Pneumonia Detection and Severity Estimation using Deep Learning and Machine Learning

A PROJECT REPORT

Submitted by

Gowripriya R (DL.AI.U4AID24113)
Yaalini R (DL.AI.U4AID24043)
Vepuri Satya Krishna (DL.AI.U4AID24140)

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY (B. Tech)
IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE (AIDS)

Under the guidance of Dr. Sakshi Ahuja and Dr. Nayana Rao

Submitted to

AMRITA VISHWA VIDYAPEETHAM
AMRITA SCHOOL OF ARTIFICIAL INTELLIGENCE
FARIDABAD – 121002

December 2025

BONAFIDE CERTIFICATE

This is to certify that this project report entitled **"Pneumonia Detection and Severity Estimation using Deep Learning and Machine Learning"** is the Bonafide work of **Gowripriya R, Yaalini R, and Vepuri Satya Krishna**, who carried out the project work under my supervision.

Signature

Dr. Sakshi Ahuja
Dr. Nayana Rao
Assitant Professor
School of AI, Faridabad

DECLARATION BY THE CANDIDATE

I declare that the report entitled **"Pneumonia Detection and Severity Estimation using Deep Learning and Machine Learning"** submitted by me for the degree of Bachelor of Technology is the record of the project work carried out by me under the guidance of **Dr. Sakshi Ahuja and Dr. Nayana Rao** and this work has not formed the basis for the award of any degree, diploma, associateship, fellowship, title in this or any other University or other similar institution of higher learning.

Gowripriya R (DL.AI.U4AID24113)
Yaalini R (DL.AI.U4AID24043)
Vepuri Satya Krishna (DL.AI.U4AID24140)

ACKNOWLEDGMNET

This project work would not have been possible without the contribution of many people. It gives me immense pleasure to express my profound gratitude to our honorable Chancellor Sri Mata Amritanandamayi Devi, for her blessings and for being a source of inspiration. I am indebted to extend my gratitude to our Principal, Dr. Lakshmi Mohandas Amrita School of Computing and Engineering, for facilitating us all the facilities and extended support to gain valuable education and learning experience.

I wish to express my sincere gratitude to my supervisors Dr. Sakshi Ahuja and Dr. Nayana Rao for their personal involvement and constant encouragement during the entire course of this work.

I am grateful to Project Supervisor, Review Panel Members, and the entire faculty of the Department of Computer Science and Engineering, for their constructive criticisms and valuable suggestions which have been a rich source to improve the quality of this work.

Gowripriya R (DL.AI.U4AID24113)
Yaalini R (DL.AI.U4AID24043)
Vepuri Satya Krishna (DL.AI.U4AID24140)

Contents

1 Introduction

Pneumonia is a serious respiratory infection that causes inflammation of the lung alveoli, leading to fluid or pus accumulation and impaired gas exchange. Chest X-ray imaging remains the most widely used diagnostic tool due to its cost-effectiveness and availability. However, manual interpretation of chest X-rays is time-consuming, subjective, and prone to inter-observer variability, particularly in resource-constrained healthcare settings.

Recent advances in deep learning have enabled automated analysis of chest X-rays using convolutional neural networks (CNNs). While end-to-end CNN models achieve high diagnostic accuracy, they often lack interpretability and require large annotated datasets. In clinical environments, model transparency and robustness are equally important as accuracy.

To address these challenges, this project proposes a hybrid artificial intelligence framework that combines deep learning, classical machine learning, lung segmentation, severity estimation, and explainable AI techniques. The system performs automated pneumonia detection using a fine-tuned DenseNet121 model enhanced with contrast-limited adaptive histogram equalization (CLAHE) preprocessing. Deep features extracted from the network are further used to train machine learning classifiers such as Support Vector Machines and Random Forests.

Beyond binary classification, the framework incorporates lung segmentation using a pretrained PSPNet model to isolate lung regions and estimate the percentage of lung involvement. This quantitative severity estimation enables risk stratification of patients into low, medium, and high-risk categories.

2 Biological Background

Understanding the biological and radiological characteristics of pneumonia is fundamental to the development of automated detection systems. This section outlines the relevant anatomy, the pathophysiology of the disease, and its manifestation in radiographic imaging, which serves as the basis for the computational features extracted in this project.

2.1 Anatomy of the Human Lung

The lungs are the primary organs of respiration, located within the thoracic cavity and protected by the rib cage. They are responsible for gas exchange, a process occurring in microscopic air sacs known as *alveoli*.

- **Structure:** The right lung is divided into three lobes (superior, middle, and inferior), while the left lung is divided into two lobes (superior and inferior) to accommodate the heart.
- **Alveolar Function:** In a healthy individual, alveoli are filled with air. Since air has a very low density, it absorbs very few X-ray photons. Consequently, healthy lung tissue appears dark (radiolucent) on a chest X-ray.

