The University of Lahore

December 30, 2025

Artificial Intelligence Project Documentation

Lung Disease Detection

Prepared by :

Eesha Javed	70136133
Sakina Batool	70132094
Zehra Batool	70132092
Imbisat Nawaz	70134374

Table of Contents

1. Introduction

Lung diseases are among the leading causes of illness and death worldwide. Pneumonia, in particular, is a severe lung infection that inflames the air sacs in one or both lungs, which may fill with fluid or pus. Early and accurate diagnosis of pneumonia is critical for effective treatment and patient recovery.

Chest X-ray imaging is one of the most commonly used diagnostic tools for detecting pneumonia. However, manual analysis of X-ray images by radiologists is time-consuming and subject to human error, especially in high-volume clinical environments.

With advancements in Artificial Intelligence and Deep Learning, automated systems can assist healthcare professionals by accurately analyzing medical images. This project focuses on developing a **Convolutional Neural Network (CNN)**–based system to automatically detect pneumonia from chest X-ray images, improving diagnostic efficiency and reliability.

2. Problem Statement

Manual interpretation of chest X-ray images for pneumonia diagnosis requires expert radiologists and considerable time. In resource-limited settings, the lack of trained professionals may delay diagnosis, leading to severe health complications.

Traditional machine learning methods rely on handcrafted features that often fail to capture complex patterns in medical images. Therefore, an automated deep learning–based approach is required to accurately classify chest X-ray images as **Pneumonia** or **Normal**, reducing diagnostic workload and improving accuracy.

3. Objectives

The main objectives of this project are:

- To develop a deep learning model for pneumonia detection using chest X-ray images
- To automate feature extraction using convolutional neural networks
- To classify X-ray images into Pneumonia and Normal categories
- To evaluate model performance using standard evaluation metrics
- To demonstrate the effectiveness of deep learning in medical image analysis

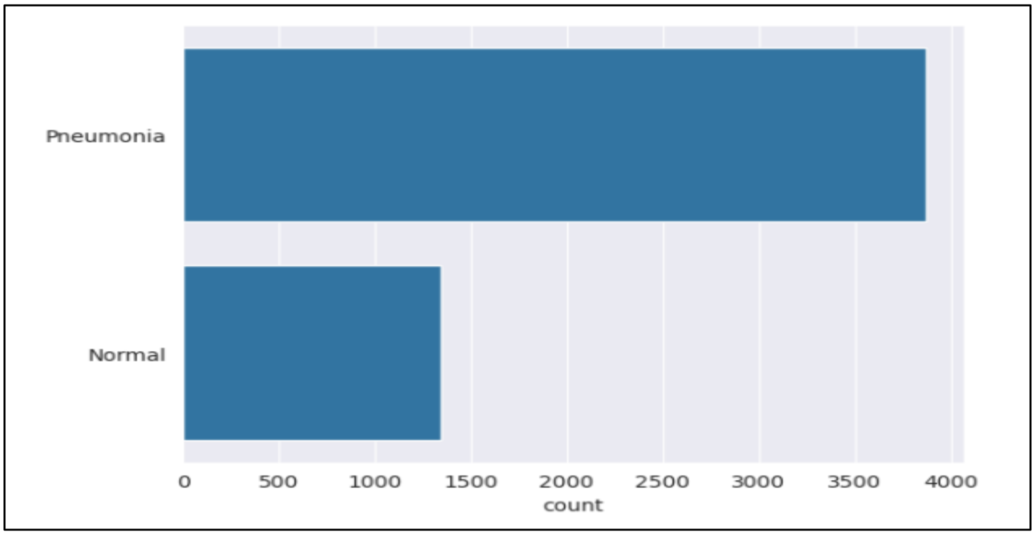
4. Dataset Description

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).**

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.

Dataset Distribution

5863 samples



This distribution ensures sufficient data for training, validation, and unbiased testing of the deep learning model.

5. Existing Solutions

Traditional pneumonia detection methods rely on radiologists manually examining chest X-ray images. Classical machine learning approaches use handcrafted features such as texture and intensity, which often fail to generalize well.

Recent studies show that deep learning models, particularly CNNs, outperform traditional methods by automatically learning discriminative features directly from image data.

6. Proposed Methodology

The proposed system uses a **Convolutional Neural Network (CNN)** to detect pneumonia from chest X-ray images. The system workflow includes data preprocessing, data augmentation, model training, and evaluation.

