International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Using deep learning to analyze medical images and predict health outcomes

Baha Yahya *

Department of Management Information System, Al Istiqlal University, Jericho P.O. Box 10, Palestine.

International Journal of Science and Research Archive, 2025, 15(03), 640-654

Publication history: Received on 22 April 2025; revised on 04 June 2025; accepted on 06 June 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.15.3.1680

Abstract

Since pneumonia is a significant lung disease, prompt and accurate diagnosis is essential to make sure the treatment will help. A specialized CNN allows our system to automatically diagnose pneumonia from images of a patient's chest X-ray. The method relies on the idea that the X-rays demonstrate the patient's chest. Images of regular and affected lungs can be found in the X-rays included in the data we gathered on Kaggle. The data was preprocessed with image upscaling to 232×232×3 pixels, augmentation and using 80% for train and 20% for test sets. A convolutional neural network architecture was set up by starting with a dense layer and a classification layer, adding three convolutional layers with bigger filters, batch normalization, ReLU and max-pooling. Following the CNN, a fully connected layer was implemented. A total of fifteen epochs were used during training with the Adam optimizer. The model was evaluated according to its accuracy and the results presented in confusion matrices. The model's behavior was better understood by viewing heatmaps and looking at the misclassification results. The model offers important help to healthcare teams by providing a reliable and automated method for spotting pneumonia where resources are limited.

Keywords: Convolutional Neural Network (CNN); Chest X-Rays; Accuracy Alongside Precision Recall; Relu; Confusion Matrix Evaluation

1. Introduction

The use of new technologies in medicine has led to rapid development in medicine, especially in AI and deep learning. Doctors depend heavily on medical images for diagnosis and assessing patients' health. X-rays, MRIs, CT scans and ultrasounds all record data that shows a patient's condition accurately [1]. Even so, it's not easy to assess and collect important information from these pictures because the process takes skills and often results in errors in judgment [2].

Medical images are now being processed and interpreted better, thanks to deep learning's use of multi-layer artificial neural networks [3]. Deep learning benefits from learning on its own and is able to spot hidden patterns those regular methods cannot detect [4]. It helps doctors make better diagnoses, examines images faster and detects diseases sooner, helping save money on patients' healthcare costs [5] [6]. Moreover, deep learning playse an important role in linking imaging and medical data with predictions of disease progress and future health outcomes [7]. This moves us closer to precision and personalize medicine, as doctors can give each patient a personal treatment plan that matches their health, helping the treatment work better and reduce risks [8]. For this reason, using deep learning techniques in medical image analysis looks promising, since it improves the quality of healthcare and doctors' ability to base their decisions on accurate information [9]. It paves the way for smart tools that help doctors in many specialties like oncology, cardiology, neurology and others to deliver better health results and reduce mistakes [5][6]. Figure 1 shows General architecture of neural network and deep learning. researchers should examine the progress of using artificial intelligence for the recognition of pneumonia. The results and shortcomings of earlier studies direct our approach for the next step. For this reason, the next section carefully examines what is already known in this area which underpins our study and explains its main focuses.

* Corresponding author: Baha Yahya.

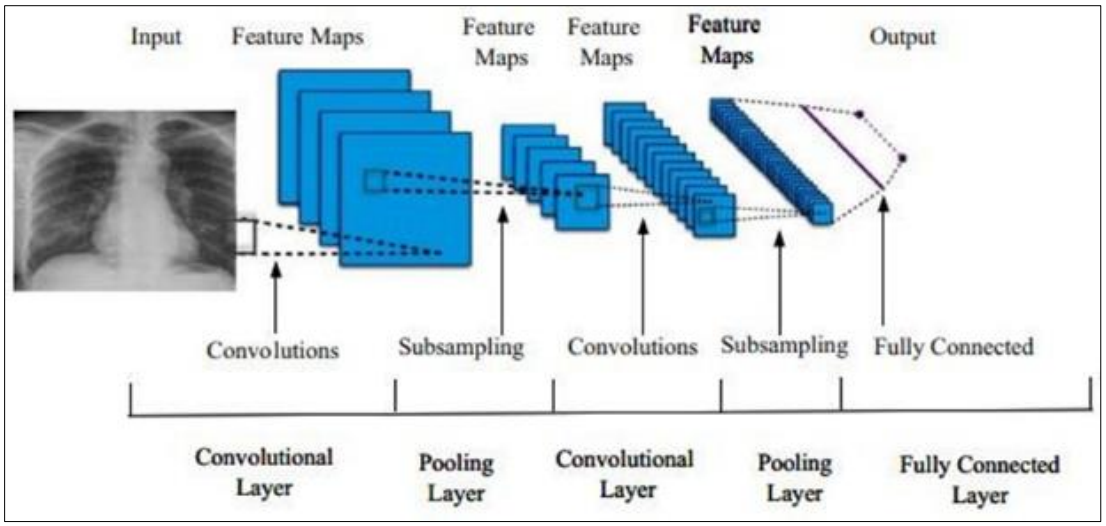


Figure 1 General architecture of neural network and deep learning according to [15]