2.2 Pathophysiology of Pneumonia

Pneumonia is an inflammatory condition affecting the alveoli, typically caused by infectious agents such as bacteria (*Streptococcus pneumoniae*), viruses (Influenza, COVID-19), or fungi.

- **Inflammation and Exudate:** When pathogens invade the lungs, the immune system responds with inflammation. This leads to the accumulation of fluid, pus, and cellular debris—collectively known as exudate within the alveolar spaces.
- **Consolidation:** As the air in the alveoli is replaced by fluid, the lung tissue solidifies in a process called consolidation. This impairs gas exchange and results in respiratory symptoms such as cough, fever, and difficulty breathing.

2.3 Radiological Manifestations

The diagnosis of pneumonia relies heavily on Chest X-rays (CXR) because the physical changes in the lung tissue result in distinct visual patterns.

- **Opacity (Whitening):** Since fluid and pus are denser than air, consolidated areas absorb more X-rays and appear white or grey (radio-opaque) against the dark background of healthy lung tissue.
- **Patterns of Infection:**
  - *Lobar Pneumonia:* Characterized by consolidation of a large, continuous section of a lung lobe.
  - *Bronchopneumonia:* Presents as patchy areas of opacity scattered around the bronchi.
  - *Interstitial Pneumonia:* Often viral, showing diffuse, net-like ("reticular") patterns rather than solid whiteness.

2.4 Relevance to Severity Estimation

The clinical severity of pneumonia is directly correlated with the extent of lung involvement. A higher percentage of opacified lung area indicates more severe disease. This biological principle underpins Phase IV of this project, where lung segmentation is used to calculate the ratio of infected tissue to total lung area, providing a quantitative metric for severity.

These radiological and biological characteristics form the foundation for automated feature extraction and severity estimation in this project. Areas of increased opacity identified within segmented lung regions directly correspond to inflamed or consolidated tissue, enabling pixel-based quantification of disease severity.

3 Objective

The primary objective is to implement and evaluate a complete hybrid AI framework for automated pneumonia detection, classification using ML models, and severity estimation from chest X-ray images, as realized in the final integrated Python code. The specific objectives are as follows:

- Automate dataset setup, download from Kaggle, and apply CLAHE contrast enhancement with data augmentation for robust preprocessing of chest X-ray images.
- Extract deep features using DenseNet121 backbone (pretrained on ImageNet) and train/evaluate five ML classifiers (SVM, Random Forest, Logistic Regression, KNN, Naive Bayes) with cross-validation and hyperparameter considerations
- Fine-tune DenseNet121 in two stages (frozen base then upper layers) for end-to-end binary pneumonia classification, achieving high test accuracy with confusion matrices and ROC analysis.
- Export predicted pneumonia cases and perform lung segmentation using pretrained PSPNet from TorchXRayVision to isolate lung regions accurately.
- Quantify pneumonia severity via opacity detection within segmented lungs, compute lung involvement percentage.
- Generate comprehensive evaluations including classification reports, precision-recall curves, and visualizations to ensure model interpretability and real-world applicability.

4 Literature Review

The development of automated systems for pneumonia detection from chest X-ray images has been a focal point of recent medical imaging research. This section reviews six significant studies that explore various deep learning architectures, ranging from Convolutional Neural Networks (CNNs) and Transformer-based models to hybrid and explainable AI frameworks.

Winarto et al. (2025) conducted a comparative analysis of CNN-based architectures (ResNet50, DenseNet121) and Transformer-based models (ViT-B/16, Swin-T) on the Kaggle chest X-ray dataset. Their study emphasized standardized preprocessing and patient-level splitting to prevent data leakage. The results demonstrated that while Transformer models like Swin-T showed significant improvement after hyperparameter tuning, the CNN-based DenseNet121 achieved the highest performance with an accuracy of 97.49% and an AUC-ROC of 99.53%, highlighting the robustness of dense connectivity in capturing radiographic features [1].

Addressing the need for model interpretability, Behera et al. (2025) proposed a framework utilizing the EfficientNetB7 architecture integrated with Squeeze-and-Excitation (SE) blocks. This attention mechanism adaptively recalibrates channel-wise feature responses to focus on relevant lung regions. Their model achieved a detection accuracy of 98.31%. Furthermore, they incorporated Explainable AI (XAI) using Grad-CAM heatmaps to visualize critical regions influencing the decision, thereby enhancing clinical trust in the automated diagnosis [2].

Bhamare et al. (2025) explored a hybrid architectural approach by combining Deep CNNs with Bidirectional Long Short-Term Memory (Bi-LSTM) networks. In their proposed model, VGG16 served as the feature extractor, while the Bi-LSTM layers processed the features sequentially to capture spatial and temporal-like dependencies. This hybrid DCNN-BiLSTM model achieved an accuracy of 93.05%, outperforming standalone architectures like ResNet50 and Inception-V3 in their comparative analysis, suggesting that combining recurrence with convolution can effectively reduce false negatives [3].

