



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

A Multiclass Deep Learning Framework for Lung Diseases Classification with Chest X-ray Images

Indushree Shetty^{1*}, Prerna Agrawal²

¹Research Scholar, FCAIT, GLS University, Ahmedabad, India. E-mail: shettyindu.05@gmail.com

²Assistant Professor, FCAIT-PG, GLS University, Ahmedabad, India. E-mail: Prerna013@gmail.com

Abstract

One of the significant global health issues is lung diseases, and they are one of the top causes of morbidity and mortality around the world. Effective diagnosis is important to optimally treat patients, thus improving their prognosis. Chest X-rays, due to their cost, speed, and accessibility, are commonly used to identify abnormal pulmonary presentations. However, there are many overlapping anatomical structures, subtle signs of disease, and other similarities among different thoracic abnormalities, making manual examination of CXR images a challenge. To handle these challenges, this research proposes an innovative deep learning-based approach using chest X-ray images for multiclass classification of lung diseases. The proposed approach uses MobileNet+Adam optimizer in conjunction with a seven convolutional layer (CNN7) feature extraction block for enhanced learning of disease related features, thereby improving classification accuracy. A balanced dataset of 7,446 chest X-ray images consisting of three classes: pneumonia, normal, and pleural thickening was used to train and evaluate the proposed multiclass classification approach. Additionally, a wise comparison of seven different optimization algorithms (e.g., Adam, SGD, and RMSProp) was performed. The performance of the proposed multiclass classification approach was assessed using F1 score, accuracy, recall, precision, confusion matrix, and AUC. The results showed that the MobileNet+Adam optimizer (with 7 layers of convolution) produced the best overall performance, resulting in respective AUC values of 0.9473 for normal, 0.9996 for pneumonia, and 0.9474 for pleural thickening. Furthermore, comparison with previously published multiclass classification methods confirmed that the proposed deep learning multiclass classification approach outperformed baseline models and previous methods in the literature. The findings indicate that the proposed methodology using a MobileNet+Adam optimizer with 7 convolutional layers provides a robust, accurate, and computationally efficient solution for automated multiclass lung disease classification and has significant potential for incorporation into intelligent computer aided diagnostic systems for clinical decision support.

Keywords: Deep Learning, Chest X-ray, Multiclass Classification, MobileNet, Convolutional Neural Network (CNN), Pneumonia, Pleural Thickening, Computer-Aided Diagnosis, Adam Optimizer

1. Introduction

Many patients suffer from respiratory diseases every year (around 3 million each year), which leads to many complications and ultimately death. Therefore, there is an urgent need for accurate diagnosis and treatment of these diseases so that the patient will not suffer further complications or die because of the disease progression. Lung disease remains a serious global health concern, as it is one of the top causes of disease and death in the world [1], [2]. The most used diagnostic imaging so far is the chest X-rays. The reasons for this include the low cost, the fast access time, the minimal radiation exposure associated with taking a CXR, and the high degree of accessibility within healthcare facilities [3]. However, the interpretation of chest X-ray images is difficult due to overlapping anatomical structures, subtle diseases, variations in quality, and similarities between X-rays of different diseases in the thoracic area [4]. Furthermore, there has been a rapid increase in the number of people receiving radiology examinations, leading to increased workloads for radiologists and greater potential for delays in diagnosis and misdiagnosis due to human error [5]. The development of systems that provide a diagnostic aid using computer aided diagnosis (CAD) tools for health professionals has created much excitement about the accuracy and speed with which health professionals can diagnose pulmonary diseases. The emergence of CAD technologies has been largely impacted by advancements in the field of AI and deep learning for automatic classification of diseases from medical images. Deep Learning based algorithms known as convolutional neural networks (CNNs) are proving to be very effective at performing a multitude of healthcare related tasks such as detecting diseases, segmenting medical images, and classifying medical images [6], [9].

Figure 1 represents the increasing burden of three important respiratory disorders, pneumonia, pulmonary edema, and pleural thickening, in India from 2020 to 2025. The year represents the X-axis, and the relative disease burden index represents the Y-axis. A higher value indicates greater disease burden. This index is used when precise nationwide counts are not available. The graph highlights that the burden

created by respiratory disorders is consistently growing and demonstrates the increasing impact of these pulmonary disorders on global healthcare systems. Pneumonia has the highest relative disease burden of the previously mentioned diseases over the entire time evaluated, with pulmonary edema and pleural thickening trailing in second and third place, respectively. The upward trend confirms the disease burden and the clinical importance of these pulmonary pathologies, requiring early diagnosis and effective management of these conditions. As the number of respiratory diseases continues to increase, healthcare professionals are confronted with more difficulties in appropriately interpreting large volumes of chest X-rays. Therefore, clinicians need intelligent computer aided diagnosis systems that will allow for early classification and detection of lung diseases. Recent research in machine learning and deep learning has shown great promise in streamlining the analysis of chest X-rays, thereby improving the accuracy of diagnosis, decreasing the workload of radiologists, and providing for better and faster clinical decision making.

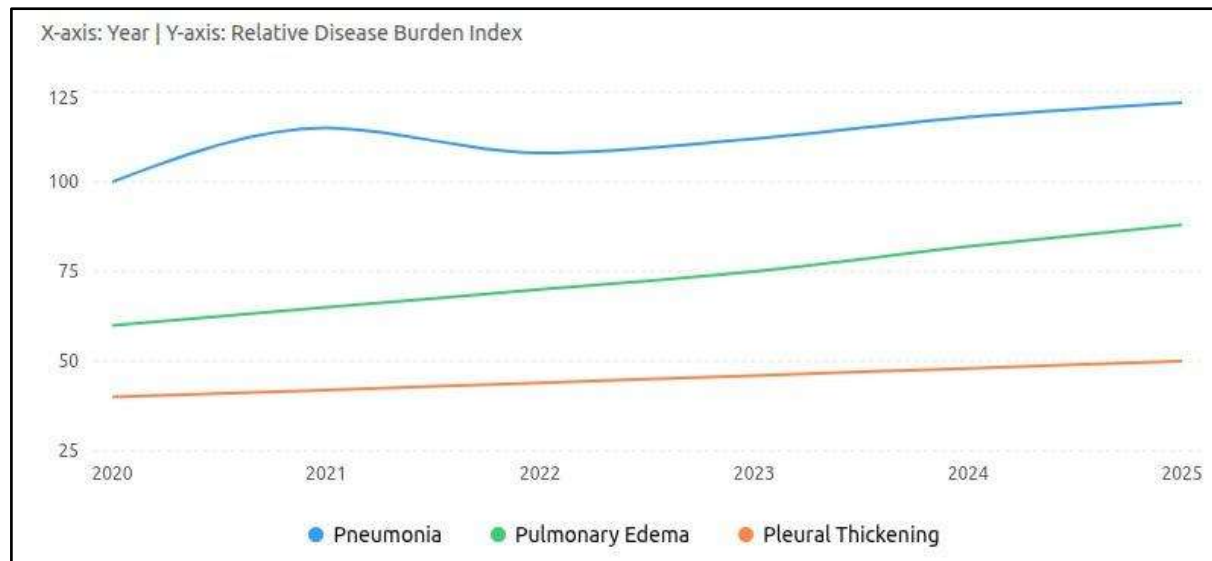


Figure 1: Respiratory Disease Burden Trend in India (2020–2025)

