A Deep Learning-Based Pneumonia Detection and Classification System Using Chest X-Ray Images

[1]M.K.V. Anvesh, [2]Varanasi Sravani, [2]Theeda Poojitha, [2]Velagala Jeevana Jyothi, [2]Yalala Lakshmi Richita

[1]Ph.D. Scholar, [2]BTech Student

[1]Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam,

[2]Department of Computer Science and Systems Engineering, Andhra University College of Engineering for Women, Visakhapatnam

[2]sravani6613.nms@gmail.com

***Abstract*— Pneumonia requires timely diagnosis to reduce mortality, but manual chest X-ray interpretation is subjective to human bias. This study leverages convolutional neural networks—EfficientNetV2B0 for pneumonia detection and VGG16 for bacterial-viral classification—trained on publicly available datasets using data augmentation, transfer learning, and class balancing techniques. The models achieve 99.45% and 97.45% accuracy, respectively, with strong performance across evaluation metrics. Grad-CAM visualization enhances interpretability by highlighting key anatomical regions influencing predictions. These findings reinforce the potential of AI in assisting radiologists with pneumonia diagnosis while improving accessibility for patients in remote areas, enabling clearer understanding of medical insights beyond clinical settings.**

***Keywords:* Pneumonia detection, Pneumonia classification, Deep learning, EfficientNetV2B0, VGG16, Grad-CAM, Chest X-ray.**

I. Introduction

Pneumonia, affecting about 7% of the global population, is a major cause of death in developing countries, particularly among children and the elderly. Mainly caused by bacteria or viruses, it inflames the alveoli by filling them with fluid. Diagnosis and treatment are challenging due to reliance on physical tests and human expertise, making delays or errors potentially life-threatening and costly.

Advances in AI, particularly computer vision and deep learning, have enabled computers to analyze medical images accurately, providing pretrained models as reliable tools for preliminary tests or second opinions. This study utilizes Convolutional Neural Networks (CNN) for pneumonia detection and classification, focusing on three objectives:

1. Detecting pneumonia from chest X-rays using EfficientNetV2B0.
2. Classifying cases into bacterial or viral using VGG16.
3. Visualizing infected regions with generalized Grad-CAM.

II. Related Works

(a) EfficientNetV2, a scalable image classification framework, has advanced medical imaging. Ghosh et al. [1] showed EfficientNetV2L's superior accuracy for pneumonia detection using progressive learning and optimized scaling. Inspired by this, we used EfficientNetV2B0 in our project for balanced performance and computational efficiency.

(b) Chouhan et al. [2] highlighted transfer learning's benefits in medical imaging tasks with limited data. Fine-tuning models like VGG16 on chest X-rays improved generalization, which we utilized for bacterial and viral pneumonia classification using ImageNet pretraining.

(c) Selvaraju et al. [3] proposed Grad-CAM, generating class-discriminative heatmaps for model prediction interpretation. We refined Grad-CAM pipeline for chest X-rays by combining TensorFlow-based heatmap generation with PyTorch-driven lung segmentation via DeepLabV3-ResNet50, ensuring spatially precise and anatomically relevant visualizations.

III. Proposed Model

Our system combines EfficientNetV2B0 for pneumonia detection and VGG16 for subtype classification, incorporating Grad-CAM to generate heatmaps that visually guide radiologists. Enhancements include contrast enhancement (CLAHE) and morphological filtering to refine visual focus, addressing gaps in prior models. This approach improves transparency and reliability for clinicians.

3.1 Dataset Description

Two publicly available chest X-ray datasets were used:

(a) Stage 1 Dataset: Includes 5,232 X-ray images (3,884 pneumonia cases, 1,349 normal), sourced from Guangzhou Women and Children's Medical Center and hosted on Mendeley Data [4]. Test data comprises 624 images.

(b) Stage 2 Dataset: Combines 15 public datasets, tailored for bacterial and viral pneumonia classification. It contains 1,656 viral and 3,001 bacterial pneumonia X-rays, sourced from Mendeley Data [5].

