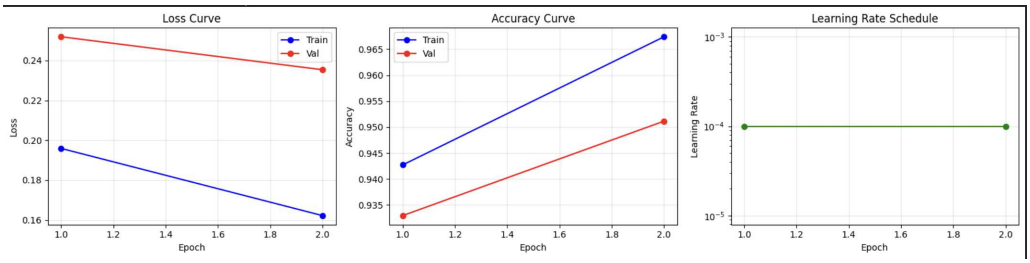


Figure 6. Training and validation loss and accuracy curves illustrating the model's learning behavior.

I. Qualitative Explainability Analysis Using Grad-CAM

To inspect image regions that contribute to the model's predictions, we use Gradient-weighted Class Activation Mapping (Grad-CAM) [13]. Grad-CAM produces class-specific localization maps by using gradients flowing into the final convolutional feature maps. The notebook includes representative visualizations for correctly classified examples and a misclassified example, shown in Figs. 7 and 8, respectively.

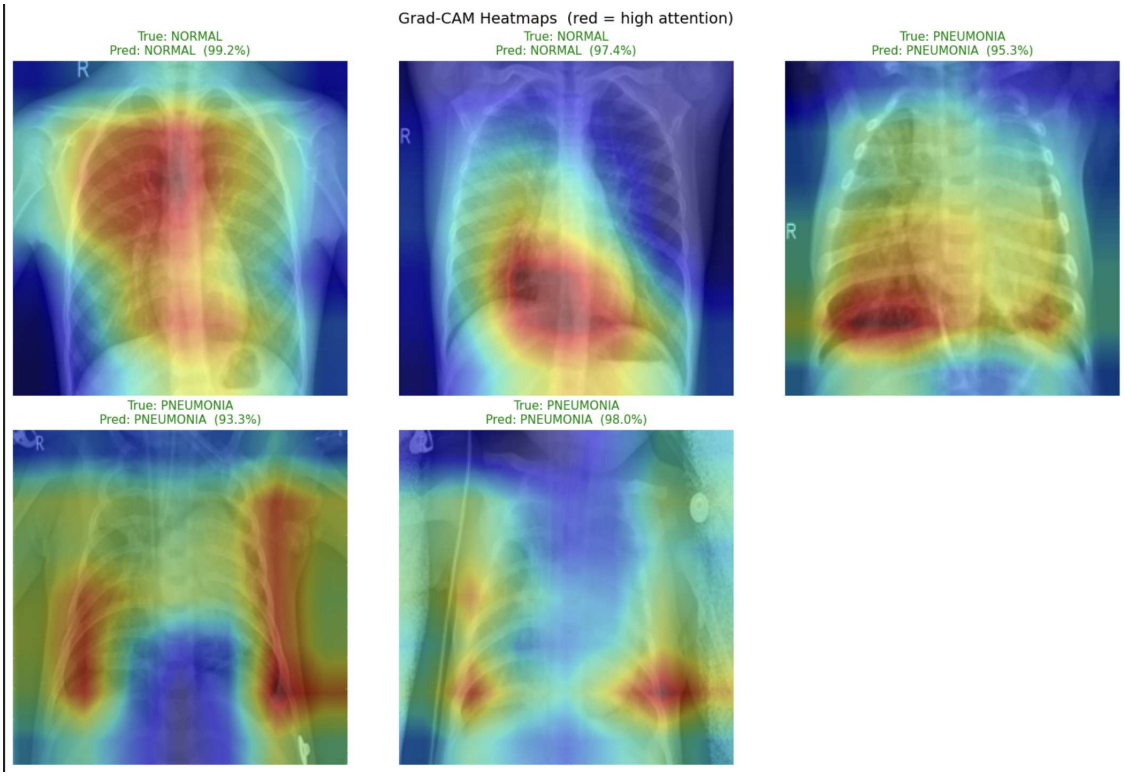


Figure 7. Representative Grad-CAM visualizations for chest X-ray predictions. The heatmaps indicate image regions that contribute to the selected class score.