6.1 Data Preprocessing

Data augmentation techniques such as rotation, zooming, width/height shifting, and horizontal flipping are applied to increase data diversity and reduce overfitting.

- Images are converted to grayscale
- All images are resized to **150 × 150 pixels**
- Pixel values are normalized to the range [0, 1]
- Invalid or corrupted images are skipped

6.2 Model Architecture

The CNN architecture consists of:

- Multiple convolutional layers for feature extraction
- Batch normalization layers to stabilize training
- Max-pooling layers to reduce spatial dimensions
- Dropout layers to prevent overfitting
- Fully connected dense layers for classification
- Sigmoid activation function for binary classification

The final output layer predicts whether the input X-ray image belongs to **Pneumonia** or **Normal** class.

6.3 Training and Validation Process

- The model is trained using the **binary cross-entropy** loss function
- **RMSprop** optimizer is used for efficient learning
- Learning rate reduction is applied using **ReduceLROnPlateau**
- The model is trained for **12 epochs**
- Validation data is used to monitor generalization performance

7. Tools and Technologies

- Programming Language: **Python**
- Deep Learning Framework: **TensorFlow / Keras**
- Development Environment: **Google Colab**
- Libraries: NumPy, Pandas, OpenCV, Matplotlib, Seaborn, Scikit-learn

8. Experimental Results

The trained CNN model demonstrated strong performance in detecting pneumonia from chest X-ray images.

8.1 Overall Performance Summary

- **Testing Accuracy: ~90.86%**
- **Testing Loss: ~0.28**

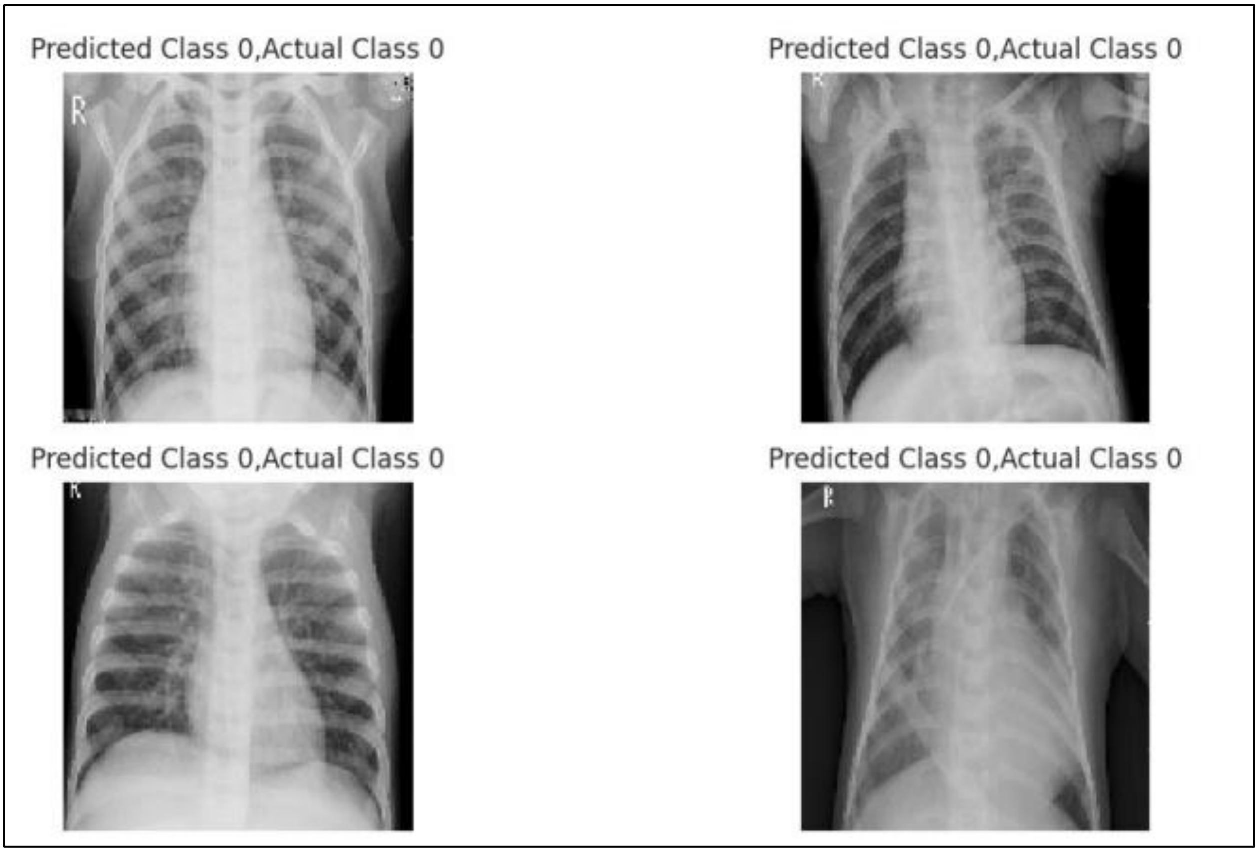
The results indicate that the model generalizes well to unseen data.