3.2 Preprocessing

Distinct pipelines were implemented for the modules:

(a) Module 1 (EfficientNetV2B0): Images resized to 224×224 pixels, normalized, and augmented. Validation and testing sets were split in an 80:10:10 ratio using TensorFlow pipelines for efficiency.

(b) Module 2 (VGG16): Images resized to 96×96 pixels, normalized via VGG-specific preprocessing, and augmented with geometric transformations. TensorFlow's API enabled efficient batching and caching.

3.3 Comparative Analysis of Architectures

Several CNN architectures were evaluated, including EfficientNetV2B0, VGG16, ResNet, InceptionNet, and DenseNet variants. After testing on the same curated datasets with identical preprocessing, EfficientNetV2B0 was selected for pneumonia detection [1], while VGG16 excelled in bacterial versus viral pneumonia classification [2]. Performance metrics, including precision, recall, F1-score, and ROC-AUC, support these choices (details in Tables 1 and 2).

Table 1: Comparative analysis of various CNN architectures for pneumonia vs. normal categorization (Module 1)

MODEL TYPE	TRAINING ACCURACY	VALIDATION ACCURACY	TESTING ACCURACY	PRECISION	RECALL	F1-SCORE	ROC-AUC SCORE
EfficientNetV2B0	98.46%	98.66%	99.45%	99%	99%	99%	99.99%
VGG16	96.90%	98.47%	98.85%	99%	99%	99%	99.93%
ResNet50V2	95.76%	98.95%	98.15%	98%	98%	98%	99.83%
InceptionNetV3	81.62%	89.69%	90.27%	92%	90%	91%	96.93%
MobileNetV2	95.74%	96.18%	77.61%	82%	78%	75%	93.68%
DenseNet121	94.15%	94.27%	78.88%	81%	78%	76%	91.74%
DenseNet201	98.94%	99.24%	75.16%	81%	75%	71%	89.61%

Table 2: Evaluation of model architectures for bacterial vs. viral pneumonia classification (Module 2)

MODEL TYPE	TRAINING ACCURACY	VALIDATION ACCURACY	TESTING ACCURACY	PRECISION	RECALL	F1-SCORE	ROC-AUC SCORE
VGG16	98.47%	98.50%	97.45%	98%	97%	97%	99.97%
EfficientNetV2B0	92.32%	91.24%	90.36%	90%	90%	90%	96.55%
ResNet50V2	95.76%	89.10%	98.15%	98%	98%	98%	92.75%
InceptionNetV3	77.48%	80.56%	83.19%	85%	83%	83%	91.8%
DenseNet201	77.96%	81.08%	81.26%	81%	81%	81%	-
DenseNet121	81.31%	78.81%	60.25%	65%	60%	56%	-

3.4 Model Architecture

To ensure robust performance and enhance the accuracy across both stages of our dual-model framework, a similar training and optimization technique was utilized. The entire procedure, in sequence, is as follows:

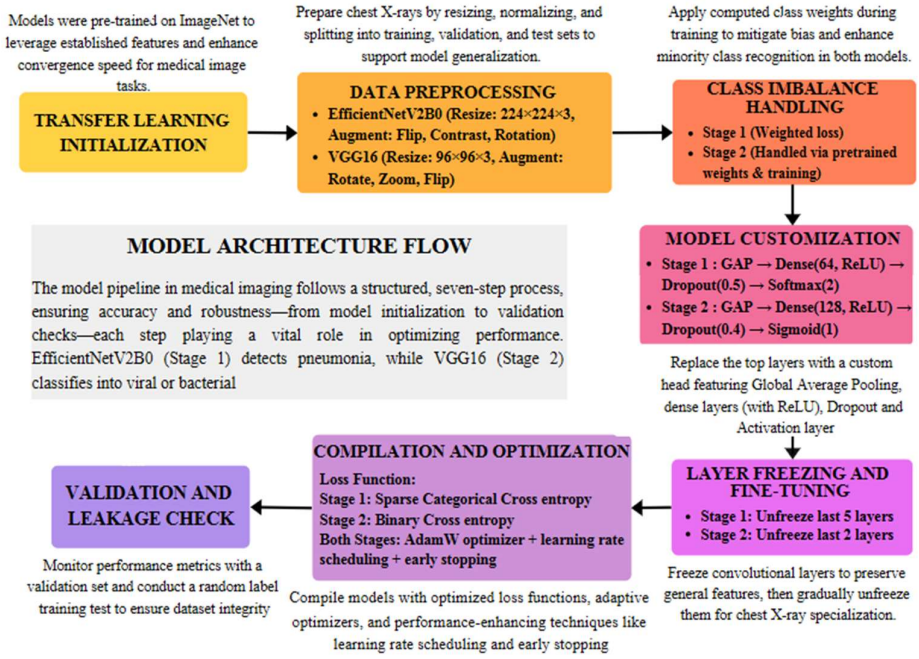

Figure 1: Model architecture flow-diagram

(a) Stage 1: Pneumonia Detection using EfficientNetV2B0:

EfficientNetV2B0 was chosen for pneumonia detection due to its superior performance. Images were resized to 224×224×3, normalized, and augmented with techniques like flipping, brightness/contrast adjustments, and rotation to simulate radiograph variability. Class imbalance was addressed using weighted loss calculations. Initially, the model's base layers were frozen, with the top five layers unfrozen for fine-tuning. A lightweight classification head—comprising GAP, Dense (64, ReLU), Dropout (0.5), and Dense (2, softmax)—was appended. Training employed the AdamW optimizer [7] (learning rate: 1e-4), with early stopping triggered by validation loss stagnation [8].

(b) Stage 2: Bacterial vs. Viral Pneumonia Classification using VGG16

In the second stage, VGG16 was used for bacterial and viral pneumonia classification due to its superior performance. Images were resized to 96×96×3 to reduce training time while retaining diagnostic features. Preprocessing involved normalization and conservative augmentation techniques like flipping, rotation, zoom, and shear to prevent overfitting. The fully connected layers were replaced with a custom classification head. Initially, all layers were frozen, with the last two unfrozen for fine-tuning. Training utilized the Adam optimizer with callbacks and class weight handling to address dataset imbalance.

The architectures of both models are represented in fig 2 and 3 respectively.

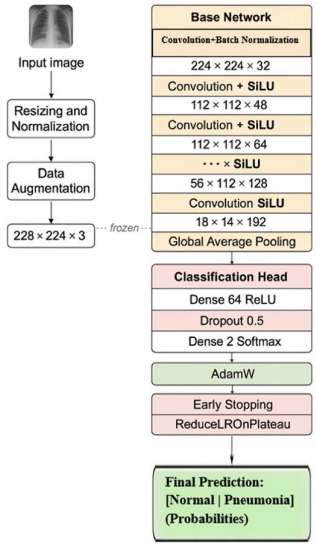

Figure 2: Architecture of Module 1 – Normal vs. Pneumonia Classification Using EfficientNetV2B0

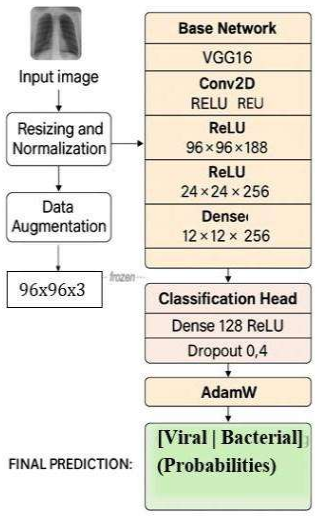

Figure 3: Architecture of Module 2 – Bacterial vs. Viral Pneumonia Classification Using VGG16

(c) Custom Classification Head

The custom classification head refines extracted features into class probabilities, optimizing task-specific predictions. Global Average Pooling (GAP) reduces spatial dimensions while preserving essential features, minimizing overfitting. Dense Layer acts as the decision-making unit, using 64 neurons for pneumonia detection and 128 neurons for bacterial-viral classification. Dropout Layer prevents overfitting by randomly deactivating neurons (50% in Module 1, 40% in Module 2). Activation Layer determines final outputs—softmax for pneumonia detection to produce probabilities for normal vs. pneumonia, and sigmoid for bacterial-viral classification, generating independent probabilities for each class.