In the domain of lightweight and optimized models, Khojare and Bhattacharjee (2025) presented a transfer learning framework utilizing MobileNetV3, ResNet50, and VGG16. They employed the Flower Pollination Algorithm (FPA) for optimizing feature extraction, addressing challenges related to data asymmetry and subtle disease patterns. Their proposed transfer learning approach achieved a training accuracy of 95.94% and a test accuracy of 88.87%, demonstrating the efficacy of integrating nature-inspired optimization

with deep learning for resource-constrained applications [4].

Kumar and Vani (2024) implemented a robust deep learning mechanism focusing on transfer learning techniques. They evaluated models such as VGG16 and Inception V3, fine-tuning them on chest X-ray datasets. Their experimental results highlighted that the proposed transfer learning model significantly outperformed standard implementations of DenseNet121 and ResNet18, achieving a remarkable accuracy of 98.53%. This study reinforces the value of fine-tuning pre-trained weights from ImageNet to overcome the limitations of limited medical datasets [5].

Finally, Menaka et al. (2025) proposed the use of the T-ResNet framework for the detection and recognition of pneumonia. Their work focused on training the model on a benchmark dataset constructed from publicly available sources, classifying images into normal, bacterial, and viral pneumonia categories. The study reported exceptional performance metrics, with the T-ResNet model achieving up to 100% accuracy in specific test scenarios, significantly surpassing traditional CNN-ELM approaches. They further emphasized the development of a software application to facilitate real-time detection in healthcare settings [6].

5 Methodology

The proposed methodology follows a structured four-phase pipeline designed to achieve accurate pneumonia detection, model comparison, and lung involvement estimation from chest X-ray images. An overview of the methodology is illustrated through sequential stages comprising preprocessing, feature extraction, classification, deep learning comparison, and lung localization.

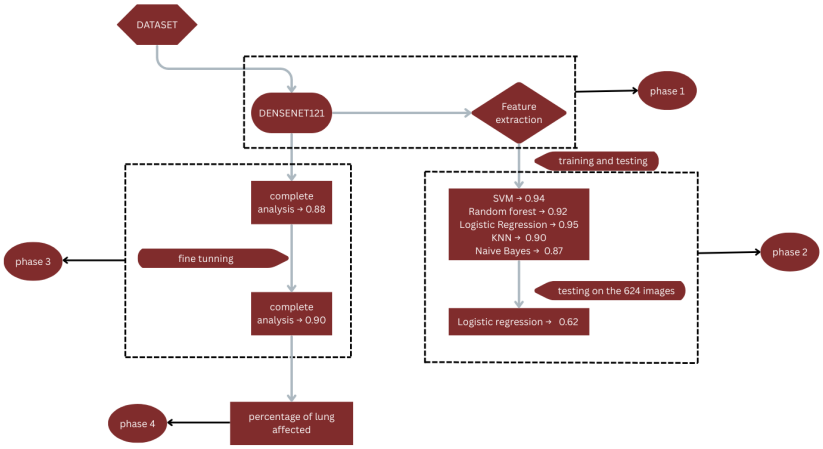

Figure 1: Overall workflow of the proposed multi-phase pneumonia detection and severity estimation system.

5.1 Dataset Description and Preprocessing

The Kaggle Chest X-Ray Images (Pneumonia) dataset is used in this study. The dataset contains labeled chest X-ray images categorized into NORMAL and PNEUMONIA classes. Images vary in resolution, contrast, and illumination, making preprocessing essential.

All images are resized to 224×224 pixels to match the DenseNet121 input requirements. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance local contrast and improve visibility of lung opacities. Pixel values are normalized to the range $[0, 1]$.

Data augmentation techniques including rotation, width and height shifting, zooming, horizontal flipping, and brightness adjustment are employed during training to reduce

overfitting and improve generalization. The dataset is split into training and validation sets using an 80:20 ratio.

5.2 Phase I: Deep Feature Extraction Using DenseNet121

DenseNet121 pretrained on ImageNet is used as the base model for pneumonia detection. Dense connectivity allows efficient feature reuse and improved gradient flow, making it well-suited for medical imaging tasks.

The network is modified by removing the top classification layers and adding a Global Average Pooling layer followed by a fully connected layer with ReLU activation and dropout for regularization. The final output layer uses a sigmoid activation for binary classification.

During training, the earlier layers of DenseNet121 are frozen, while the last 20 convolutional layers are fine-tuned using a low learning rate. Early stopping is applied to prevent overfitting.

5.3 Phase II: Machine Learning Classification and Optimization

Deep features are extracted from the final convolutional layers of the trained DenseNet121 model. These feature maps are spatially averaged to form fixed-length feature vectors.

Support Vector Machine (SVM) with RBF kernel and Random Forest classifiers are trained on these deep features. Feature scaling is applied using standardization. Class imbalance is handled using class-weighted learning.

Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrices.

5.4 Phase III: Optimized DenseNet121-Based Pneumonia Classification

In Phase 3, an optimized deep learning–based pneumonia classification model was developed using a transfer learning approach with DenseNet121. The Chest X-ray Pneumonia dataset was loaded and split into training, validation, and test sets. To enhance lung

visibility and contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied during preprocessing, followed by normalization and resizing of images to 224×224 pixels. Data augmentation techniques such as horizontal flipping and zooming were employed during training to improve generalization. DenseNet121 pre-trained on ImageNet was used as the backbone network with its top layers removed, and a custom classification head consisting of global average pooling, batch normalization, fully connected layers, and dropout was added for binary classification.

The model was trained using a two-stage strategy to improve performance and stability. In the initial stage, the backbone network was frozen and only the custom layers were trained using class-weighted loss and early stopping to handle class imbalance and prevent overfitting. In the second stage, selected upper layers of the DenseNet121 network were unfrozen and fine-tuned using a reduced learning rate. The final model was evaluated on the test dataset using metrics such as classification report, confusion matrix, ROC curve, and precision–recall curve. Chest X-ray images predicted as pneumonia were then automatically extracted and saved for further analysis in subsequent project phases.

5.5 Phase IV: Severity Estimation

Lung segmentation is performed using a pretrained PSPNet model from the TorchXRayVision library. The model generates probability maps for anatomical structures, from which left and right lung masks are extracted.

The lung masks are resized to match the original image dimensions. Opacity detection is performed within the lung region using adaptive thresholding and global thresholding techniques. Morphological operations are applied to remove noise.

Disease severity is calculated as:

Lung Involvement (%) = (Opacity Pixels/Total Lung Pixels) × 100
This quantitative measure enables objective assessment of pneumonia severity.

The final phase of the proposed framework focuses on estimating the severity of pneumonia rather than limiting the system to binary classification. While pneumonia detection determines the presence of disease, it does not indicate the extent of lung involvement. To address this, lung regions are first segmented using a pretrained deep learning–based segmentation model, ensuring that analysis is restricted only to the lungs and excluding irrelevant anatomical structures such as ribs and the heart. Within the segmented lung regions, opacity detection is performed to identify pneumonia-affected areas. The severity

of infection is then quantified by calculating the percentage of lung area affected, using pixel-level analysis of opacity regions relative to the total lung area.

5.6 Evaluation Metrics

Model performance is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provide a comprehensive assessment of model effectiveness and reliability.

6 Results and Discussion

6.1 Performance Comparison of the Five ML Models

Support Vector Machine (SVM)

As illustrated in Fig.1, the Support Vector Machine (SVM) classifier achieved a validation accuracy of 95.5%, accompanied by a high cross-validation F1-score of 0.9768 ± 0.0034, indicating strong generalization and stability across folds. The confusion matrix demonstrates that the model correctly classified 260 NORMAL and 504 PNEUMONIA samples. Only 28 pneumonia cases were misclassified as normal, highlighting the model's effectiveness in minimizing false negatives. The high recall of 0.95 for the pneumonia class is particularly significant for clinical diagnosis, where early and accurate disease detection is critical.

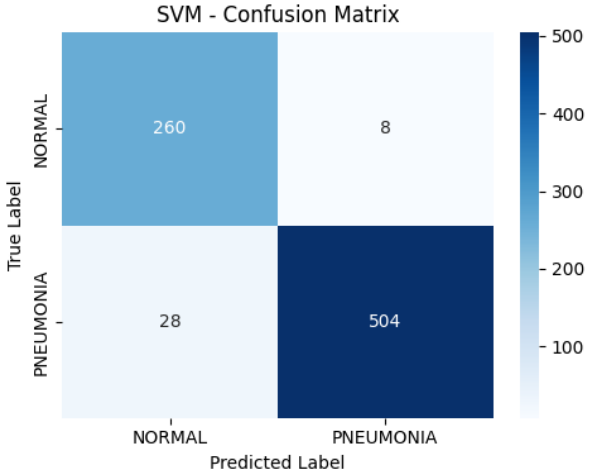

Figure 2: Confusion matrix of the SVM

Random Forest

The performance of the Random Forest classifier is presented in Fig.2 . The model achieved a validation accuracy of 91.5% with a cross-validation F1-score of 0.9596 ± 0.0059. From the confusion matrix, it is observed that 500 pneumonia cases were correctly identified, while 36 normal samples were incorrectly classified as pneumonia. Although the model demonstrates robust detection capability, the increased false positive rate for normal cases indicates comparatively lower specificity than SVM and Logistic Regression.