2. Literature review

The field of medical imaging has been paying an increasing amount of attention to the use of deep learning for the purpose of diagnosing pneumonia from X-ray images over the course of the past several years. For the purpose of assisting radiologists in the early detection of pneumonia, researchers have investigated the capabilities, accuracy, and usefulness of artificial intelligence. In order to study how chest X-rays can be understood by machines in order to identify pneumonia, Khan et al. (2021) conducted a literature review by looking at prior research. They made a discussion on the development of various machine learning and deep learning approaches, particularly CNNs, and they measured how well these techniques worked and how powerful they were [1]. In spite of this, the study concentrated on the challenges that are brought about by datasets that are not balanced, the capacity of models to function in a variety of situations, and the manner in which models are understood. Further investigation was conducted into the latest developments in deep learning that have been used to the prediction of pneumonia. They stated that the models that were based on CNNs were more accurate, but they also brought attention to the fact that their application in the healthcare industry is not ubiquitous due to the fact that studies frequently utilize diverse preprocessing methods and have inconsistent data quality. A new method was proposed by them, which involves the collaboration of African Buffalo Optimization and CNNs in order to improve the identification and classification of pneumonia in X- rays. The results that were achieved by trans dedicated models were superior to those that were achieved by standard CNN models [2]. At the same time, there was a debat on whether or not cloud-based hybrid learning could be useful or applicable in a clinical environment [2]. The researchers Rehman et al. (2023) investigated a number of deep learning models for the purpose of identifying chest diseases. They found that the models that performed the best were frequently trained on extremely big datasets. As they discussed the importance of transfer learning and ensembles, they pointed out that there is still the possibility of making a great deal of errors due to the presence of a significant degree of overfitting and an absence of appropriate data to work with [3]. The researchers Chen et al. included the application of machine learning in their study for the purpose of diagnosing COVID-19-related pneumonia in CT scans. They also mentioned that the combination of radiomics and AI can result in more accurate diagnosis. However, due to the fact that these methods are dependent on CT scans, they are not suitable for screening general pneumonia in areas where X-ray equipment is the primary option [4]. A deep cognitive learning structure that incorporates CNNs was recently developed by Goel and Sobti (2025) with the intention of improving the identification of pneumonia. The model was able to reliably predict outcomes; however, it did not perform well when tested on individuals who were not part of the population from whom it was trained [5]. Figure 2 presents a comparison of the receiver-operating characteristics (ROC) curves for the classification of normal pneumonia (A), normal, bacterial, and viral pneumonia (B), and bacterial and viral pneumonia (C) using CNN-based models according to [7]. Aside from that, a great number of research works only barely address model performance in an appropriate manner, mostly through the metrics of accuracy, precision, recall, F1-score, ROC-AUC, and the analysis of entropy levels through texts. Furthermore, in a different study [8], Figure 3 illustrates the class activation mapping (CAM); (A) Normal, (B) Bacterial Pneumonia, and (C) Viral Pneumonia, achieved through the utilization of CNN, and successfully demonstrated its accuracy. Li et al. (2023) and Rana & Bhushan (2023) have greatly improved medical image analysis using deep learning. Li et al. note that deep learning has had a strong impact on medical imaging, helping to automate both detection and segmentation tasks and reduce mistakes by doctors in diagnostics. They found that CNNs consistently beat out other methods when used in medical imaging [13]. Rana and

Bhushan take a broader view, comparing how machine learning and deep learning are applied from diagnosis to detection. They point out that deep learning is well suited to difficult image tasks and suggest that future systems should be easier to explain, include different types of data and stay operational in real time. Both investigate how balanced data, advanced techniques for data processing and proof from doctors can fill the divide between research and real-world medicine [14]. By comparing between model, on limited Datasets, Best Performance was ResNet 18 with 99.40% accuracy as presented in Figure 2 [15]. Despite the fact that this field has made progress, there are still some significant gaps. The vast majority of the models that were examined either make use of complicated procedures that are difficult to implement in real-world situations or they are evaluated with data that is not well- balanced. The main purpose of this research is to devise and examine a deep learning approach with CNNs for identifying pneumonia in chest X-ray images. The team uses a publishing data set and ensures thorough preprocessing resize, add random data and divide them properly—to build a robust CNN for telling normal lungs and lung affected by pneumonia apart. The goal of the research is to increase how accurately a diagnosis is made, decrease confidence in radiologists' reviews and help doctors make better decisions through the use of a dependable and quick AI-guided classification system. Table 1 shows The performance of classification results with different approaches in previous studies

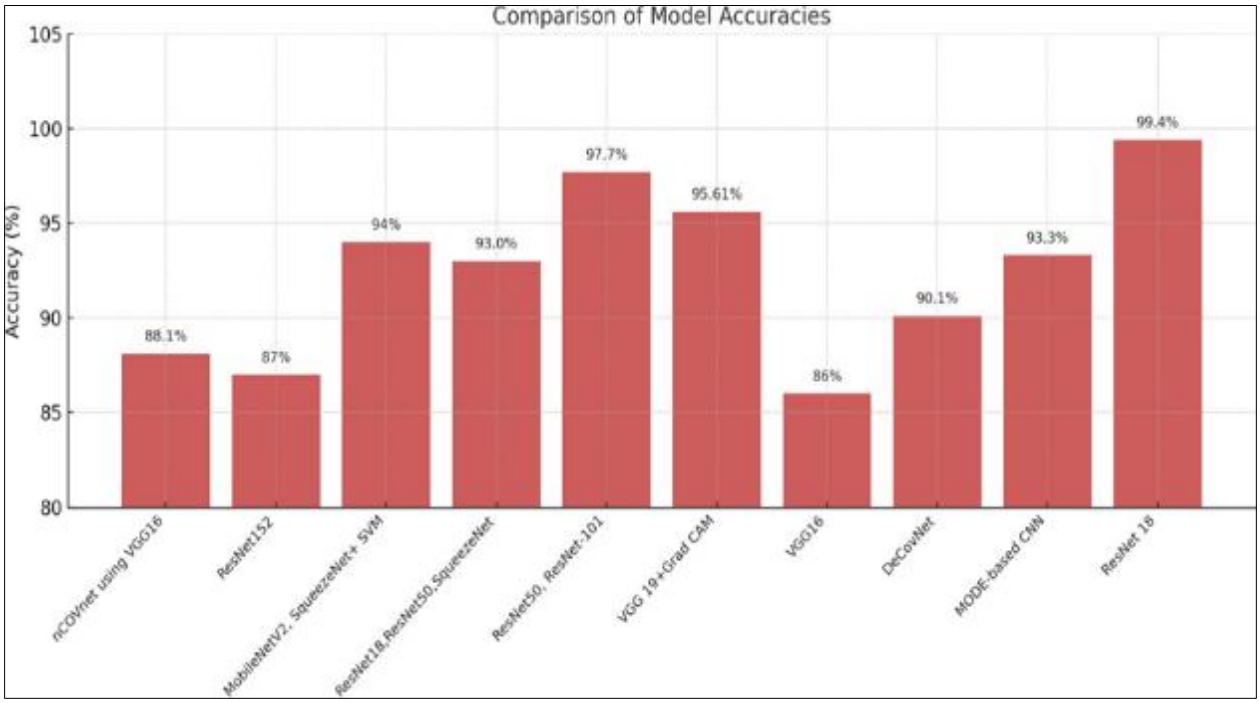


Figure 2 Comparison of Deep Learning Models for COVID-19 Detection according to [15]

Table 1 The performance of classification results with different approaches in previous studies