Although most deep learning-based chest X-ray classification studies have reported good classification results, nearly all prior studies have focused on binary classification tasks such as identifying normal and abnormal images or identifying a single type of disease [10-13]. In real world clinical settings, however, multiclass classification systems will be necessary to identify multiple pulmonary diseases simultaneously. Multiclass classification is particularly problematic because most pulmonary diseases share similar X-ray findings, creating confusion between classes and reducing the accuracy of multiclass classification [14]. In addition, most multiclass architectures rely only on transfer learning, rather than developing new methods for extracting relevant features, and there have been few studies investigating the effects of optimization on multiclass classifications of lung diseases [15], [16]. To fill these research gaps, this paper introduces the deep learning framework for performing multiclass classification with chest X-ray images. This framework combines MobileNet with a seven layer convolutional feature extraction block, which enhances the quality of extracted features (disease specific) and reduces the level of confusion associated with classifying X-ray images of diseased patients. The proposed framework will classify the X-ray images of patients as having either normal, pneumonia, or pleural thickening. The experimentation includes an extensive evaluation of 7 optimization algorithms (Adam, RMSProp, SGD, Adagrad, Nadam, Adadelta, and Adamax) to measure the effects of these algorithms on classification results. This study contributes a new method, MobileNet+Adam optimizer using 7 convolutional layers for multiclass classification, providing an extensive evaluation of optimization methods. This paper incorporates various research contributions:

- Classification of normal, pneumonia, and pleural thickening chest X-ray images.
- Utilizing a dataset of 7,446 chest X-ray images that is balanced.
- Enhanced feature extraction using a seven-layer CNN block.
- Evaluation of seven optimization methods for deep learning algorithms.
- Assessed performance using recall, accuracy, F1-score precision, AUC, confusion matrix, and ROC.
- Experimentation results showed that the MobileNet+Adam optimizer with the 7 convolutional layers is best for classifying multiclass lung disease.
- Compared results of MobileNet+Adam optimizer with the 7 convolutional layers with existing multiclass classification systems.

- Proved that multiclass lung disease classification is strong and reliable.
- Proposed a deep learning framework for multiclass classification of lung diseases using chest X-rays.

This research paper contains six sections. In section 2, a literature review pertaining to previous research on chest X-ray classification based upon deep learning. Section 3 proposes a deep learning framework for multiclass classification. Section 4 describes the dataset used for performing multiclass classification. Section 5 discusses the experimentation methodology. Section 6 discusses the results of the experiments conducted for multiclass classification. Section 7 concludes the paper with a conclusion and future research.

2. State Of The Art Review

In upcoming years, deep learning has made tremendous progress in the autonomous analysis of chest X-ray images for the diagnosis and classification of lung diseases. There is much research that has studied multiclass thoracic disease identification based on deep learning frameworks. Table 1 provides a comparative analysis of current deep learning methodologies for the classification of diseases in chest X-rays. Irtaza et al. [3] introduced a multi-label deep learning system utilizing the ChestX-ray14 and CheXpert datasets, attaining AUC values of 0.69 for pneumonia and 0.87 for pleural thickening. Ho and Gwak [4] developed a complete feature integration technique to classify thoracic illnesses, producing an AUC of 0.74 for pneumonia and an AUC of 0.85 for pleural thickening. Gong et al. [5] also produced Deformable Gabor Feature Networks (DGFN) with an AUC of 0.77 for pneumonia and an AUC of 0.81 for pleural thickening. Ho and Gwak [6] also performed an extensive review of the methodologies used for knowledge distillation and produced an AUC of 0.76 for pneumonia and an AUC of 0.78 for pleural thickening. Baltruschat et al. [7] compared several deep learning architectures and produced a pneumonia AUC of 0.69 and a pleural thickening AUC of 0.73. In comparison, Gündel et al. [8] produced location aware dense networks and reported an AUC of 0.76 for pneumonia and an AUC of 0.94 for pleural thickening. Shamrat et al. [9] constructed a modified MobileNetV2 architecture to classify multi-class lung illnesses, having produced a classification accuracy of 96.97%. Sharma and Guleria [10] implemented a VGG16 based framework for pneumonia diagnosis and produced a combined accuracy, precision, recall, and F1 score of 95.4%. CNN based classifiers were used by Pranaya et al. [11] and Racic et al. [12] to classify pneumonia, generating accuracies of 94.00% and 93.73%, respectively. Ali et al. [13] introduced a framework utilizing EfficientNetV2L and achieved 94.02% accuracy for pneumonia identification. Goyal and Singh [14] identified machine learning and deep learning frameworks for classifying respiratory related diseases and produced 96.20% accuracy classification. Hole et al. [15] implemented a deep learning framework to classify pneumonia, achieving training accuracy of 83.32% and testing accuracy of 74.20%. Usman et al. [16] evaluated ML methodology for pneumonia classification and produced 79% accuracy, 76% precision, 73% recall, and 74% F1 score. Agrawal and Gandhi [17] most recently examined optimizer specific methods for pneumonia classification using DenseNet121, ResNet50, and CNN architectures, achieving an AUC of 0.9868. Likewise, Shetty et al. [19] performed an optimizer specific assessment for pulmonary edema detection, obtaining an AUC of 0.9999.

Previous research has proven encouraging results in analyzing chest X-rays; however, there are still numerous limitations. There is a lack of research on the impact of various optimization techniques on the efficacy of multi class classification. Furthermore, the radiographic characteristics of thoracic abnormalities such as pleural thickening and pneumonia are frequently comparable, which leads to inter class confusion and a decrease in classification accuracy. To overcome these limitations, we propose an improved deep learning framework by incorporating MobileNet with seven layer convolutional feature extraction using the Adam optimizer and conducting a comprehensive optimizer wise evaluation for multiclass classification of normal, pneumonia, and pleural thickening chest X-ray images. In the next section the proposed deep learning framework for multiclass classification of lung diseases is discussed.

Paper	Dataset Used	Methodology	Application	Results	Limitation
Irtaza et al. [3]	ChestX-ray14 & CheXpert datasets	Multi-label Deep Learning Framework	Multi-label Lung Disease Classification	Pneumonia AUC: 0.69 Pleural Thickening AUC: 0.87	No optimizer-wise evaluation.
Ho and Gwak [4]	ChestX-ray14	Multiple Feature Integration	Thoracic Disease Classification	Pneumonia AUC: 0.74 Pleural Thickening AUC: 0.85	Limited feature enhancement mechanism.