Training details for both models are in Table 3, and their customized classification heads are in Table 4.

Table 3: Pneumonia Detection (EfficientNetV2B0) and Bacterial vs. Viral Classification (VGG16)

PARAMETER	DETAILS	
Model Used	EfficientNetV2B0	VGG16
Input Size	224 × 224 × 3	96 × 96 × 3
Preprocessing	Basic-Pixel normalization [0,1] (Standard Float Conversion)	VGG16-specific normalization
Augmentation Techniques	Brightness, contrast, mirror rotation	Rotation, shift, zoom, flipping
Class Imbalance Handling	Class weights computed	Using Pre-Trained Weights and Training Techniques
Layers (in total)	~482 layers in total	16 layers in total (13 convolutional +3 fully connected layers)
Fine-Tuning Strategy	Last 5 layers unfrozen during later training	Last 2 layers unfrozen during later training
Classification Head	GlobalAveragePooling2D → Dense(64, ReLU) → Dropout(0.5) → Dense(2, Softmax)	GlobalAveragePooling2D → Dense(128, ReLU) → Dropout(0.4) → Dense(1, Sigmoid)
Optimizer	AdamW	AdamW
Learning Rate	1e-4	3e-4
Loss Function	Sparse Categorical Cross entropy	Binary Cross entropy

Table 4: Classification Head Details

FEATURE	EFFICIENTNETV2B0	VGG16
Dense Layer Size	64 neurons	128 neurons
Dropout Rate	50%	40%
Activation Function	Softmax (2 output classes)	Sigmoid (1 output class)
Output	Class probabilities (healthy vs pneumonia)	Single probability for pneumonia

(d) Stage 3: Grad-CAM–Driven Explainability Framework

Grad-CAM [3] enhances transparency by localizing activation regions in chest X-rays, generating interpretable heatmaps refined through anatomical priors and morphological filtering. EfficientNetV2B0 and VGG16 were integrated post-training, using gradients from the final convolutional layer to produce class-discriminative heatmaps overlaid on X-rays, with Gaussian filtering applied to remove artifacts. DeepLabV3-ResNet50 [9] performed lung segmentation, ensuring Grad-CAM highlights focused on pulmonary structures, while PyTorch-generated masks were refined using morphological filtering for precision. Heatmaps were optimized through Gaussian Centering to enhance key activations and Exponential Smoothing for improved transition clarity. Final heatmaps were blended onto X-rays via OpenCV, balancing raw image visibility with infection zone highlights for clearer diagnostics.

IV. Model Evaluation

The two-model system was evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC [6]. They were computed separately both modules and they followed a consistent evaluation strategy. Data was split into **80:10:10** (train-validation-test), with class imbalance addressed using a weighted loss function. Performance tracking involved confusion matrices, ROC curves, classification reports, and validation loss trends, with early stopping applied [8]. The best-performing model weights, selected based on validation loss, were tested on the held-out test set to ensure unbiased reporting.

Grad-CAM visualizations were generated to qualitatively highlight key regions influencing predictions in chest X-rays for both models. Integrated into a Flask-based web interface, users can toggle between original X-rays and Grad-CAM views for better interpretation. Key features include a colorbar legend indicating activation intensity (blue to red), toggle switches for raw and interpreted outputs, and PDF report downloads with embedded heatmaps. This intuitive interface ensures accessibility for both technical and non-technical users, fostering trust and transparency in model decisions.