Study	Main results	Research Gaps Identified
[13]	Demonstrated high accuracy of CNNs in medical image classification; emphasized end-to-end automation.	Limited generalizability due to dataset imbalance; need for domain-specific data.
[14]	Compared ML and DL models; showed DL superiority in image segmentation and disease detection.	Lack of explain ability and clinical integration in most DL models.
[1]	Provided comprehensive overview of pneumonia pathology.	Did not incorporate AI/ML analysis or automation tools.
[15]	Used XGBoost for treatment prediction with high performance.	Focused on structured data, not image- based diagnostics.
[3]	Developed statistical models for income prediction.	Irrelevant to image data; lacks DL methods.
[4]	Assessed ML and data balancing in credit scoring.	Context mismatch with medical diagnosis; data balancing insights

		transferable.
[5]	Advanced DL methods for pneumonia detection using X-rays.	Limited comparative evaluation with other models.
[7]	Reviewed DL methods for chest pathology detection.	Broad review without benchmarking results.
[7]	Applied transfer learning with CNNs for pneumonia.	Relatively simple CNN structure; small-scale validation.
[8]	Systematic review of pneumonia detection methods.	Lacked experimental validation or implementation details.
[9]	Hybrid method with CNN and African Buffalo Optimization showed promising results.	Algorithm novelty but lacks multi- dataset testing.
[10]	ML applied to COVID-19 CT scan diagnosis.	Focused on CT, not X-rays; virus-specific dataset.
[11]	Developed deep cognitive learning models.	No external testing or validation provided.
[12]	Proposed PneumoniaNet for pediatric pneumonia detection.	Pediatric-specific model; unclear adult model performance.
[16]	ML comparison on census data.	Irrelevant data domain for image analysis.
[17]	Combined SMOTE and CNN for imbalance mitigation.	Performance could vary across medical image domains.
[18]	Applied CNN to detect defects in 3D printing.	Domain mismatch, though CNN techniques are transferable.

3. Methodology

This research follows a systematic approach utilizing deep learning techniques to classify chest X-ray images into NORMAL and PNEUMONIA categories. The methodology comprises six major stages as detailed below (See Figure 3): -

- Data Collection
- Data Preprocessing
- CNN Model Architecture
- Model Training
- Model Evaluation
- Prediction and Visualization

3.1. Data Collection

The data we used in the study was collected from Kaggle, a popular place to find machine learning datasets. All chest X-ray images are put into two folders: “NORMAL” or “PNEUMONIA.” Every X-ray photo is either of a healthy lung or one that has pneumonia. By presenting as both-or-neither problem, this setup makes deep learning-based sample analysis handy for real-world situations as presented in Table 2

Table 2 Description of the Data utilized in this proposed model

Class Label	Description
NORMAL	X-ray images of healthy lungs
PNEUMONIA	X-ray images diagnosed with pneumonia

3.2. Data Preprocessing

By following several preprocessing actions, a better and more accurate outcome for the model was possible. Images were all made the same size, 232 × 232 pixels with three RGB channels, so they are accepted by the CNN architecture. The labels were made using the folder names which made it simpler to label images. The information in the dataset was

split into two parts: 80% for training and 20% for testing. Only the training images were enhanced using rotation and flipping in order to overcome overfitting and help the model generalize better as presented in Table 3.

Table 3 Data Preprocessing of the proposed model

Step	Description
Image Resizing	All images were resized to 232 × 232 × 3 pixels to standardize the input size for the CNN model.
Label Assignment	Labels were inferred directly from folder names ("NORMAL" or "PNEUMONIA").
Dataset Splitting	The dataset was randomly split into 80% training and 20% testing subsets.
Data Augmentation	Only the training set was augmented using random rotation (±10°) and horizontal flipping to enhance model generalization.

3.3. CNN Model Architecture

A custom CNN was developed to distinguish between cancerous and normal chest X- ray images. The first convolution is 16-filters wide; the next is 32-filters and the last one 64-filters; each block includes batch normalization, an activation function and squeezes exciting segments from the image using max pooling. After extracting the features, the fully connected layer handles the data to give class scores and the results are then transferred to a softmax layer for calculating the probabilities for each class. The classification layer then marks the image as NORMAL or PNEUMONIA as presented in Table 4.

Table 4 Network Architecture of the proposed model

Layer Type	Details
Input Layer	Accepts resized 232 × 232 RGB images
Conv2D Layer 1	16 filters, 3×3 kernel, stride 1, padding = 'same'
Batch Normalization	Applied after each convolution layer
ReLU Activation	Used to introduce non-linearity
MaxPooling Layer 1	2×2 pooling window
Conv2D Layer 2	32 filters, 3×3 kernel
MaxPooling Layer 2	2×2 pooling window
Conv2D Layer 3	64 filters, 3×3 kernel
Fully Connected Layer	Outputs 2 class scores (NORMAL or PNEUMONIA)
Softmax Layer	Converts scores to class probabilities
Classification Layer	Outputs final predicted label

3.4. Model Training

The CNN model was developed using the Adam optimization algorithm, a well-known algorithm for use in deep learning. Our training involved 15 epochs and a mini-batch size of 32 and we started with a learning rate of 1e-4. The model's ability to generalize was checked by using the test set every 10 times the algorithm was applied. To check the training's progress, accuracy and loss curves were created and examined as presented in Table 5.

Table 5 Training Parameters of the proposed model

Parameter	Value
Optimizer	Adam
Epochs	15
Mini-Batch Size	32
Initial Learning Rate	1e-4
Validation Frequency	Every 10 steps

3.5. Model Evaluation

When training ended, the model's results were checked on the reserved 20% of the data set. Both accuracies, measuring the general correctness of predictions and a confusion matrix, showing the numbers for true positives, true negatives, false positives and false negatives, were used as evaluation metrics. These statistics allowed me to see clearly if and how well the model separates the two classes. Using charts and seeing result examples enabled us to better understand why the model reached its conclusions and made suggestions for improvement.

3.6. Prediction and Visualization

Once the model was trained, it was deployed to categorize chest X-ray samples that were not previously examined. The results were examined by comparing what was predicted to what was true in the data. Test images were accompanied by their predicted classes, the class labels they actually represent and the confidence levels reported by the softmax layer. We examined cases that fit outside the prediction range of the model to see what issues exist. Combining visualizations allows us to confirm a model's abilities in the real world and to explain the outcomes to other medical experts.