Gong et al. [5]	INbreast dataset and ChestX-ray14 dataset.	Deformable Gabor Feature Networks (DGFN)	Biomedical Image Classification	Pneumonia AUC: 0.77 Pleural Thickening AUC: 0.81	Not focused on multiclass chest X-ray classification.
Ho and Gwak [6]	ChestX-ray14	Knowledge Distillation-Based Deep Learning	Chest X-ray Abnormality Classification	Pneumonia AUC: 0.76 Pleural Thickening AUC: 0.78	Focused on knowledge transfer rather than feature enhancement.
Baltruschat et al. [7]	ChestX-ray14	Comparative Deep Learning Models	Multi-label Chest X-ray Classification	Pneumonia AUC: 0.69 Pleural Thickening AUC: 0.73	Performance limitations for certain thoracic diseases.
Gundel et al. [8]	ChestX-ray14	Location-Aware Dense Networks	Thoracic Abnormality Detection	Pneumonia AUC: 0.76 Pleural Thickening AUC: 0.94	No optimizer analysis.
Shamrat et al. [9]	ChestX-ray14	Customized MobileNetV2	Multiclass Lung Disease Classification	Achieved 96.97% classification Accuracy , AUC not defined	Relied mainly on transfer learning.
Sharma and Guleria [10]	ChestX-ray14	VGG16 + Fully Connected Neural Network	Pneumonia Detection	Achieved 95.4% accuracy, 95.4% precision, recall, and F1-score; AUC not defined	Binary classification only.
Pranaya et al. [11]	Kaggle Dataset	Convolutional Neural Network (CNN)	Pneumonia Detection	Achieved 94.00% Accuracy, AUC not defined	No multiclass evaluation.
Racic et al. [12]	Heath Care Center	CNN-Based Architecture	Automated Pneumonia Diagnosis	Achieved 93.73% Accuracy, AUC not defined	Limited to pneumonia classification.
Ali et al. [13]	ChestX-ray14	EfficientNetV2L	Pneumonia Detection	Achieved 94.02% Accuracy, AUC not defined	No multiclass disease categorization.
Goyal and Singh [14]	ChestX-ray14	Machine Learning and Deep Learning Models	Respiratory Disease Classification	Achieved 96.20% Accuracy, AUC not defined	Disease-specific classification.
Hole et al. [15]	Chest X-ray dataset, RSNA Pneumonia Detection Dataset.	Deep Learning Framework	Pneumonia Detection	Achieved 83.32% training Accuracy and 74.20% testing accuracy; AUC not defined	Limited to pneumonia diagnosis.
Usman et al. [16]	RSNA Pneumonia Detection Challenge Dataset.	Machine Learning Algorithms	Pneumonia Classification	Achieved 79% Accuracy, 76% Precision, 73% recall, and 74% F1-score; AUC not defined	Did not address multiclass thoracic disease classification.

Agrawal & Gandhi [17]	ChestX-ray14	DenseNet121, ResNet50, CNN with Optimizer Analysis	Pneumonia Detection	AUC Pneumonia: 0.9868	for Binary classification only; no multi-class thoracic disease classification.
Shetty et al. [19]	ChestX-ray14	Deep Learning Models with Optimizer-wise Evaluation (Adam, SGD, RMSProp, Adagrad, Adadelata, Adamax, and Nadam)	Pulmonary Edema	AUC for Edema: 0.9999	Focused only on binary classification (pulmonary edema vs. normal) and did not investigate multiclass thoracic disease classification.

Table 1: Comparison of Existing Studies on Chest X-ray Classification [3-17, 19]

3. Proposed Deep Learning Framework For Multiclass Classification

The entire proposed multiclass classification process of lung diseases through chest X-ray imaging is shown in Figure 2. The proposed deep learning architecture for multiclass lung disease classification is structured into three main modules as depicted in the figure. They are 1) binary classification, 2) multiclass classification, and 3) XRAYASSIST development of an intelligent chest X-ray analysis system. In the first module, binary classification tests are performed using a combination of different deep learning and hybrid architectures combined with a seven layer convolutional feature extraction block and with seven optimizers named Adam, Adagrad, SGD, RMSProp, Adadelata, Nadam, and Adamax. The binary classification was done between edema vs. normal, pneumonia vs. normal, and pleural thickening vs. normal individually. The experimental results of these binary classification modules are shown for individual studies conducted for the detection of pneumonia [17], [19], pulmonary edema [18], and pleural thickening [20], where the impact of different deep learning architectures and optimization algorithms on classification performance was extensively evaluated using chest X-ray images.

The second module focuses on multiclass classification, wherein appropriate deep learning architectures are shown to simultaneously categorize chest X-ray pictures into three classifications: pneumonia, pleural thickening, and normal. A variety of optimizers are assessed to identify the most efficient setup for multiclass disease classification. The incorporation of the CNN7 layered feature extraction block seeks to augment disease specific feature learning, refine feature representation, and minimize inter class confusion among visually analogous thoracic abnormalities.



Figure 2: Proposed Deep Learning Framework for Multiclass Classification of Lung Diseases Using Chest X-ray Images

This last module targets the development of XRAYASSIST, a computerized system to interpret chest x-rays that will transform research findings into a working clinical decision support system. XRAYASSIST is designed as an integrated system that combines the most successful binary classifiers with one of the best multiclass classifiers. The complete design of this system includes components for preprocessing of images, predicting diseases automatically, generating confidence scores, and producing reports. The binary classifiers for detecting pneumonic conditions, pleural thickening, and normality will all be integrated with multiclass classification capabilities to produce a complete and reliable automated chest x-ray interpretation system that provides a robust platform for diagnosing lung diseases. In the next section the dataset used for the multiclass classification experimentation is discussed.

4. DATASET USED

The images of chest X-ray shown in Figure 3 were taken from three classes utilized in this classification study: the three classes were (a) Normal, (b) Pneumonia, and (c) Pleural Thickening to visually illustrate and identify categorical variations among the classes being classified as multiclass classification. Table 2 explains the research data using chest X-ray images containing a total number of 7,446 X-ray images collected from publicly accessible datasets. These chest X-ray images are divided into three different categories: pneumonia, pleural thickening, and normal. Each of these categories contains 2,484 images for use in research and will therefore create a balanced sample size within each category. In addition to splitting the images into three balanced samples, each sample has also been divided into test and training datasets.

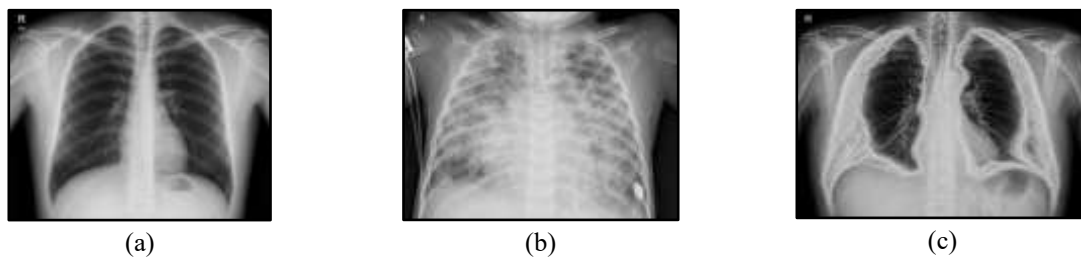


Figure 3: Representative chest X-ray samples of (a) normal, (b) pneumonia, and (c) pleural thickening

Class	Training Images	Testing Images	Total Images
Normal	2182	302	2484
Pneumonia	2182	302	2484
Pleural Thickening	2182	302	2484
Total	6546	906	7446

Table 2: Multiclass Classification Dataset [21-28]

The training images are further divided into training and validation in a ratio of 80:20, 80% for training and 20% for validation. The number of images designated for training in each category was 2,182, while the number of images set aside for testing in each category was 302. Therefore, the training data contained a total number of 6,546 images, and the test data contained 906 images. In order to perform the multi class classification and measure the ability of the model to differentiate between chest X-ray images of normal, pneumonia, and pleural thickening, the entire dataset was used. In the next section, the experimentation methodology for carrying out multiclass classification experiments is discussed.