V. Results

The comparative results of all tested architectures are summarized in Tables 1 and 2. These highlight the superiority of EfficientNetV2B0 in binary pneumonia detection and the competence of VGG16 in handling the subtler subclassification task. Module 1 (EfficientNetV2B0) achieved 99.45% test accuracy with a ROC-AUC of 0.9999 [6], demonstrating high sensitivity, specificity, and minimal false positives or negatives (F1-score: 99%). Module 2 (VGG16) attained 97.45% test accuracy and ROC-AUC of 99.97%, effectively distinguishing bacterial and viral pneumonia despite their similar radiographic features, with balanced precision and recall (97%–98%), ensuring reliable classification.
The confusion matrix and ROC-AUC curve results of both models, shown in fig 4 and 5 further validate their performance.

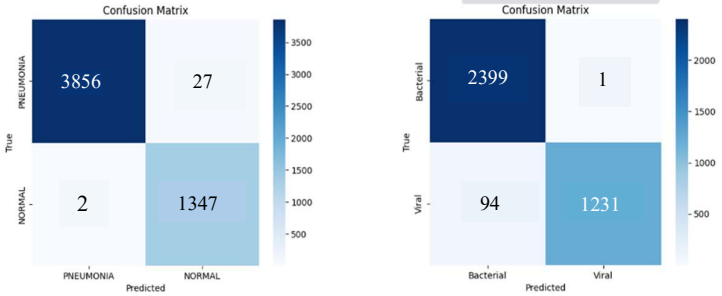

Figure 4: Confusion Matrix results for EfficientNetV2B0 and VGG16 Models respectively

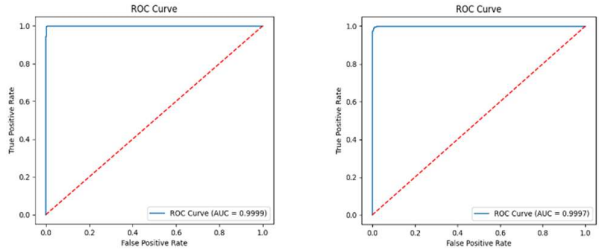

Figure 5: ROC-AUC Performance results for EfficientNetV2B0 and VGG16 Models respectively

The accuracy and loss curves for both architectures are illustrated in fig 6 and 7. Stage 1 (EfficientNetV2B0) achieved 98.46% training and 99.45% testing accuracy, with minimal overfitting and early stopping at epoch 15. Stage 2 (VGG16) demonstrated 98.47% training and 97.45% testing accuracy, with smooth loss convergence and low overfitting risk, validated by final losses of 0.0435 (training) and 0.0656 (testing).

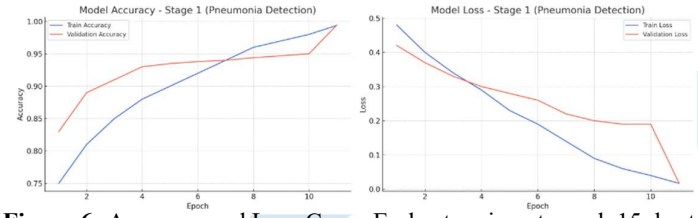

Figure 6: Accuracy and Loss Curve Early stopping at epoch 15; best weights restored from epoch 11.

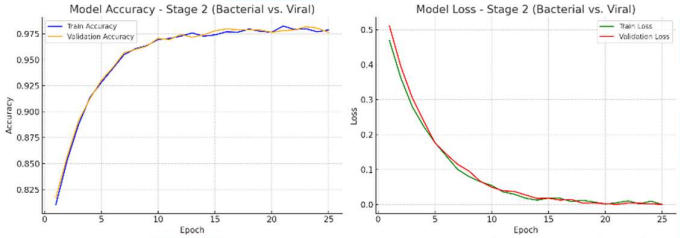

Figure 7: Accuracy and loss curve of Stage over **25** epochs

Grad-CAM visualizations were generated only for pneumonia cases in Stage 1 to prevent clutter in normal cases. While highlighting key regions, they mostly focused on general lung areas rather than infection sites. Heatmaps appeared similar within each class, indicating limited spatial resolution or insufficient localization. This could be due to reliance on final convolutional features, coarse pattern learning, or lack of pathological variation in training data. Though the model targets relevant lung regions, better interpretability techniques or higher-quality pixel-level annotations could enhance localization accuracy.
The Grad-CAM results, shown in fig 8 and 9, highlight the key activation regions that influenced the model's predictions.