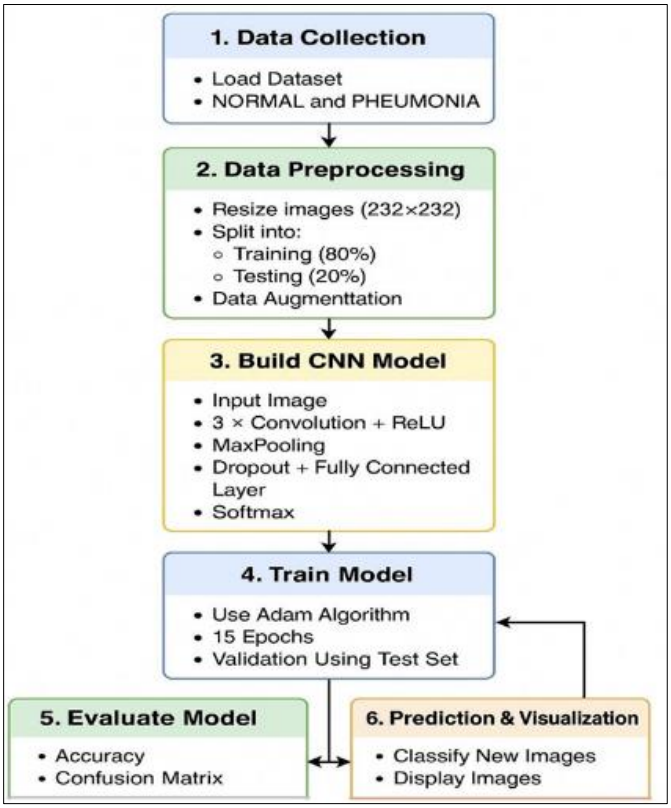


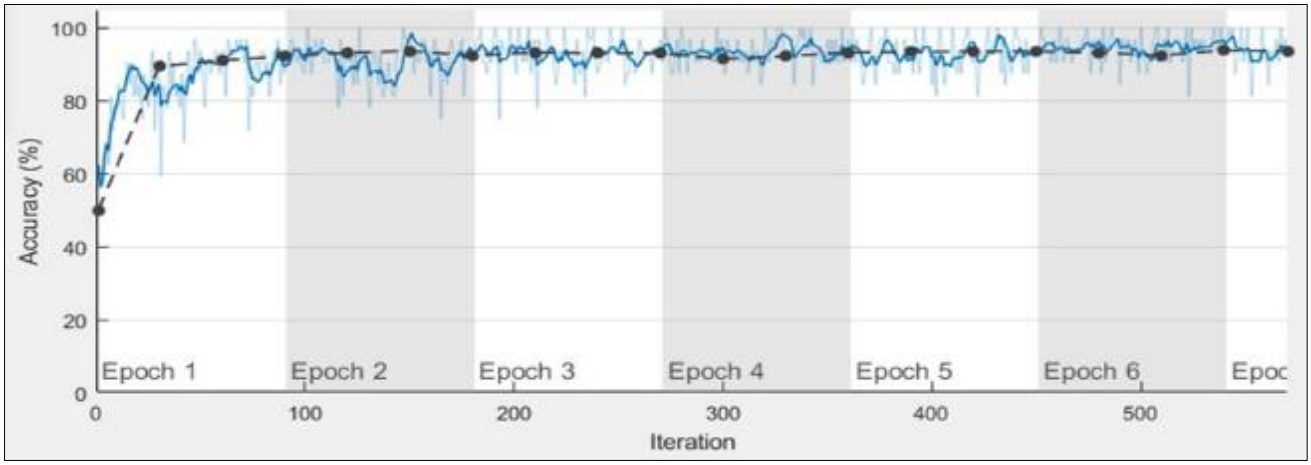
Figure 3 procedures of the proposed model

4. Results and Discussion

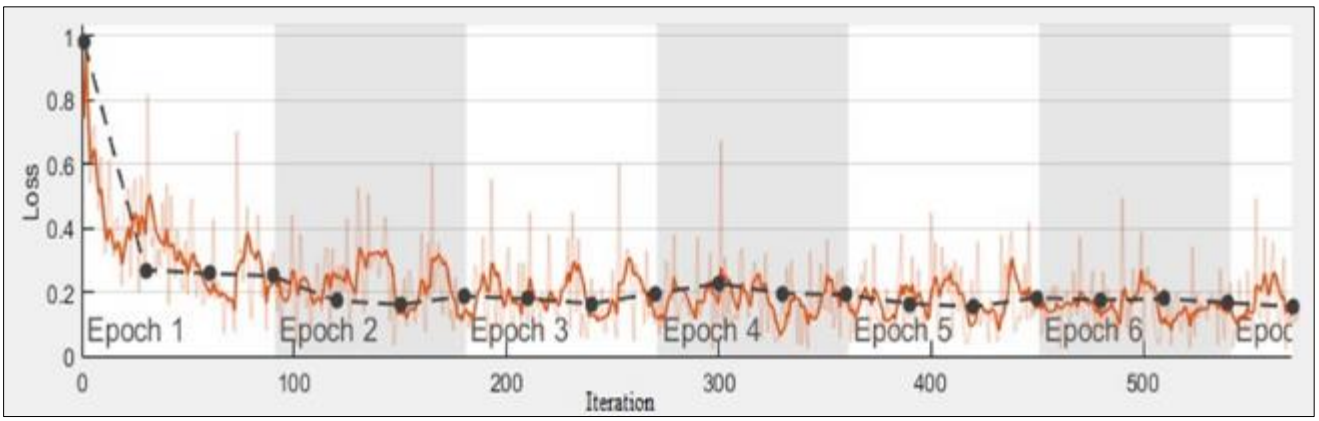
In this section, the complete results of the model training and evaluation are presented. Particular attention is paid to the progression of accuracy and loss over time, as well as final performance indicators such as validation accuracy, and a comparative study with earlier studies.

The findings of the CNN model at the beginning stage (Epoch 7) and again during the mid-to-late phases (Epoch 10) are depicted in Figures 4 and 5, respectively. In the beginning stages, the accuracy of both training and validation continues to improve, eventually reaching about 85% and approximately 90% by the time the 580th iteration has been completed. The losses that are being exhibited by training and validation are falling at a rapid rate, which indicates that the model is learning quickly and performing

effectively. With regard to Epoch 10 and the 930th iteration, the training accuracy reached a plateau of approximately 98%, while the validation accuracy remained in the vicinity of 94%. This demonstrates that the model is learning in a quick and effective manner. The fact that the model is converged is demonstrated by the fact that both the training and validation losses are very near to zero. It depicts how the architecture style and the manner in which it is educated, in conjunction with regularization, ultimately led to the production of this result. In addition, the fact that the validation results are comparable indicates that the model is not overfitting, which means that it performs well even when applied to data that has not been seen before. The model is improving steadily and functioning consistently, despite the fact that it takes so much time because it only uses a single CPU.