8.2 Class-wise Performance

Class	Precision	Recall	F1-Score	Support
Pneumonia	0.94	0.91	0.93	390
Normal	0.86	0.91	0.88	234
Overall Accuracy			**0.91**	624

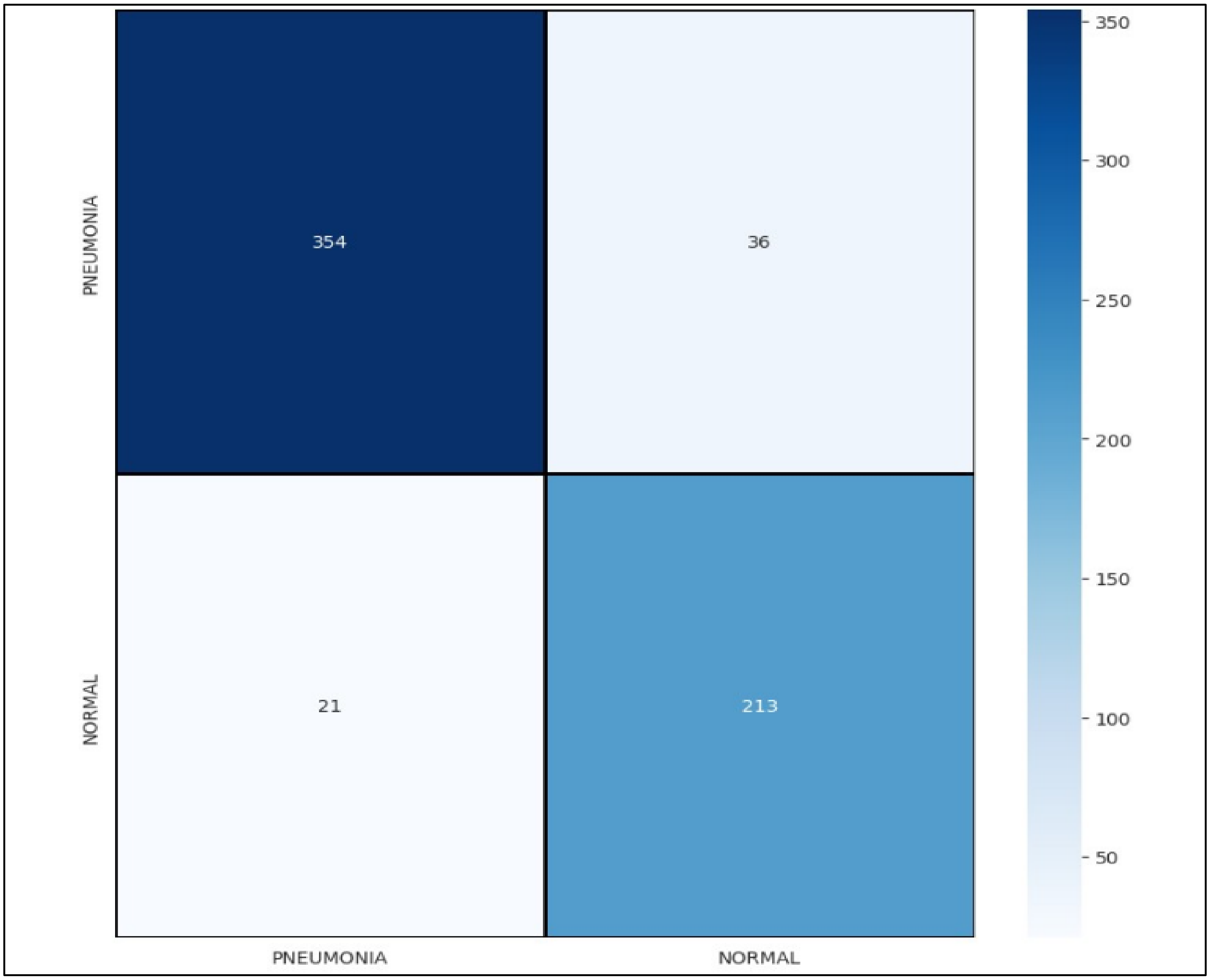
8.3 Sample Output Predictions

Sample Lungs X-Ray Images Input:



The model outputs the predicted tumor class along with probability scores for all classes, providing insight into prediction confidence.