5. Experimentation Methodology

5.1 Experimental Setup

In this section the setup of an experiment to develop a multiclass classification system used to assess the performance of the proposed system is described. All tests were performed on an AMD Ryzen™ AI 5 340 computer (64 bit) using an AMD Radeon™ 840M (2 GHz), running under the Microsoft 64 bit Windows operating system. The software used for data preprocessing, model building, training, and testing included TensorFlow and Keras, while scientific computing and visualization libraries such as Pandas, NumPy, Matplotlib, and Seaborn were used for analysis of data, evaluation of the model, and visualization of results.

5.2 Data Preprocessing

The X-ray images of the chest are taken from multiple datasets that are publicly available and have different image sizes, resolutions, and pixel intensity distributions before model training; therefore, various preprocessing activities were created to ensure that all chest X-ray images meet specific guidelines to allow for consistent model performance, which will subsequently lead to improved chest X-ray image classification performance. Figure 4 shows different data preprocessing steps carried on chest X-rays. The first of the preprocessing steps applied to all X-ray images was normalization. A normalization factor of $1/255$ was adopted for each of the chest X-ray images to convert each pixel in an X-ray image from a value scale of 0-255 to a value scale of 0-1. All images of chest X-rays were then resized into an image dimension of 150×150 pixels in size in order to facilitate using the same image input sizes to improve the algorithm performance. The use of normalization of images is a technique that will improve numerical stability by supporting the rapid combination of the model during the training phase of the learning algorithm. After all of the preprocessing steps of normalization and resizing was completed; the dataset was then divided into separate testing and training sets. The use of image generators was also used to assist in efficiently loading batches of images during the training of the model.

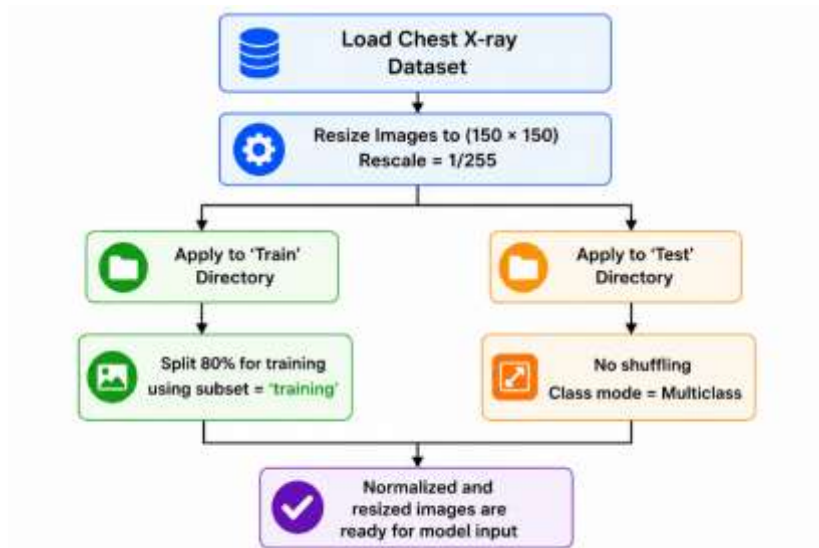


Figure 4: Chest X-ray preprocessing flowchart diagram [17-18]

5.3 Hyperparameters Used

Hyperparameter	Value
Rescale	$1/255$ (Normalizes pixel values from 0–255 to 0–1)
Target Size	150×150 pixels
Batch Size	32
Class Mode	Multiclass (Normal, Pneumonia, Pleural Thickening)
Optimizers Used	SGD, Adam, Adagrad, RMSProp, Adadelata, Nadam, Adamax
Output Layer Function	Softmax
Number of Convolutional Layers	7
Epochs	5

Table 3: Hyperparameters Configuration [17-18]

While fitting the model, some of the following hyperparameters have been chosen to determine the capability of the proposed multiclass classification framework. Table 3 explains all the hyperparameters used for the multiclass experiments. A rescaling factor of $1/255$ was used to rescale all of the chest X-ray images to 0-1, which is the one responsible for translating pixel intensity numbers from the range of 0-255 to just 0-1. All of the images were then resized in height and width to 150 pixels so that all the inputs had the same dimensions and reduced the complexity of the algorithm. The batch size of 32 is a result of effective memory usage and reliable gradient updates. Normal, pneumonia, and pleural thickening are the three categories for which the proposed framework was constructed to classify chest X-ray images. To study the effects of various optimization methods on classification performance, seven different algorithms were used. These are Adadelata, SGD, Adam, Nadam, Adagrad, RMSProp, and Adamax. For the purpose of feature extraction, a total of seven convolutional layers are included in the network. For building probability distributions to all three of the target classes, the output layer used is softmax, the activation function. The model was trained using 5 epochs, which were then evaluated using a variety of classification measures.

5.4 Proposed Methodology

The first step requires loading the chest X-ray images and preprocessing them by normalizing and scaling them to the same dimensions of 150×150 px. The CNN7 Block will consist of seven layers of convolution to create a unique method for extracting identifying behaviours of each disease from the non-identical images present. The unique optical aspects contain disease identifying behaviors that assist with modeling better predictions. Optimizers will be used for the duration of the training of the model to determine the best performing optimization method (Adam, SGD, RMSProp, Adagrad, Adadelata, Adamax, and Nadam). These optimizers will be used individually during the evaluation of specialist classifiers to measure their classification success. Global average pooling of feature maps and then the use of common dense and dropout layers will also reduce the chance of overfitting and also improve the model's generalized performance. The model will train on categorical cross entropy losses with the use of model checkpointing and early stopping methods to form the best performing of all the classifiers in measure of their measurement. The model will be employed to determine class type after the training process is completed, with the use of assessment being the ground truth. The predicted probabilities will then be converted to the actual class types, and to understand and compare each optimizer to the complete, use a confusion matrix and utilize the performance metrics of their performance on their confusion score metric (TP, TN, FP, and FN). Finally, the results from the performance of all optimizers will be recorded in a CSV or spreadsheet file, and the optimizers will be compared to determine which optimizer's performance is the best for the purpose of creating a multi class lung disease classifier from image data of chest X-rays. In the next section the results of the experiments carried out for multiclass classification are discussed.

Proposed Algorithm for Multiclass Classification

STEP 1: Begin

STEP 2: Load training and testing datasets from the specified directories.

STEP 3: Normalize pixel values using rescaling ($1/255$) and then resize all input images to 150×150 pixels.

STEP 4: Initialize image data generators for training and testing datasets.

STEP 5: Extract class labels and determine the total number of classes.

STEP 6: Define the set of optimizers including Adam, SGD, RMSProp, Adagrad, Adadelata, Adamax, and Nadam.

STEP 7: For each optimizer, initialize the deep learning model architecture and configure the feature extraction layers.

STEP 8: Add the proposed CNN7 feature extraction block consisting of convolutional layers, batch normalization, ReLU activation, and max pooling operations with progressively decreasing filter sizes.

STEP 9: Apply Global Average Pooling followed by a dense layer and dropout for regularization.

STEP 10: Add a final Softmax layer to perform multiclass classification.

STEP 11: Compile the model using categorical cross-entropy loss and the selected optimizer.

STEP 12: Train the model using the training dataset while validating on the testing dataset.

STEP 13: Apply model checkpointing to save the best-performing model and use early stopping to prevent overfitting.