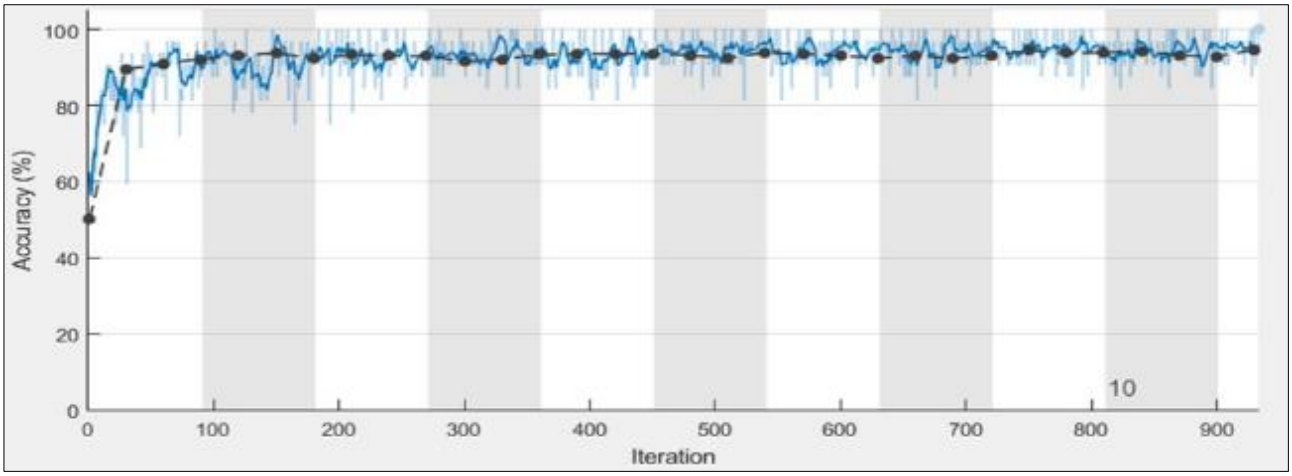


(a)

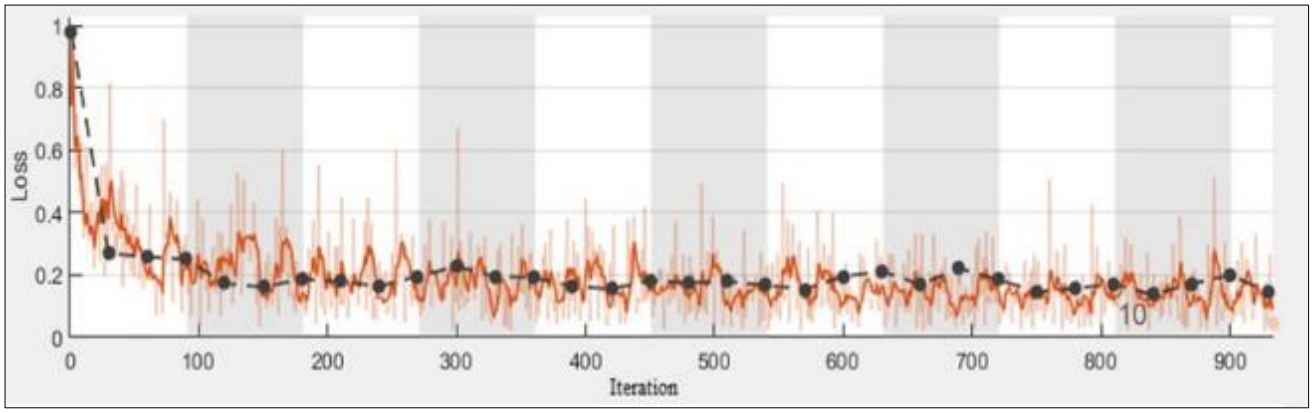


(b)

Figure 4 The accuracy % for training, testing, and validation samples for 580 of 1350 iteration and Epoch 7



(a)



(b)

Figure 5 The accuracy % for training, testing, and validation samples for 930 of 1350 iteration

As can be observed from the training progress plot (Figure 6), the accuracy of training as well as the accuracy of validation both rise at a quick rate at the beginning of the process. The precision continues to improve until around 300 iterations have passed, at which point it reaches a plateau, culminating in a validation accuracy of 94.03%. Due to the fact that it is stable, the model appears to have learned the fundamental structure of the data without relying excessively on specific samples (See Table 6).

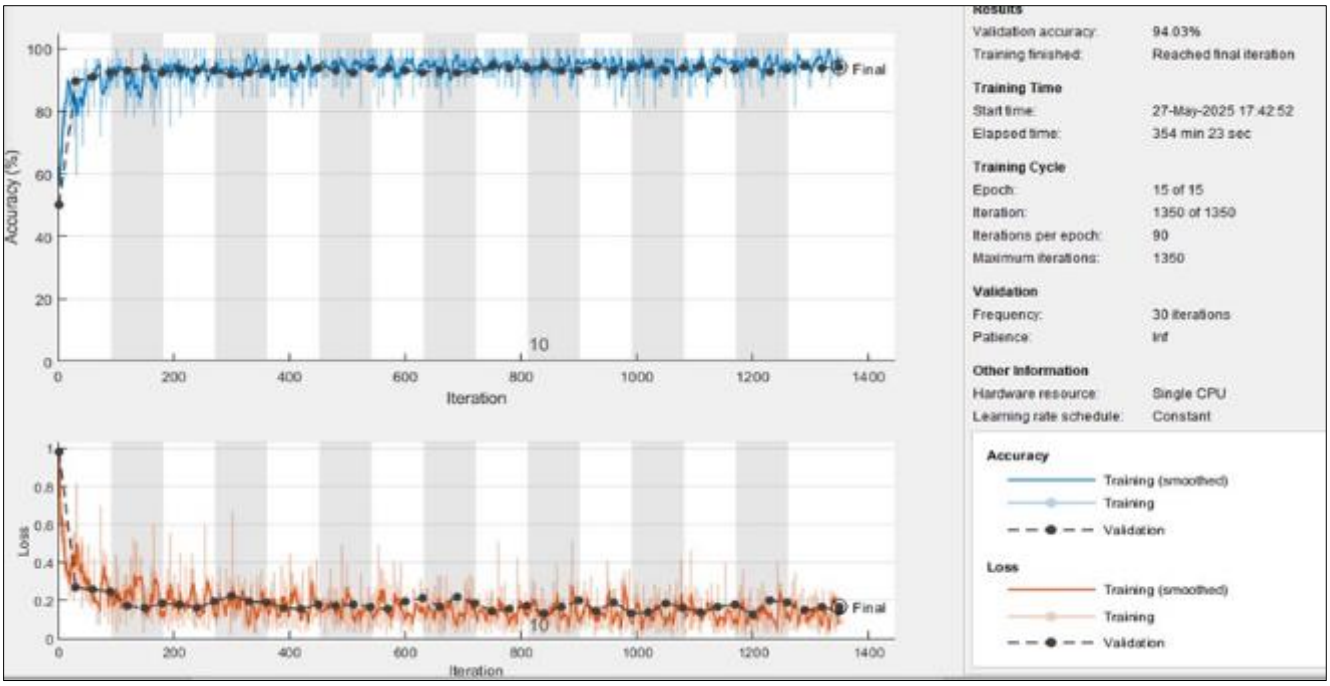


Figure 6 The accuracy % for training, testing, and validation samples for 1350 of 1350 iteration

Table 6 Data Preprocessing of the proposed model

Metric	Value
Final Validation Accuracy	94.03%
Epochs	15
Total Iterations	1350
Training Duration	354 minutes
Learning Rate Schedule	Constant
Hardware	Single CPU

Both the training loss and the validation loss begin at a high level, but they quickly decrease and eventually reach a value that is consistent and low. The fact that the loss curves are exhibiting relatively little shifts is consistent with the fact that the model is capable of learning how to generalize without requiring significant adjustments.