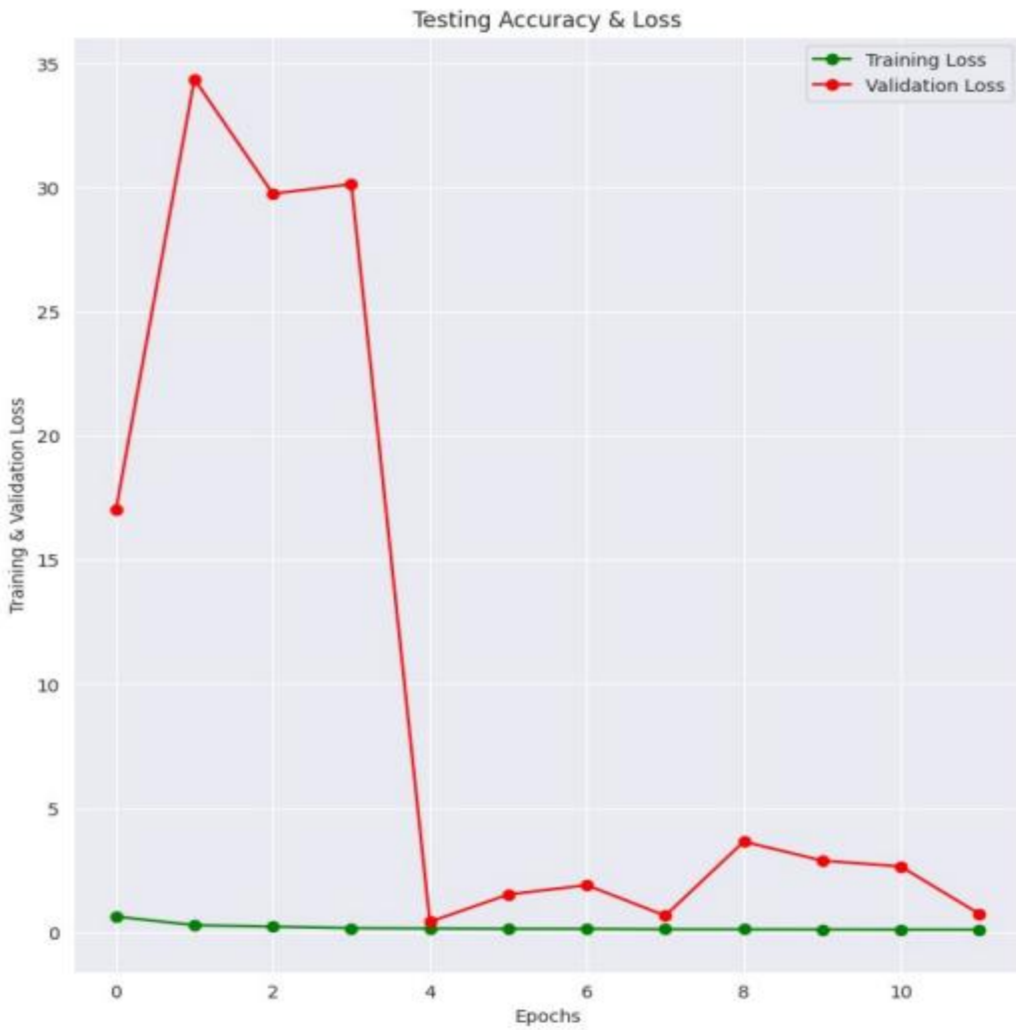
8.4 Confusion Matrix



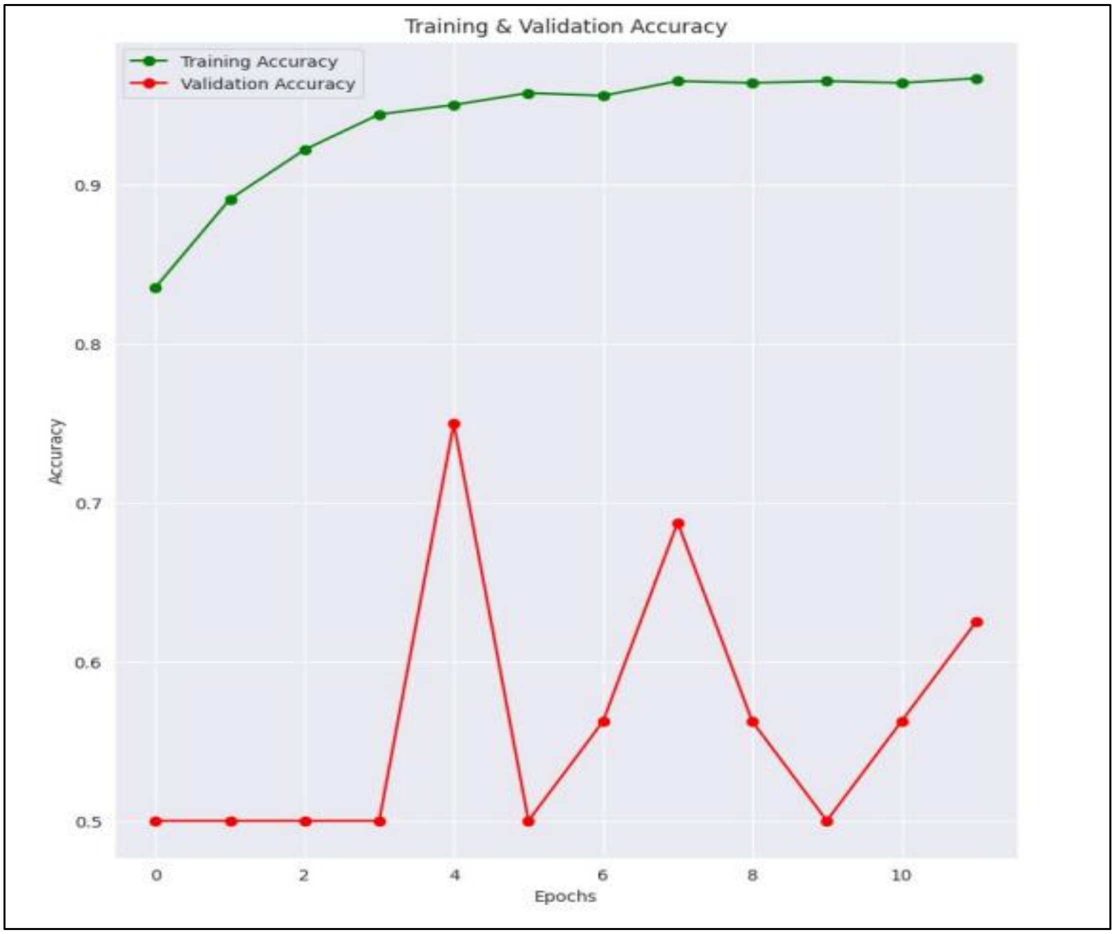
The confusion matrix shows that most predictions lie along the diagonal, indicating a very high number of correct classifications and minimal misclassification between pneumonia categories.

8.5 Training and Validation Curves

Testing Accuracy & Loss vs Training & Validation Loss Graph:



Training vs Validation Accuracy Graph:



The curves show stable convergence and minimal overfitting throughout the training process.

9. Evaluation Metrics

The following metrics were used to evaluate model performance:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**
- **Confusion Matrix**

10. Discussion

The experimental results demonstrate that CNN-based deep learning models are highly effective for pneumonia detection from chest X-ray images. Automated feature extraction allows the model to capture complex lung patterns that are difficult to identify using traditional techniques.

Despite the dataset imbalance, data augmentation helped improve generalization. However, model performance may vary depending on image quality and dataset size.

11. Conclusion

This project successfully presents a deep learning–based approach for automated pneumonia detection using chest X-ray images. The proposed CNN model achieves high accuracy and demonstrates the potential of artificial intelligence in medical diagnostics.

The system can serve as a reliable decision-support tool for healthcare professionals, assisting in early diagnosis and reducing diagnostic workload.

12. References

1. Chest X-Ray Pneumonia Dataset – Kaggle
2. TensorFlow & Keras Documentation
3. National Library of Medicine