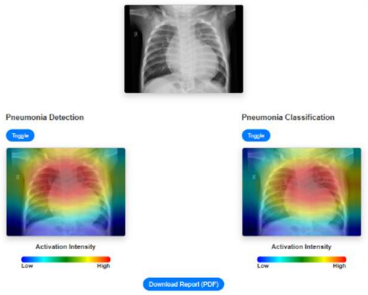

Figure 8: Grad-CAM highlighting the viral pneumonia-affected zone in the lungs.

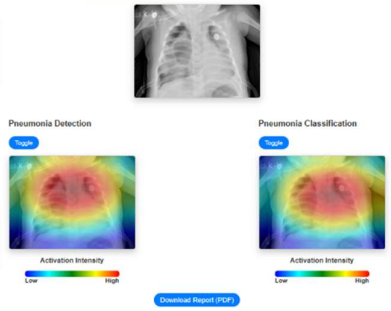

Figure 9: Grad-CAM visualization showing infection region for a bacterial pneumonia case.

VI. CONCLUSION

This study presents an AI-based pneumonia detection and classification system using deep convolutional neural networks, achieving high diagnostic accuracy (99.45% for pneumonia detection and 97.45% for bacterial-viral classification) while ensuring interpretability through Grad-CAM visualizations. Leveraging EfficientNetV2B0 and VGG16 with transfer learning, data augmentation, and class balancing, the system enhances accessibility for non-specialist users and supports clinical decision-making, particularly in resource-limited settings. Grad-CAM visualizations provide insight into the model's predictions, highlighting relevant anatomical regions, though their spatial resolution and localization capacity require further refinement. Future work may focus on improving generalizability with diverse datasets, integrating real-time hospital systems, and exploring advanced interpretability techniques to enhance the reliability of visualization-based explanations.

VII. REFERENCES

[1] Ghosh, A. Verma, M. Nath, and D. Das, "Pneumonia Detection Using Chest Radiographs with Novel EfficientNetV2L Model," in Proc. IEEE Int. Conf. Comput., Commun. and Intell. Syst. (ICCCIS), pp. 154–160, 2023.
[2] V. Chouhan, S. Kaul, U. Singh,A. Khamparia, and D. Gupta,"A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images,"Applied Sciences, vol. 10, no. 2, p. 559, 2020.
[3] R. R. Selvaraju, M. Cogswell,A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in Proc. IEEE Int. Conf. Comput. Vision (ICCV), pp. 618–626, 2017.
[4] K. Paul, J. L. Sornapudi, A. S. Abhilash, and K. Rajalakshmi, "Chest X-ray Images (Pneumonia)," Mendeley Data, V3, 2018. [Online]. Available: https://data.mendeley.com/datasets/rscbjbr9sj/3
[5] N. Sharma, A. Aggarwal, and N. Goyal, "A Deep-Learning Based Multimodal System for COVID-19 Diagnosis Using Breathing Sounds and Chest X-ray Images," Applied Soft Computing, vol. 113, p. 107522, 2021.
[6] D. M.W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," J. Mach. Learn. Technol., vol. 2, no. 1, pp. 37–63, 2011.
[7] Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," arXiv preprint arXiv:1711.05101, 2017.
[8] F. Chollet et al., "Keras: The Python Deep Learning Library," [Online].Available: https://keras.io/api/callbacks/
[9] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in Proc. ECCV, 2018, pp. 801–818.