STEP 14: Perform predictions on the testing dataset and obtain predicted probabilities.

STEP 15: Convert predicted probabilities into final class labels.

STEP 16: Generate a confusion matrix for each optimizer and save it as an image.

STEP 17: Compute ROC curves for each class and calculate the corresponding AUC values, then save the ROC plots.

STEP 18: For each class, compute evaluation metrics including Accuracy, Precision, Recall, F1-Score, AUC, True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) values.

STEP 19: Store all computed results in a structured format.

STEP 20: Save the results into a CSV file.

STEP 21: Repeat the process for all optimizers.

STEP 22: Return the trained models, confusion matrices, ROC curves, and evaluation metrics.

STEP 23: End.

6 Results, Findings, And Discussion

The experiments were carried out using 7 convolutional layers and without them also, so the results of both processes are discussed here. A total number of 21 deep learning models were evaluated, including DenseNet121, DenseNet121+AlexNet, DenseNet121+NASNet, DenseNet121+ResNet101, DenseNet121+VGG16, DenseNet121+VGG19, MobileNet, AlexNet, ResNet101, NASNet, VGG16, VGG19, ANN, DenseNet121+ANN, DenseNet121+EfficientNet, DenseNet169, DenseNet201, EfficientNet, GoogleNet, InceptionNet, and DenseNet121+InceptionNet. Various optimizers were also used in both the processes. They are Adam, SGD, RMSProp, Adagrad, Adadelata, Adamax, and Nadam. The following criteria were employed to assess each model: accuracy, precision, recall, F1-score, area under receiver operating characteristic curve (AUC), TN, TP, FN, and FP.

6.1 Evaluation Parameters

A complete evaluation of the multiclass classification can be obtained by integrating the confusion matrix with accuracy, precision, recall, F1-score, and area under the curve (AUC). Accuracy is an overall measure of how correct the classifications are, while precision and recall are indicators of how well the model is able to reduce the number of erroneous predictions and identify positive disease cases. The F1-score is utilized for an average of precision and recall. The confusion matrix is used to classify each of the diseases separately based on the performance of the classifier. In addition to this, the AUC score measures the ability of the model to distinguish between the classes separately. Together these evaluation metrics provide a comprehensive and accurate overall assessment of the multi class classification experiments.

6.2 Performance Analysis without CNN7 Convolutional Layers

Initially, experiments were conducted using baseline deep learning architectures without the integration of the CNN7 feature extraction block. Table 4 presents the performance comparison of the top seven optimizers evaluated using deep learning models without the proposed CNN7 feature extraction block. The results indicate that the Adam optimizer achieved the best overall performance among all optimization techniques considered in this study.

Sr. No.	Algorithm Used	Class	Accuracy	Precision	Recall	F1 Score	AUC	TN	TP	FN	FP	Optimizer
1	DenseNet 121 + NASNet	Normal	0.8097	0.5523	0.8046	0.655	0.9042	846	243	59	197	Adam
		Pneumonia	0.9978	0.9974	0.9949	0.9961	1	954	388	2	1	
		Pleural Thickening	0.8171	0.6733	0.5771	0.6215	0.8907	897	202	148	98	
2	DenseNet 121 + NASNet	Normal	0.8112	0.567	0.6722	0.6152	0.8871	888	203	99	155	SGD
		Pneumonia	0.9955	0.9923	0.9923	0.9923	0.9999	952	387	3	3	
		Pleural Thickening	0.7948	0.5974	0.6486	0.6219	0.8678	842	227	123	153	
3	DenseNet 121 + NASNet	Normal	0.8312	0.7095	0.4205	0.5281	0.8986	991	127	175	52	RMSProp
		Pneumonia	0.9963	0.9949	0.9923	0.9936	0.9999	953	387	3	2	
		Pleural Thickening	0.7784	0.5461	0.88	0.674	0.8852	739	308	42	256	
4	DenseNet 121 + NASNet	Normal	0.7502	0.4684	0.8344	0.6	0.8657	757	252	50	286	Adagrad
		Pneumonia	0.9859	0.9649	0.9872	0.9759	0.9992	941	385	5	14	
		Pleural Thickening	0.7814	0.625	0.4	0.4878	0.8507	911	140	210	84	
5	DenseNet +	Normal	0.6364	0.3775	0.9536	0.5408	0.8016	568	288	14	475	Adadelata

	Inception Net	Pneumonia	0.9242	0.8064	0.9718	0.8814	0.9895	864	379	11	91	
		Pleural Thickening	0.7442	0.5517	0.0914	0.1569	0.7749	969	32	318	26	
6	DenseNet 121 + NASNet	Normal	0.8245	0.5948	0.6854	0.6369	0.9003	902	207	95	141	Adamax
		Pneumonia	0.997	0.9974	0.9923	0.9949	1	954	387	3	1	
		Pleural Thickening	0.8074	0.626	0.6457	0.6357	0.8892	860	226	124	135	
7	DenseNet 121 + NASNet	Normal	0.7881	0.5183	0.798	0.6284	0.898	819	241	61	224	Nadam
		Pneumonia	0.9978	1	0.9923	0.9961	1	955	387	3	0	
		Pleural Thickening	0.803	0.6471	0.5343	0.5853	0.8834	893	187	163	102	

Table 4: Top 7 Optimizer Wise Results of Deep Learning Models without CNN7 Convolutional layers

Furthermore, the DenseNet121 + NASNet architecture optimized using Adam demonstrated superior multiclass classification performance compared to the other optimizer configurations, highlighting the importance of optimizer selection in achieving effective chest X-ray disease classification. The multiclass dataset classifies with performance revealing high accuracy, precision, recall, F1-score, and AUC for the model architecture DenseNet121+NASNet with the Adam optimizer. The results show that the baseline architecture can successfully perform multiclass classification and learn the chest X-ray features. However, the class prediction performance was steeply influenced by the type of optimization technique chosen for training, as shown in the performance fluctuations of the various optimizers. Adadelta performed the worst of all optimizers tested, while Adam, SGD, RMSprop, Adamax, and Nadam performed significantly better.

6.3 Performance Analysis with CNN7 Convolutional Layers

The top 7 optimizer wise performance for multiclass classification of chest X-ray images into Normal, Pneumonia, and Pleural Thickening categories is presented in Table 5. The Adam optimizer showed the best balanced and stable performance with high values of accuracy, precision, recall, F1-score, and AUC for all classes among the seven optimizers tested. In particular, the Adam optimizer obtained AUC values of 0.9473, 0.9996, and 0.9474 for the Normal, Pneumonia, and Pleural Thickening categories, respectively, suggesting a superb discriminative capacity. Likewise, SGD, RMSProp, Adagrad, Adamax, and Nadam also achieved good classification performance, especially for the pneumonia class, where consistently high AUC values were reached. Among these optimizers, Adagrad earned the best AUC value of 0.9571 for pleural thickening classification, while Adamax attained the perfect classification performance for the pneumonia class with the AUC of 1.0000.