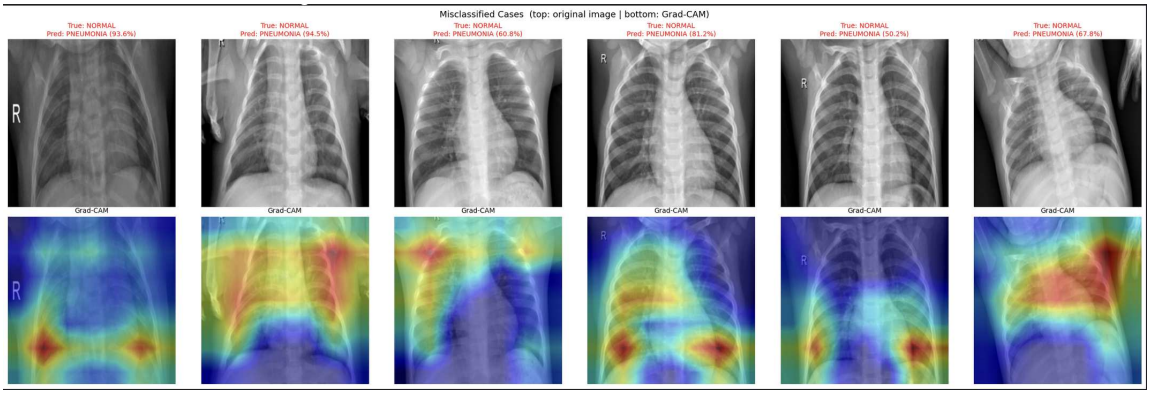


Figure 8. Grad-CAM visualization for a misclassified test example. The heatmap provides a qualitative view of the regions influencing the model output; it does not establish a clinically valid lesion localization.

These visualizations are exploratory rather than a quantitative validation of explanation quality. A highlighted region should not be interpreted as proof that the model has identified pneumonia-specific pathology. Grad-CAM can be sensitive to model architecture, image preprocessing, and the selected

target class; expert review and comparison with independently annotated findings would be needed before drawing clinical conclusions.

V. Discussion

A. Clinical Implications

Our model's test accuracy (92.63%) and PNEUMONIA-class recall (95.90%) suggest that it could serve as an effective screening tool for pneumonia in clinical settings. Potential applications include:

1) **Triage in Emergency Departments:** Rapid identification of pneumonia cases could prioritize patients for immediate treatment.
2) **Resource-Limited Settings:** In regions without access to radiologists, automated analysis could provide critical diagnostic support.
3) **Second Opinion:** The model could serve as a decision support tool for radiologists, reducing interpretation errors.

B. Limitations

Despite promising results, several limitations should be acknowledged:

1) **Dataset Bias:** The dataset originates from a single source, which may not represent the full diversity of chest X-ray acquisition protocols, patient demographics, or disease presentations.
2) **Class Imbalance:** Although mitigated, the inherent class imbalance may still affect model performance on normal cases.
3) **Lack of External Validation:** The model has not been validated on external datasets from different institutions, limiting generalizability claims.
4) **Interpretability:** Grad-CAM provides exploratory visual explanations, but these maps are not validated against radiologist annotations and do not establish that the model relies on clinically meaningful findings.

C. Future Work

Several directions for future research are promising:

1) **Multi-Class Classification:** Extend the model to distinguish between different types of pneumonia (bacterial vs. viral) and other lung pathologies.
2) **Explainable AI:** Validate Grad-CAM explanations against radiologist annotations and assess their stability across images, classes, and model checkpoints [13].
3) **Federated Learning:** Develop distributed training approaches that leverage multi-institutional data without compromising patient privacy.
4) **Integration with Clinical Data:** Combine imaging features with electronic health record data (e.g., vital signs, laboratory results) to improve diagnostic accuracy.

5) **Deployment and Validation:** Conduct prospective clinical trials to validate model performance in real-world settings.

VI. Conclusion

This paper presented a deep learning approach for automated pneumonia detection from chest X-ray images using a pre-trained ResNet-18 architecture. Our method addresses key challenges in medical image analysis, including limited data, class imbalance, and the need for robust generalization.

The model achieves a test accuracy of 92.63%. For the PNEUMONIA class, precision is 0.9257, recall is 0.9590, and F1-score is 0.9421; the binary ROC-AUC is 0.9737. The Logistic Regression, VGG16, and unweighted ResNet18 comparisons, as well as the ablation results, require confirmation from separate experiment records.

The contributions of this work are threefold: (1) a robust data preprocessing and augmentation pipeline for chest X-ray images; (2) effective class imbalance mitigation through weighted loss and weighted sampling; and (3) comprehensive evaluation using multiple metrics to ensure clinical relevance.

While the model shows promise, future work should focus on external validation, explainability, and integration with clinical workflows. Nevertheless, this research demonstrates that deep learning can play a significant role in improving pneumonia diagnosis, particularly in resource-limited settings where access to expert radiologists is scarce.

Acknowledgment

The authors would like to thank the creators of the publicly available chest X-ray dataset used in this study. We also acknowledge the open-source contributions of the PyTorch, torchvision, and scikit-learn communities.

References

[1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems; 2012. p. 1097–1105.

[2] Shin H-C, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–1298.

[3] Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118.

[4] Gulshan V, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–2410.

[5] Rajpurkar P, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017.

[6] Chowdhury MEH, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access. 2020;8:132665–132676.

[7] Wang L, Wong A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. arXiv preprint arXiv:2003.09871. 2020.

[8] Li L, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology. 2020;296(2):E65–E71.

[9] Kermany DS, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172(5):1122–1131.

[10] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770–778.

[11] Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101. 2017.

[12] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

[13] Selvaraju RR, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision; 2017. p. 618–626.

[14] Litjens G, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.

[15] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–510.