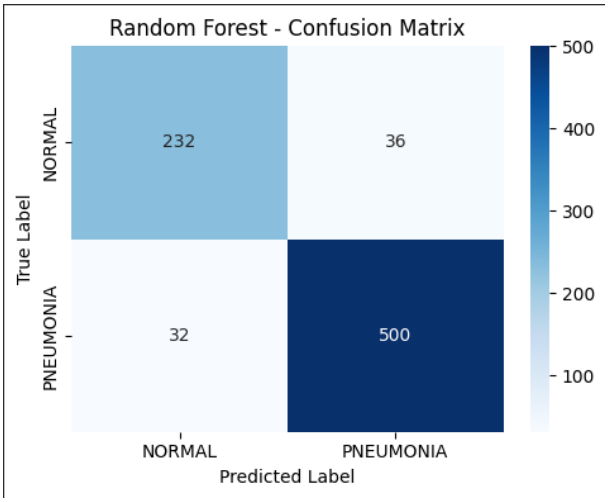

Figure 3: Confusion matrix of the Random Forest

Logistic Regression

As shown in Fig.3, Logistic Regression yielded the best overall performance among the evaluated classifiers. The model achieved the highest validation accuracy of 96.0% and a cross-validation F1-score of 0.9784 ± 0.0054. The confusion matrix indicates correct classification of 510 pneumonia and 258 normal cases, with only 22 pneumonia samples misclassified. The balanced precision and recall values across both classes suggest that Logistic Regression effectively exploits the discriminative power of CNN-extracted features and generalizes well to unseen validation data.

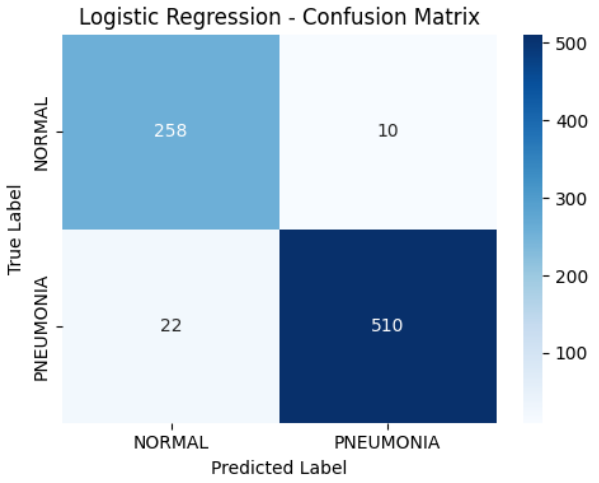

Figure 4: Confusion matrix of the Logistic Regression

K-Nearest Neighbors (KNN)

The confusion matrix for the K-Nearest Neighbors classifier is depicted in Fig.4 . The model achieved a validation accuracy of 89.9% with a cross-validation F1-score of 0.9488 ± 0.0056. Despite achieving high recall (0.94) for the NORMAL class, the model misclassified 66 pneumonia cases as normal. Such false negatives reduce clinical reliability and can be attributed to KNN's sensitivity to neighborhood selection and its limitations in handling high-dimensional deep feature spaces.

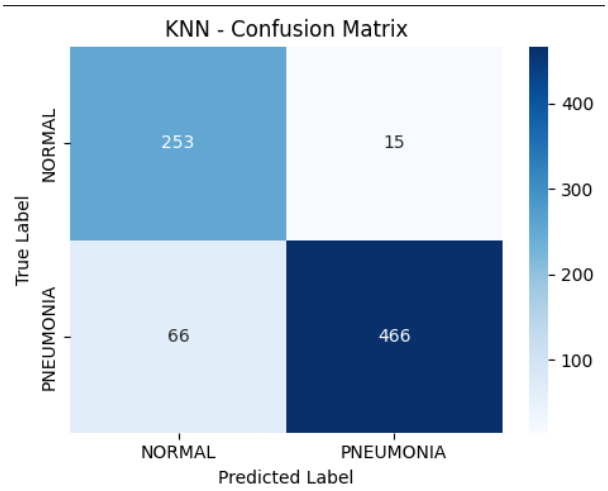

Figure 5: Confusion matrix of the KNN

Naive Bayes

As illustrated in Fig.5 , the Naive Bayes classifier recorded the lowest validation accuracy of 88.3%, with a cross-validation F1-score of 0.9200 ± 0.0068. The confusion matrix reveals a high recall of 97% for the NORMAL class but a reduced recall of 84% for pneumonia cases. This performance degradation is expected, as the Naive Bayes classifier assumes conditional independence among features—an assumption that does not hold for complex, correlated representations derived from deep convolutional neural networks.