According to the ROC curve, the model classification procedure is of a high grade, as demonstrated in Figure 7. The top-left corner of the curve is where a true positive is most likely to occur, which means that there are just a few false positives. Since the area under the curve (AUC) is close to 1, the model successfully differentiates between NORMAL and PNEUMONIA as presented in Table 7.

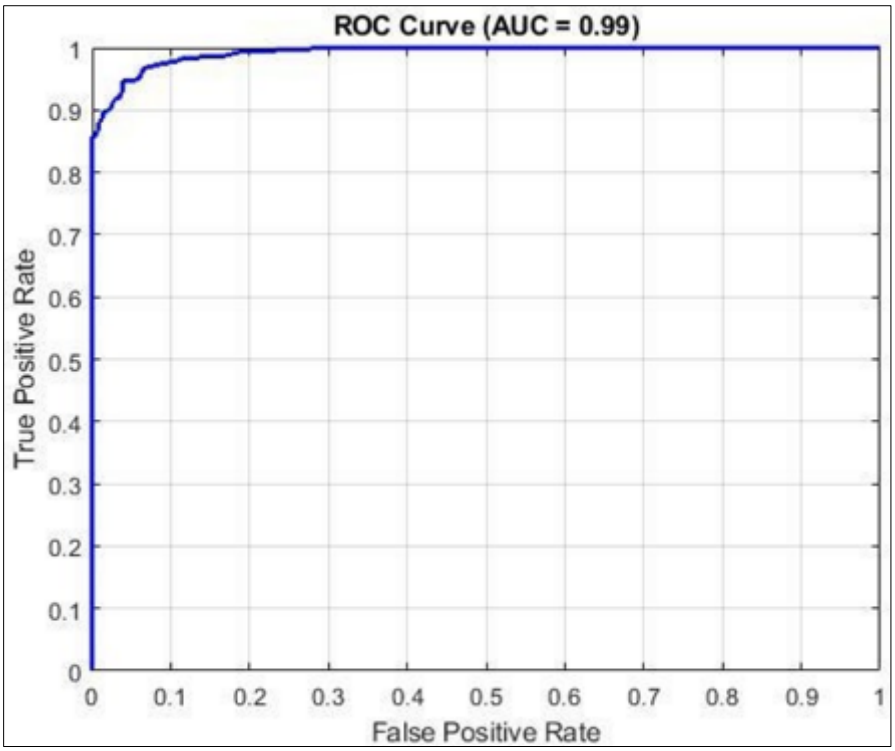


Figure 7 The ROS result of this study

Table 7 ROC analysis for the perusal

Metric	Value
AUC (Area Under Curve)	0.99
Classification Capability	Excellent
ROC Curve Shape	Steep, top-left

The outcomes of the evaluation metrics provide insight into the degree to which the model is able to adapt to circumstances. accurate results for PNEUMONIA. More than 98% of instances that are considered to be NORMAL are correctly identified by the model, while it is able to correctly identify over 90% of patients who are diagnosed with PNEUMONIA. By the time the F1-score reaches 0.94, both precision and recall are almost identical to one another (Figure 8) . In Figure 8, you can clearly see how many samples from each group were classified correctly and incorrectly by the model, as NORMAL or PNEUMONIA. Both healthy and diseased cases were well recognized by the model, as it achieved a high accuracy rate of 354 for healthy cases and 323 for sick cases. Still, 6 NORMAL cases were predicted as PNEUMONIA, when they actually didn't have pneumonia. Such errors don't need to be a concern in medical situations, since they can encourage more tests, not missed care. It is especially serious that of the 37 cases, 37 were falsely classified as normal, but they actually had PNEUMONIA. Errors of this type often result in misdiagnosis and delays in providing treatment. Even so, the results indicate that the model stays reliable for diagnosis, mainly because it doesn't mislabel examples often. Ensuring fewer false negatives should be a key part of improving future clinical devices.

In terms of classifying chest X-ray images, our model does an outstanding job of determining whether they are NORMAL or PNEUMONIA (See Figure 9). It is possible to accurately identify individuals who do not have pneumonia thanks to the technique's high area under the curve (AUC) and specificity. Additionally, due to the technique's high sensitivity, it is also able to consistently diagnose individuals who do have pneumonia. The accuracy of the model is quite high, coming in at 94.03%. It shows outstanding accuracy in labeling NORMAL cases correctly and requires little false detection. There is good detection of PNEUMONIA cases with a recall of 0.90, though a small number were overlooked. The strong F1 score which is 0.94, demonstrates that the model achieves equal accuracy in medical diagnosis (as presented in Table 8).