Sr. No.	Algorithm Used	Class	Accuracy	Precision	Recall	F1 Score	AUC	TN	TP	FN	FP	Optimizers
1	MobileNet	Normal	0.8937	0.7461	0.798	0.7712	0.9473	961	241	61	82	Adam
		Pneumonia	0.9993	1	0.9974	0.9987	0.9996	955	389	1	0	
		Pleural Thickening	0.8818	0.7418	0.8371	0.7866	0.9474	893	293	57	102	
2	MobileNet	Normal	0.8669	0.6836	0.7583	0.719	0.9238	937	229	73	106	SGD
		Pneumonia	0.9978	0.9924	1	0.9962	1	952	390	0	3	
		Pleural Thickening	0.8833	0.7443	0.84	0.7893	0.9463	894	294	56	101	

3	MobileNet	Normal	0.8528	0.6469	0.7583	0.6982	0.9259	918	229	73	125	RMSProp
		Pneumonia	0.9993	1	0.9974	0.9987	1	955	389	1	0	
		Pleural Thickening	0.8491	0.7221	0.6829	0.7019	0.9245	903	239	111	92	
4	MobileNet	Normal	0.8848	0.8419	0.5993	0.7002	0.9412	1009	181	121	34	Adagrad
		Pneumonia	0.9993	1	0.9974	0.9987	1	955	389	1	0	
		Pleural Thickening	0.8595	0.6557	0.9686	0.782	0.9571	817	339	11	178	
5	MobileNet	Normal	0.6892	0.4076	0.8477	0.5505	0.8348	671	256	46	372	Adadelta
		Pneumonia	0.9807	0.9396	0.9974	0.9677	0.9995	930	389	1	25	
		Pleural Thickening	0.739	0.4974	0.2714	0.3512	0.8178	899	95	255	96	
6	MobileNet	Normal	0.8647	0.7206	0.649	0.6829	0.9208	967	196	106	76	Adamax
		Pneumonia	1	1	1	1	1	955	390	0	0	
		Pleural Thickening	0.8565	0.6919	0.8086	0.7457	0.9283	869	283	67	126	
7	MobileNet	Normal	0.861	0.765	0.5497	0.6397	0.9257	992	166	136	51	Nadam
		Pneumonia	0.9993	1	0.9974	0.9987	1	955	389	1	0	
		Pleural Thickening	0.8164	0.6	0.8829	0.7145	0.9219	789	309	41	206	

Table 5: Top 7 Optimizer Wise Results of Deep Learning Models with CNN7 Convolutional Layers

The Adadelta, on the other hand, performed somewhat worse in all assessment criteria, implying that it is less effective for the proposed multiclass classification problem. The overall findings show that the integration of the CNN7 feature extraction block greatly improved the feature learning and classification performance. The Adam optimizer yielded the best consistent and balanced results in comparison to all optimization techniques.

6.4 Proposed Methodology Results

The multi classification confusion matrix using MobileNet + Adam optimizer with 7 convolutional layers is shown as the best model using the proposed methodology. Figure 5 shows the multi class classification confusion matrix. The confusion matrix explains the success of the model in classifying the images of normal, pneumonia, and pleural thickening properly. It also shows the cases of misclassification.

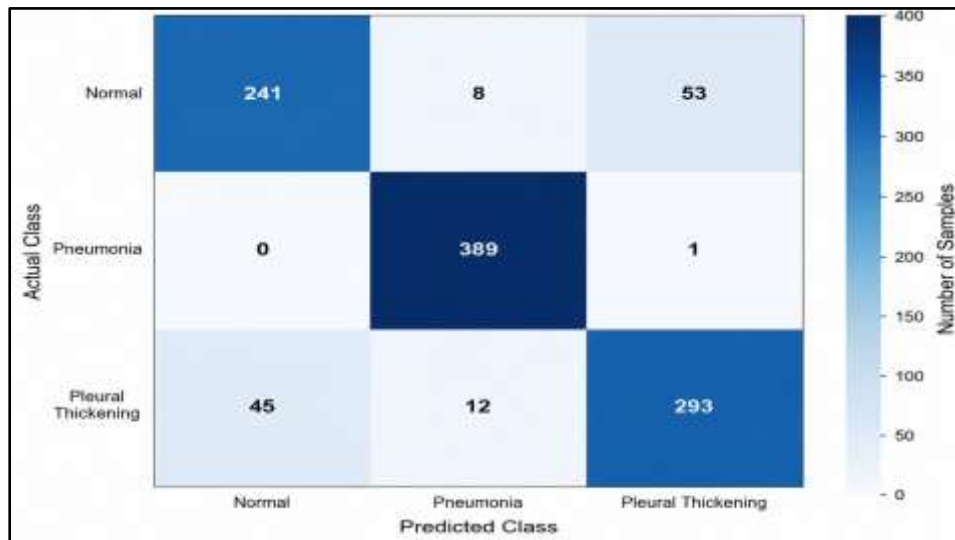


Figure 5: Confusion Matrix for Best Model (MobileNet + Adam with 7 Convolutional layers)

The diagonal members of the matrix signify that a total of 241 normal images, 389 pneumonia images, and 293 pleural thickening images were correctly identified. Of all three different classifications, pneumonia has the best ability to be accurately identified from other types of diseases. This can be seen since only 1 out of 255 images of pneumonia was incorrectly identified as pleural thickening, and no images of pneumonia were classified as normal. Therefore, this finding supports the ability of the model to accurately identify the radiographic characteristics associated with pneumonia.

For normal, there were 241 properly identified images; 8 misidentified images were labeled as pneumonia, and 53 were labeled as pleural thickening and thus incorrectly identified. For pleural thickening, there were 293 images properly identified, 12 were classified incorrectly as pneumonia, and 45 images were classified incorrectly as normal. The majority of the incorrect classifications occurred between the normal and pleural thickening categories because of the overlap in the radiographic characteristics, which made it difficult to differentiate between the two categories. When there are similar traits between classes, it can lead to confusion and subsequently reduces the overall classification performance. Still, even with these limitations, the MobileNet CNN7 layer performed extremely well and demonstrated strong discriminative capabilities for all three classes. These results demonstrate that MobileNet+ Adam optimizer using 7 convolutional layers can be used effectively for multiclass chest disease classification using X-ray images.

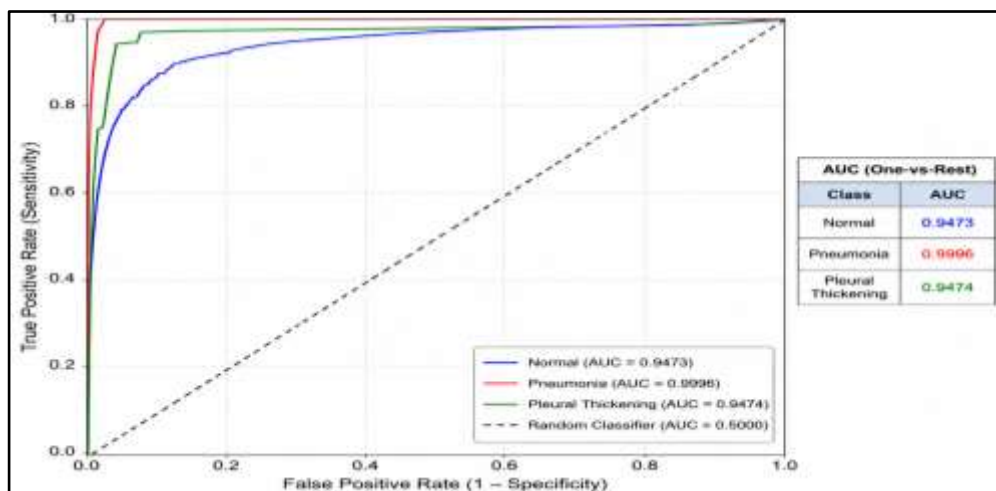


Figure 6: ROC Curve for Best Model (MobileNet + Adam with 7 Convolutional layers)