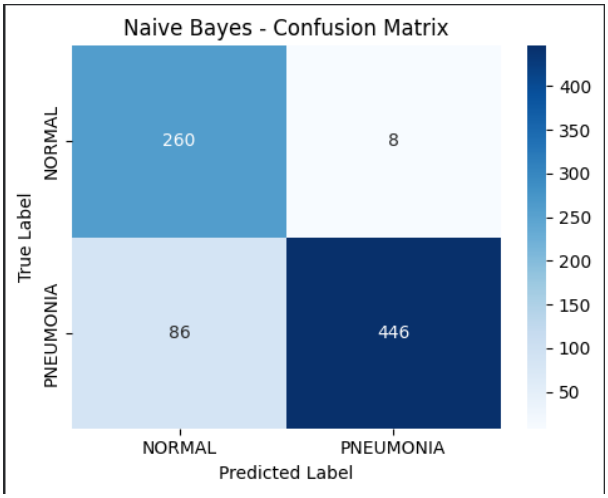

Figure 6: Confusion matrix of the Naive Bayes

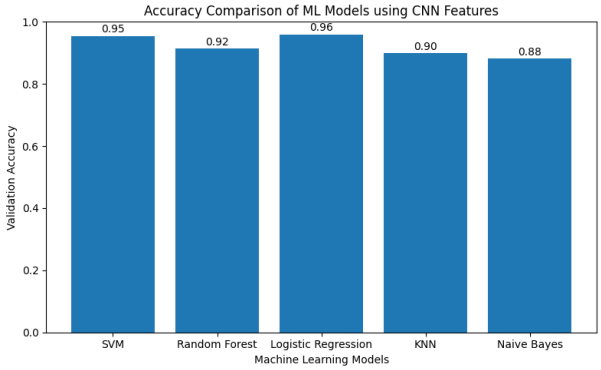

Figure 7: Performance Comparison of ML Models

6.2 Performance of the DENSENET121 Model

To evaluate the effectiveness of the end-to-end deep learning approach, the DenseNet121 model was trained in two stages: initial feature extraction (frozen base) and fine-tuning.

Initial Training (Without Fine-tuning)

In the first stage, the convolutional base of DenseNet121 was frozen, and only the custom fully connected layers were trained. Fig. 7 displays the training and validation metrics over the epochs. The model showed steady convergence, but the performance was limited by the generic features learned from ImageNet.

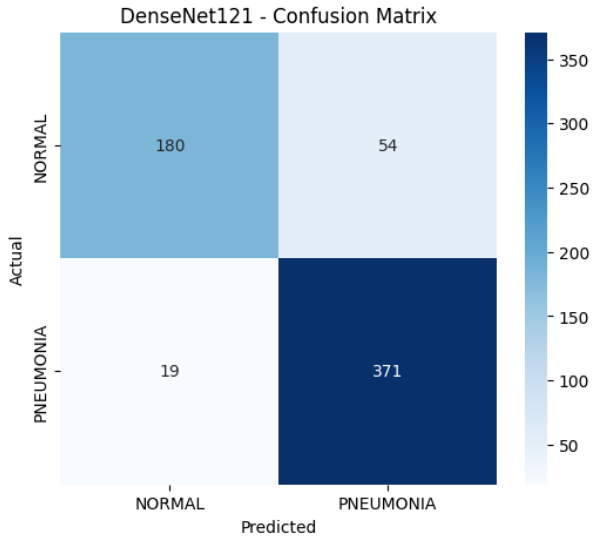

Figure 8: Training and validation metrics without hyperparameter fine-tuning

Fine-tuning Performance

In the second stage, the last convolutional block of the DenseNet121 architecture was unfrozen, and the model was re-trained with a lower learning rate. As shown in Fig. 8, fine-tuning significantly improved the validation accuracy and reduced the loss, allowing the model to adapt specific radiographic features of pneumonia.