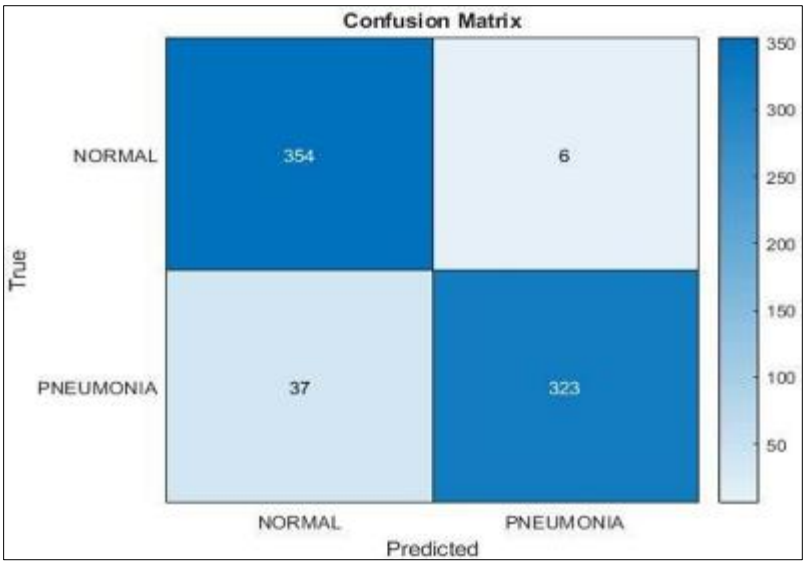


Figure 8 The Confusion Matrix result of this study

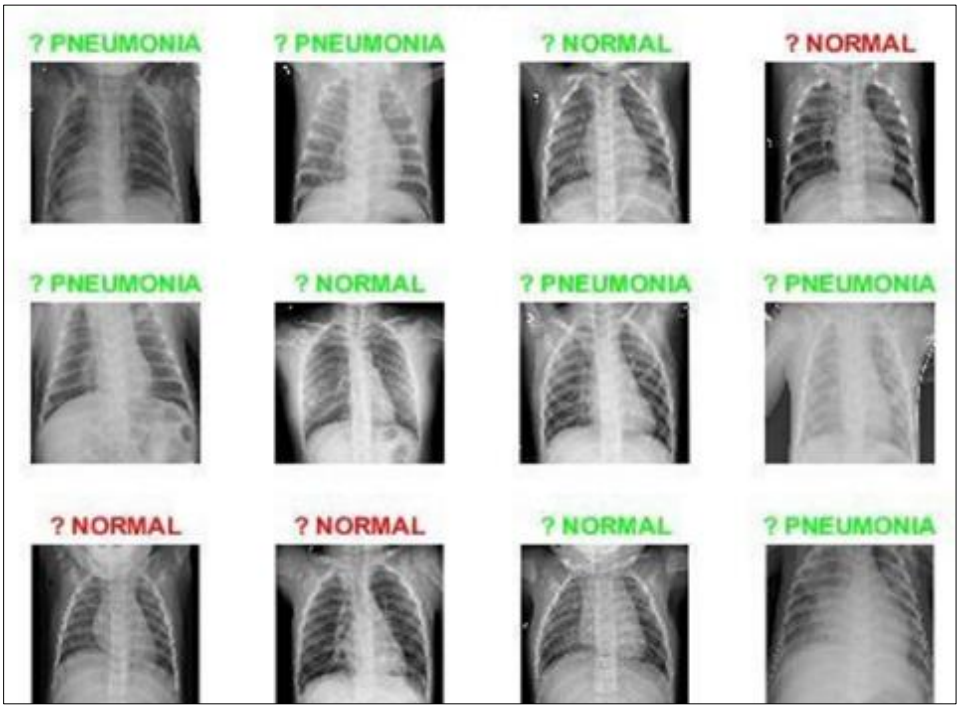


Figure 9 prediction results, green color (correct prediction), red color (false prediction)

Table 8 Classification Result of this study

Metric	Value
Accuracy	94.03%
Sensitivity (Recall)	0.90
Specificity	0.98
Precision	0.98
F1 Score	0.94

Despite this, the fact that false negatives (37) are mentioned indicates that there is a possibility that real cases of PNEUMONIA could be missed, which could have major implications for the patient's health.

In contrast to other studies, the existing model performs well in many evaluation areas. While the accuracy in Goel et al. (2025) and Rana et al. (2023) was 91–92%, the proposed model reached 94.03%. Additionally, it attained a precision of 0.98, greatly minimizing false positives and an F1 score of 0.94, suggesting great harmony between its accuracy and its recall. While Rahman (2020) got similar accuracy with transfer learning, our model surpasses it in precision and specificity and keeps recall strong.

Consequently, this shows that the method proposed not only delivers high accuracy but also stronger diagnostic reliability which is better for medical use as presented in Table 9.

Table 9 Comparison of this study with previous studies

Study	Year	Methodology	Accuracy	Precision	Recall (Sensitivity)	F1 Score	Notable Features
[15]	2025	Custom CNN + Cognitive Feature Extraction	91.2%	0.89	0.87	0.88	Cognitive features improved reasoning, but moderate sensitivity.
[11]	2020	Transfer Learning (DenseNet- 121)	93.1%	0.94	0.89	0.91	Leveraged pre- trained model; good overall but slightly lower precision.
[17]	2023	Hybrid CNN + SVM	92.6%	0.92	0.90	0.91	Combined classical ML with DL; computationally more complex.
Current Study	2025	Custom CNN, optimized training	94.03%	0.98	0.90	0.94	High precision and specificity; excellent balance; low false positives.

5. Conclusion

This research introduces a strong CNN that accurately identifies chest X-ray images as having NORMAL or PNEUMONIA problems. The model performed extremely well, hitting an accuracy rate of 94.03% and a recall, precision and F1-score of 0.90, 0.98 and 0.94, respectively and outperforming several new top approaches. Additional validation by the confusion matrix and ROC analysis indicated it was highly efficient and could be used in clinical practice. While lots of similar studies depend more on ready-made or mixed models, the new research finds that a custom-built CNN performs well for pneumonia detection. It demonstrated strong ability to generalize by not overfitting too much which you can see from both the training and validation metrics. Although the results are promising, the study admits that there are false negatives in medical diagnosis. For future work, techniques such as ensemble learning, using different data augmentation techniques and cost-sensitive training should be studied to make the model less likely to miss anything during testing and help it work more

reliably. All in all, this study introduces a powerful, streamlined and relevant deep learning model well-suited for practical medical imaging, mainly helping with early and accurate diagnosis of pneumonia from chest X-rays.

Compliance with ethical standards

Acknowledgments

The Author declares that there is no conflict of interest.

Statement of ethical approval

The present study followed national, and institutional guidelines for humane animal treatment and complied with relevant legislation from institute ethical guideline

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] Alalwan and et al, "A Hybrid Classification and Identification of Pneumonia Using African Buffalo Optimization and CNN from Chest X-Ray Images," CMES-Computer Modeling in Engineering, vol. 138, no. 3, 2024.

[2] Alsharif and et al, "PneumoniaNet: Automated detection and classification of pediatric pneumonia using chest X-ray images and CNN approach," Electronics, vol. 10, no. 23, p. 2949, 2021.

[3] Bramesh and et al, "Comparative Study of Machine Learning Algorithms on Census Income Data Set," Journal of Engineering Research and Application, pp. 78-81, 2019.

[4] Chakrabarty and et al, "A statistical approach to adult census income level prediction," International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pp. 207-212, 2018.

[5] Jason Brownlee, "Imbalanced Classification with the Adult Income Dataset," Imbalanced Classification, 2020.

[6] Joloudari and et al, "Effective class-imbalance learning based on SMOTE and convolutional neural networks," Applied Sciences, vol. 13, no. 6, p. 4006, 2023.

[7] Kozhay and et al, "Convolutional neural network-based defect detection technique in FDM technology," Procedia Computer Science, vol. 231, pp. 119-128, 2024.

[8] Mateo and et al, "Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis," Biocybernetics and Biomedical Engineering, vol. 41, no. 2, pp. 792-801, 2021.

[9] Li and et al, "Medical image analysis using deep learning algorithms, Frontiers in public health," p. 1273253, 2023.