The ROC curve for the highest performance model (MobileNet + CNN7 with Adam optimizer) is shown in Figure 6. The true positive rate (TPR) is plotted on the Y-axis against the false positive rate (FPR) on the X-axis based on the accuracy of classification across each class, providing a strong visual indication of how well discriminating the model was able to perform in distinguishing classes from one another based upon the attributes of their respective classes. Furthermore, the pneumonia class attained an AUC of 0.9996, demonstrating outstanding classification capability and strong separability among the classes. The ROC curves for the normal and pleural thickening classes indicated AUC values of 0.9473 and 0.9474, respectively, signifying both classes were also highly effective at separating their various features through

classification, thus demonstrating that the model was effective in its ability to classify chest X-rays and achieve accurate classifications across multiple classes by learning features that differentiate them. The very high AUC values also provide additional support for the performance of the Adam optimizer as a training mechanism for optimal performance of the proposed architecture and for its efficient use for automatically detecting the presence or absence of pneumonia, normal, and pleural thickening in chest X-rays.

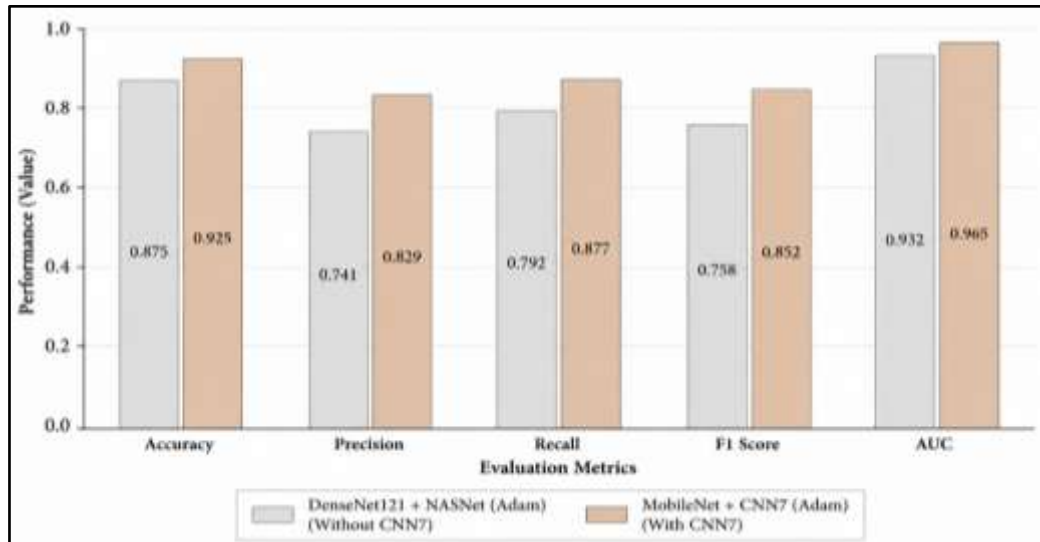


Figure 7: Performance comparison of the best models with and without CNN7

Figure 7 presents a comparative performance analysis of the best performing models obtained from experiments conducted with and without the CNN7 layered feature extraction block. The DenseNet121 + NASNet model optimized using the Adam optimizer achieved average accuracy, precision, recall, F1-score, and AUC values of 0.875, 0.741, 0.792, 0.758, and 0.932, respectively. In contrast, the proposed MobileNet model integrated with the CNN7 layer feature extraction block and optimized using the Adam optimizer achieved superior performance, attaining accuracy, precision, recall, F1-score, and AUC values of 0.925, 0.829, 0.877, 0.852, and 0.965, respectively. The findings provide compelling evidence of how well the CNN7 architecture works for improving feature extraction as well as multiclass classification performance. While all backbone architectures benefited from a performance increase provided by the CNN7 block, the most substantial improvements were observed in the MobileNet, which produced a more compatible feature set than those matched with deeper layered architectures. This allowed for improved learning of disease specific radiographic patterns and minimized inter-class confusion between normal, pneumonia, or pleural thickening chest X-ray images with the combination of MobileNet and CNN7 and using the Adam optimizer was the best performing configuration overall.

6.5 Comparison between Multiclass Classification with Existing Models

Performance of our proposed method was evaluated against previously published multi class chest X-ray classification methods outlined in the literature; Table 6 illustrates the differences between the AUC scores achieved by other multi class classification studies and our proposed method for pneumonia and pleural thickening classification purposes.

Other Approaches/Models	Pneumonia	Pleural Thickening
M. Irtaza et al. [3]	0.69	0.87
T.K.K. Ho and J. Gwak [4]	0.74	0.75
X. Gong et al. [5]	0.77	0.81
T.K.K. Ho and J. Gwak [6]	0.76	0.78
Baltruschat et al. [7]	0.69	0.73
S. Gundel et al. [8]	0.76	0.78
Proposed Methodology* (MobileNet + Adam optimizer) using 7 convolutional layers	0.99	0.94

Table 6: Comparison between Multiclass Classification with Existing Models

The literature indicates AUC values for pneumonia are between 0.69 and 0.77, and for pleural thickening, AUC values are from 0.73 to 0.94. In contrast, the developed method produced an AUC score of 0.9996 for pneumonia and 0.9474 for pleural thickening, demonstrating much better classification performance than existing methods. Based on this evidence, it is clear that the MobileNet+Adam Optimizer utilizing 7 convolutional layers successfully captures the radiographic features that are specific to each disease and provides superior discrimination ability compared to previously published multi-class classification methods. Additionally, when evaluated on an optimizer-based approach, the Adam optimizer was found to be the most effective optimizer, contributing to an overall increase in classification performance & model robustness. Collectively, these results provide strong evidence that our proposed method with a MobileNet+Adam optimizer utilizing 7 convolutional layers provides a reliable and accurate means of automatically classifying thoracic abnormalities from chest X-ray images into multiple classes. The next section discusses the conclusion and the future directions.

7 Conclusion And Future Outlook

This research developed a unique, deep learning model to classify lung diseases via chest X-ray images. The proposed multiclassification deep learning framework has a seven-layer convolutional neural network (CNN7) feature extraction block and performs a comprehensive optimizer wise evaluation for the classification of normal, pneumonia, and pleural thickening cases from a balanced dataset containing 7446 X-ray images.

The results showed that MobileNet + CNN7 architecture, using the Adam optimization algorithm, resulted in the most balanced and reliable performance (0.9473 normal, 0.9996 pneumonia, 0.9474 pleural thickening). Compared to the baseline models (without CNN7) and multiple previously published multiclass chest X-ray classification models, the proposed framework provided superior performance. Results showed that not only does CNN7 block significantly improve pulmonary disease specific feature learning, but it also decreases inter-class confusion between visually similar thoracic abnormalities. In addition, analysis of the optimizers showed the Adam optimization algorithm provided the best overall performance by increasing the rate of convergence and accuracy during classification. Ultimately, the proposed framework has the potential to be used as a reliable and accurate tool for computer aided diagnostic support to enhance the accuracy of lung disease screening and support clinical decision making. In the future, the proposed framework may be expanded to other thoracic diseases (pulmonary edema, tuberculosis, COVID-19, lung nodules, and lung cancer) in order to develop a more comprehensive multiclass classification system. Additional research may be conducted to examine cutting-edge attention-based techniques (i.e., transformer architecture) as well as advanced explanation methods (XAI) and ensemble learning to improve classification ability and interpretation. Additionally, validation utilizing large, many center clinical data sets and deployment into the proposed XRAYASSIST platform will provide support for radiologists to efficiently and accurately interpret chest X-ray images in the clinical setting.