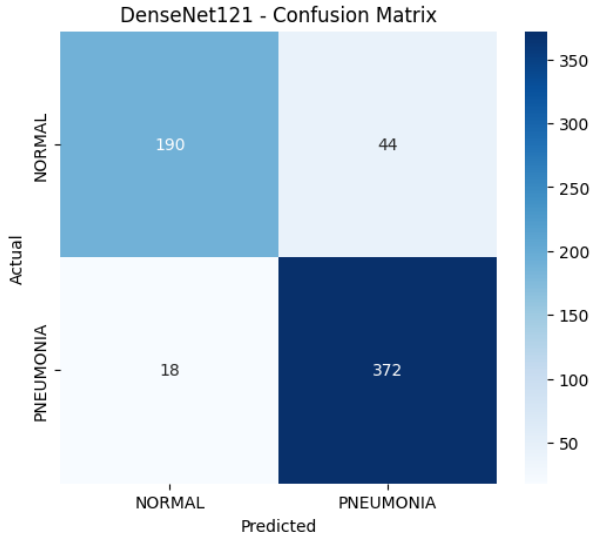

Figure 9: Training and validation metrics after fine-tuning

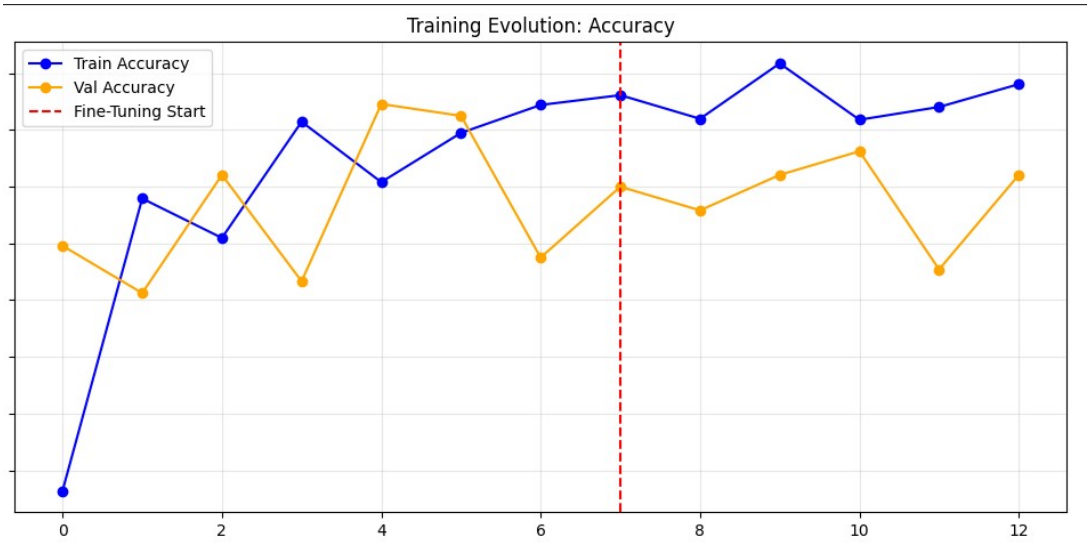

Figure 10: Combined training and validation accuracy evolution of the DenseNet121 model, with the fine-tuning phase indicated.

6.3 Final Evaluation on Test Set

Based on the comparative analysis of the machine learning classifiers, Logistic Regression was identified as the most robust model. It was subsequently evaluated on the complete unseen test dataset to assess real-world performance.

```
Extracting features from the REAL Test Set (624 images) for ML evaluation...

==================================================
FAIR TEST RESULT: LOGISTIC REGRESSION ON HARD TEST SET
==================================================
Test Accuracy: 0.6250

Classification Report:
              precision    recall  f1-score   support

      NORMAL       0.00      0.00      0.00       234
   PNEUMONIA       0.62      1.00      0.77       390

    accuracy                           0.62       624
   macro avg       0.31      0.50      0.38       624
weighted avg       0.39      0.62      0.48       624
```

Figure 11: Confusion matrix of Logistic Regression on the Test Set

Final Evaluation of DenseNet121 Model

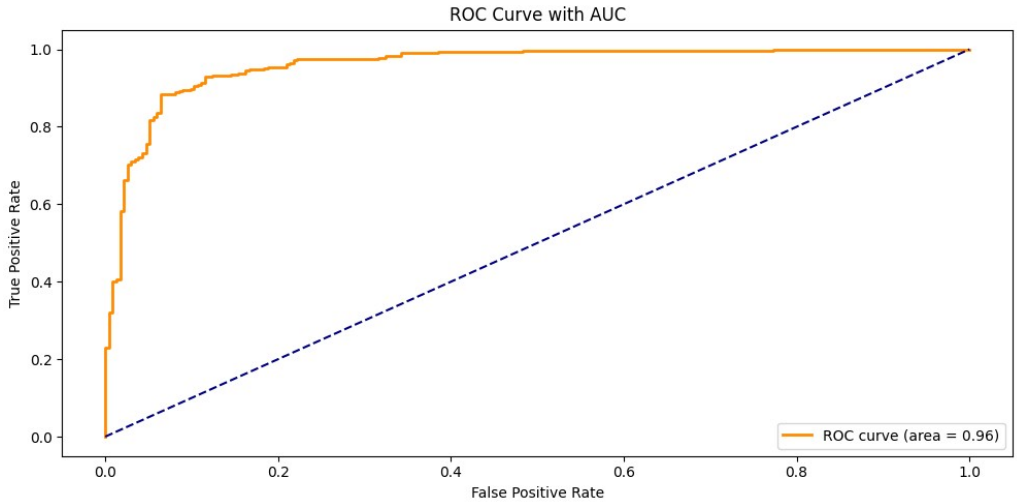

Figure 12: ROC curve of the fine-tuned DenseNet121 model evaluated on the test dataset.

The above figure illustrates the Receiver Operating Characteristic (ROC) curve of the fine-tuned DenseNet121 model on the unseen test dataset. The high Area Under the Curve (AUC) value of approximately 0.96 demonstrates strong discriminative capability and confirms superior generalization performance compared to classical machine learning models.