[10] Hasan and et al, "Recent advancement of deep learning techniques for pneumonia prediction from chest X-ray image," Medical Reports, p. 100106, 2024.

[11] Rehman and et al, "Review on chest pathogies detection systems using deep learning techniques," Artificial Intelligence Review, vol. 56, no. 11, pp. 12607-12653, 2023.

[12] Suganyadevi and et al, "A review on deep learning in medical image analysis, International Journal of Multimedia Information Retrieval," vol. 11, no. 1, pp. 19-38, 2022.

[13] Khan and et al, "Intelligent pneumonia identification from chest x-rays: A systematic literature review," IEEE Access, vol. 9, pp. 51747-51771, 2021.

[14] Chen and et al, "Machine learning techniques for CT imaging diagnosis of novel coronavirus pneumonia: a review," Neural Computing and Applications, pp. 181-199., 2024.

[15] Goel and et al, "Deep Cognitive Learning for Enhanced Pneumonia Detection: Employing CNNs," Innovative Computing and Communications: Proceedings of ICICC, vol. 3, p. 213, 2025.

[16] Rahman, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," Applied Sciences, vol. 10, no. 9, p. 3233, 2020.

[17] Rana and et al, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection, Multimedia Tools and Applications," pp. 26731-26769, 2023.

[18] Lim and et al, "Pneumonia—overview," Encyclopedia of Respiratory Medicine, p. 185, 2021.

[19] Hussin Adam Khatir and et al, "Machine learning models and data-balancing techniques for credit scoring: What is the best combination?" Risks, vol. 10, no. 9, p. 169, 2022.

[20] Mateo and et al, "Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis," Biocybernetics and Biomedical Engineering, vol. 41, no. 2, pp. 792-801, 2021.

Appendix

The MATLAB code of two models of this study is as following

```
clc;
clear;
close all;

%% Set the path to your dataset
baseFolder = 'D:\New folder (2)\Marwa All\Research\1\datasets';

% Category names (ensure these folders exist inside baseFolder)
categories = {'NORMAL', 'PNEUMONIA'};

% Check if category folders exist
for i = 1:length(categories)
    categoryFolder = fullfile(baseFolder, categories{i});
    if ~isfolder(categoryFolder)
        error('Folder not found: %s', categoryFolder);
    end
end

%% Set image size
imageSize = [232 232 3];

%% Load images with labels
imds = imageDatastore(fullfile(baseFolder, categories), ...
    'LabelSource', 'foldernames', ...
    'IncludeSubfolders', true);

%% Split dataset into training and testing sets
[imdsTrain, imdsTest] = splitEachLabel(imds, 0.8, 'randomized');

%% Prepare data augmenter for training only
augmenter = imageDataAugmenter( ...
    'RandRotation', [-15 15], ...
    'RandXReflection', true, ...
    'RandXScale', [0.9 1.1], ...
    'RandYScale', [0.9 1.1]);

augTrain = augmentedImageDatastore(imageSize, imdsTrain, 'DataAugmentation', augmenter);
augTest = augmentedImageDatastore(imageSize, imdsTest);

%% Define CNN architecture
layers = [
    imageInputLayer(imageSize)

    convolution2dLayer(3, 32, 'Padding', 'same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2, 'Stride', 2)

    convolution2dLayer(3, 64, 'Padding', 'same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2, 'Stride', 2)

    convolution2dLayer(3, 128, 'Padding', 'same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2, 'Stride', 2)

    dropoutLayer(0.5)

    fullyConnectedLayer(256)
    reluLayer
    dropoutLayer(0.5)

    fullyConnectedLayer(2)  % number of classes
    softmaxLayer
    classificationLayer
];
```

```
%% Set training options
options = trainingOptions('adam', ...
    'InitialLearnRate',1e-4, ...
    'MaxEpochs',15, ...
    'MiniBatchSize',32, ...
    'ValidationData',augTest, ...
    'ValidationFrequency',30, ...
    'Verbose',false, ...
    'Plots','training-progress');

%% Train the network
disp('Training the model...');
trainedNet = trainNetwork(augTrain, layers, options);

%% Evaluate the model
disp('Evaluating the model...');
YPred = classify(trainedNet, augTest);
YTrue = imdsTest.Labels;

accuracy = mean(YPred == YTrue);
disp(['Accuracy: ', num2str(accuracy * 100, '%.2f'), '%']);

confMat = confusionmat(YTrue, YPred);
figure;
heatmap(categories, categories, confMat, ...
    'Title', 'Confusion Matrix', ...
    'XLabel', 'Predicted', 'YLabel', 'True');

% Performance metrics
TP = confMat(2,2); TN = confMat(1,1);
FP = confMat(1,2); FN = confMat(2,1);

sensitivity = TP / (TP + FN);
specificity = TN / (TN + FP);
precision = TP / (TP + FP);
f1_score = 2 * (precision * sensitivity) / (precision + sensitivity);

disp(['Sensitivity: ', num2str(sensitivity, '%.2f')]);
disp(['Specificity: ', num2str(specificity, '%.2f')]);
disp(['Precision:   ', num2str(precision, '%.2f')]);
disp(['F1 Score:    ', num2str(f1_score, '%.2f')]);

%% Display prediction results
figure;
idx = randperm(numel(imdsTest.Files), min(12,numel(imdsTest.Files)));
for i = 1:numel(idx)
    subplot(3,4,i);
    img = readimage(imdsTest, idx(i));
    pred = YPred(idx(i));
    trueLabel = YTrue(idx(i));
    imshow(img);
    if pred == trueLabel
        title(['✓ ', char(pred)], 'Color','g');
    else
        title(['✗ ', char(pred)], 'Color','r');
    end
end
sgtitle('Prediction Results');

%% ROC Curve and AUC
[~, scores] = classify(trainedNet, augTest);
trueBin = double(YTrue == 'PNEUMONIA');  % Assuming 'PNEUMONIA' is the positive class

[fpRate, tpRate, ~, AUC] = perfcurve(trueBin, scores(:,2), 1);

figure;
plot(fpRate, tpRate, 'b', 'LineWidth', 2);
xlabel('False Positive Rate');
ylabel('True Positive Rate');
title(['ROC Curve (AUC = ' num2str(AUC, '%.2f') ')']);
grid on;

%% Save the model and results
save('trainedPneumoniaNet.mat', 'trainedNet');

resultsTable = table(YTrue, YPred, scores(:,1), scores(:,2), ...
    'VariableNames', {'TrueLabel', 'PredictedLabel', 'NormalProb', 'PneumoniaProb'});
writetable(resultsTable, 'prediction_results.xlsx');
```

Figure A 1 MATLAB code of proposed Model