References

- 1 Degefu, N., Jambo, A., Nigusse, S., Dechasa, M., Gashaw, T., & Getachew, M. (2023). The burden and contributing factors of cardiogenic pulmonary edema among acute heart failure patients admitted to tertiary hospital, Eastern Ethiopia. *Open Access Emergency Medicine*, 15, 405–414. <https://doi.org/10.2147/OAEM.S436352>
- 2 Yan, T., et al. (2023). Burden, trends, and inequalities of heart failure globally, 1990 to 2019: A secondary analysis based on the Global Burden of Disease 2019 study. *Journal of the American Heart Association*, 12(6). <https://doi.org/10.1161/JAHA.122.027852>
- 3 Irtaza, M., Ali, A., Gulzar, M., & Wali, A. (2024). Multi-label classification of lung diseases using deep learning. *IEEE Access*, 12, 124062–124080. <https://doi.org/10.1109/ACCESS.2024.3454537>
- 4 Ho, T. K. K., & Gwak, J. (2019). Multiple feature integration for classification of thoracic disease in chest radiography. *Applied Sciences*, 9(19), 4130. <https://doi.org/10.3390/app9194130>
- 5 Gong, X., Xia, X., Zhu, W., Zhang, B., Doermann, D., & Zhuo, L. A. (2021). Deformable Gabor feature networks for biomedical image classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 4004–4012). <https://doi.org/10.1109/WACV48630.2021.00407>
- 6 Ho, T. K. K., & Gwak, J. (2020). Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities. *IEEE Access*, 8, 160749–160761. <https://doi.org/10.1109/ACCESS.2020.3019935>
- 7 Baltruschat, I., Nickisch, H., Grass, M., Knopp, T., & Saalbach, A. (2019). Comparison of deep learning approaches for multi-label chest X-ray classification. *Scientific Reports*, 9, 6381. <https://doi.org/10.1038/s41598-019-42294-8>

- 8 Gündel, S., Grbic, S., Georgescu, B., Liu, S., Maier, A., & Comaniciu, D. (2018). Learning to recognize abnormalities in chest X-rays with location-aware dense networks. In *Proceedings of the Iberoamerican Congress on Pattern Recognition* (pp. 757–765). Springer.
https://doi.org/10.1007/978-3-030-13469-3_89
- 9 Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., & De Boer, F. (2023). High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. *Computers in Biology and Medicine*, 155, 106646.
<https://doi.org/10.1016/j.compbimed.2023.106646>
- 10 Sharma, S., & Guleria, K. (2023). A deep learning-based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks. *Procedia Computer Science*, 218, 357–366. <https://doi.org/10.1016/j.procs.2023.01.018>
- 11 Pranaya, A., Sowmya, D. V., Poojitha, L., Grace, P., Bhavya, K., & Ganapathi Raju, N. V. (2023). Pneumonia detection using deep learning. *E3S Web of Conferences*, 391, Article 01067.
<https://doi.org/10.1051/e3sconf/202339101067>
- 12 Račić, L., Popović, T., Čakić, S., & Šandi, S. (2021). Pneumonia detection using deep learning based on convolutional neural network. In *2021 25th International Conference on Information Technology (IT)* (pp. 1–4). IEEE. <https://doi.org/10.1109/IT51528.2021.9390137>
- 13 Ali, M., et al. (2024). Pneumonia detection using chest radiographs with novel EfficientNetV2L model. *IEEE Access*, 12, 34691–34707. <https://doi.org/10.1109/ACCESS.2024.3372588>
- 14 Goyal, S., & Singh, R. (2023). Detection and classification of lung diseases for pneumonia and COVID-19 using machine and deep learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, 14, 3239–3259. <https://doi.org/10.1007/s12652-021-03464-7>
- 15 Hole, S. R., Salotagi, S., Kolluru, V., Sagar, A., Sawale, G. J., & Y, J. (2025). Chest X-ray pneumonia detection using deep learning. In *2025 International Conference on Biomedical Engineering and Sustainable Healthcare (ICBMESH)* (pp. 1–6). IEEE.
<https://doi.org/10.1109/ICBMESH66209.2025.11182215>
- 16 Usman, C., Rehman, S. U., Ali, A., Khan, A. M., & Ahmad, B. (2025). Pneumonia disease detection using chest X-rays and machine learning. *Algorithms*, 18(2), 82.
<https://doi.org/10.3390/a18020082>
- 17 Agrawal, P., & Gandhi, S. (2025). *PneumoSense: Smart pneumonia detection using deep learning*. In *Proceedings of the 10th International Conference on ICT and Sustainable Development (ICT4SD 2025)*. Springer.
- 18 Shetty, I., Agrawal, P., & Gandhi, S. (2025). Deep learning based pulmonary edema detection: Performance evaluation of optimizers. *Eudoxus Press*.
<https://www.eudoxuspress.com/index.php/pub/article/view/3453/2484>
- 19 P. Agrawal and S. Gandhi, “PneumoSense++: An Enhanced Pneumonia Detection with Deep Learning and Comparative Optimizer Analysis,” *Procedia Computer Science*, vol. 283, pp. 3407–3423, 2026, doi: 10.1016/j.procs.2026.06.404
- 20 Shetty, I., Agrawal, P., & Gandhi, S. (2026). *Deep learning based pleural thickening detection: Performance evaluation of optimizers*. ICECI 2026, Springer. (Under publication).
- 21 National Institutes of Health. (n.d.). *NIH Chest X-rays dataset*. Kaggle.
<https://www.kaggle.com/datasets/nih-chest-xrays/data>
- 22 Rahimanshu. (n.d.). *COPD processed data*. Kaggle.
<https://www.kaggle.com/datasets/rahimanshu/copd-processed-data>
- 23 Sachin Kumar. (n.d.). *COVID, pneumonia, and normal chest X-ray images*. Kaggle.
<https://www.kaggle.com/datasets/sachinkumar413/covid-pneumonia-normal-chest-xray-images>
- 24 Fernando2rad. (n.d.). *X-ray lung diseases images (9 classes)*. Kaggle.
<https://www.kaggle.com/datasets/fernando2rad/X-ray-lung-diseases-images-9-classes>
- 25 Kolas, A. (n.d.). *Chest X-ray classification code repository*. Kaggle.
<https://www.kaggle.com/artiomkolas/code>
- 26 Dincer, T. (n.d.). *Labeled chest X-ray images*. Kaggle.
<https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images>
- 27 Mooney, P. (n.d.). *Chest X-ray images (Pneumonia)*. Kaggle.
<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- 28 Mehrparvar, F. (n.d.). *Chest X-ray dataset*. Kaggle.
<https://www.kaggle.com/fatemehmehrparvar/datasets>