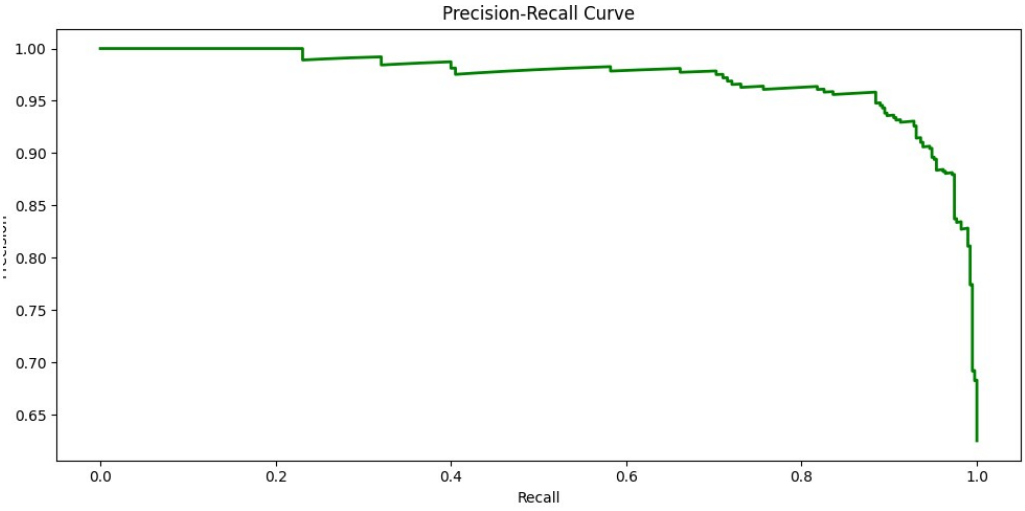

Figure 13: Precision–Recall curve of the fine-tuned DenseNet121 model evaluated on the test dataset.

The above figure shows the Precision–Recall curve of the DenseNet121 model on the test dataset. The consistently high precision across a wide range of recall values indicates reliable pneumonia detection performance, particularly under class imbalance conditions commonly observed in medical imaging datasets.

7 Conclusion

This project proposed a hybrid artificial intelligence framework for automated pneumonia detection, localization, and severity assessment using chest X-ray images, addressing key challenges of manual diagnosis such as subjectivity, time consumption, and variability. By integrating CLAHE-based preprocessing, transfer learning with a fine-tuned DenseNet121 model, and deep feature–based classical machine learning classifiers, the system achieved robust and reliable classification performance, with fine-tuning significantly improving generalization as validated through confusion matrix analysis, ROC, and precision–recall curves. Beyond binary detection, the framework incorporated lung segmentation and pixel-level opacity analysis to quantitatively estimate lung involvement and stratify patients into clinically meaningful risk categories, enhancing practical applicability. The inclusion of Grad-CAM–based visual explanations further improved interpretability and clinical trust, demonstrating the potential of the proposed framework as a scalable, accurate, and clinically relevant decision-support system for real-world pneumonia diagnosis and management.

Bibliography

[1] A. K. Winarto, K. Aprilian, A. A. S. Gunawan, and R. C. Pradana, "Smart Diagnosis of Pneumonia: A Comparative Analysis of CNN and Transformer-Based Models on Chest X-Ray Images," *2025 1st International Conference on Artificial Intelligence Technology (ICoAIT)*, Jakarta, Indonesia, 2025, pp. 60-65.

[2] S. K. Behera, K. M. Gopal, and S. B. Punuri, "A Deep Learning-Based Pneumonia Detection System with Explainable AI for Medical Decision Support," *2025 11th International Conference on Communication and Signal Processing (ICCSP)*, Odisha, India, 2025, pp. 694-699.

[3] R. Bhamare, S. Kamble, V. Pandey, S. Paithanka, Z. Sayyed, and R. Walia, "Building a Deep CNN-BiLSTM Model for Pneumonia Diseases Prediction on Chest X-Ray Images," *2025 IEEE International Conference on Advances in Computing Research On Science Engineering and Technology (ACROSET)*, Mumbai, India, 2025.

[4] S. Khojare and S. Bhattacharjee, "Pneumonia Identification based on Transfer Learning over Chest X-Ray Images," *2025 International Conference on Emerging Trends in Industry 4.0 Technologies (ICETI4T)*, Jhunjhunu, India, 2025.

[5] M. R. Kumar and B. G. Vani, "Implementation of Deep Learning Mechanisms for Detection and Classification of Pneumonia on Chest X-Ray Dataset," *2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST)*, Puttur, India, 2024.

[6] K. Menaka, S. Rao G, Bala P. K, Ramprawin U, and S. Prabhu S, "Detection and Recognition of Pneumonia in Chest X-ray Images using Deep Learning Methods," *2025 International Conference on Emerging Technologies in Computing and Communication (ETCC)*, Chennai, India, 2025.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261-2269.