Report: Pneumonia Detection Using Deep Learning

Internship – DevelopersHub Corporation

Week 3, Project 2

# 1. Introduction

This report presents the development and implementation of a pneumonia detection system using deep learning techniques, specifically Convolutional Neural Networks (CNNs). The project was undertaken as part of Week 3, Project 2 of the internship at DevelopersHub Corporation. The primary objective of the system is to classify chest X-ray images into two categories: Normal and Pneumonia.

# 2. Dependencies

The following libraries were required for implementation:  
- kagglehub – dataset acquisition from Kaggle  
- tensorflow – development of deep learning models  
- matplotlib – visualization of training metrics  
- seaborn – visualization of the confusion matrix  
- scikit-learn – evaluation of classification performance

# 3. Dataset

The dataset used in this project was the Chest X-ray Pneumonia dataset obtained from Kaggle. It contains chest X-ray images categorized into Normal and Pneumonia. The dataset was divided into three subsets:  
- Training set – for model training  
- Validation set – for model tuning and validation  
- Testing set – for final evaluation

# 4. Data Preprocessing

To ensure consistency and improve training performance, the following preprocessing steps were applied:  
- All images were resized to 128 × 128 pixels.  
- Pixel values were normalized to a range of 0 to 1.  
- Data was loaded in batches of 32 images to optimize computational efficiency.

# 5. Model Architecture

A Convolutional Neural Network was designed with the following architecture:  
1. Convolutional Layer (32 filters, 3×3 kernel, ReLU activation)  
2. MaxPooling Layer (2×2)  
3. Convolutional Layer (64 filters, 3×3 kernel, ReLU activation)  
4. MaxPooling Layer (2×2)  
5. Flatten Layer  
6. Dense Layer (128 units, ReLU activation)  
7. Dropout Layer (rate: 0.5)  
8. Dense Output Layer (1 unit, Sigmoid activation)  
  
The model was compiled using the Adam optimizer and the Binary Crossentropy loss function.

# 6. Model Training

The model was trained for 10 epochs. The training set was used for learning, while the validation set was employed to monitor generalization performance. Training accuracy and loss were recorded across all epochs.

# 7. Model Evaluation

The trained model was evaluated using the test dataset. The following metrics were applied:  
- ROC Curve and AUC Score – assessment of overall classification performance  
- Confusion Matrix – visualization of prediction accuracy  
- Classification Report – detailed performance metrics, including precision, recall, F1-score, and accuracy

# 8. Results

- The ROC Curve achieved a high AUC score, indicating strong discriminative capability.  
- The confusion matrix confirmed effective classification of both Normal and Pneumonia cases.  
- The classification report demonstrated high precision, recall, and F1-score, validating the robustness of the model.

# 9. Conclusion

The CNN-based pneumonia detection system successfully classified chest X-ray images into Normal and Pneumonia categories with a high degree of accuracy. The results highlight the potential of deep learning in medical image classification and demonstrate its applicability as a decision-support tool for healthcare professionals in diagnostic processes.