



Grad-CAM guided transfer learning with optimizer evaluation for multi-disease chest x-ray classification

Sreena Vasantha Gopinathan^{1,2}

Deivendran Narain Ponraj³

Deepa Pushparajan Latha⁴

Kulandairaj Martin Sagayam⁵

Hatira Gunerhan⁶⁺

^{1,2}Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.

^{1,2}Email: sreenavg.ec@marian.ac.in

³Email: narainpons@karunya.edu

²Electronics and Communication Engineering, Marian Engineering College, Trivandrum, Kerala, India.

⁴Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, Trivandrum, Kerala, India.

⁴Email: deepa.pl@mbcet.ac.in

⁵Electronics and Communication Engineering, SRM TRP Engineering College (Autonomous), Trichy, Tamil Nadu, India.

⁵Email: martinsagayam.k@gmail.com

⁶Department of Mathematics, Kafkas University, Kars, Turkey.

⁶Email: gunerhanhatira@gmail.com



(+ Corresponding author)

ABSTRACT

Article History

Received: 22 June 2026

Revised: 14 September 2026

Accepted: 25 September 2026

Published: 23 October 2026

Keywords

Chest X-ray

Deep learning

DenseNet201

Grad-CAM

Medical image analysis

Optimizer

Transfer learning.

Lung diseases are serious health conditions that affect normal human life and are considered life-threatening. The timely diagnosis of lung infections is essential to ensure treatment and thus improve recovery rates. Clinicians most commonly use chest radiographs to detect conditions such as tuberculosis, lung opacity, COVID-19, lung cancer, and pneumonia. Interpreting chest X-rays manually is both time-consuming and dependent on individual expertise, which can introduce variability. In this context, this study aims to examine the impact of optimizer selection on transfer learning for multiclass classification of chest X-ray images while incorporating explainability through visualization. Five pretrained deep learning network models, including InceptionV3, InceptionResNetV2, ResNet50, DenseNet, and NASNetMobile, were fine-tuned using transfer learning. The COVID QU-Ex dataset, comprising chest X-ray images categorized as COVID-19, normal, and non-COVID (pneumonia), was used to conduct the experiments. The models were trained and evaluated using competing optimizers, including Adaptive Moment Estimation (Adam), Stochastic Gradient Descent (SGD), and Root Mean Squared Propagation (RMSProp), to determine the most effective optimization strategy. Experimental results demonstrate that InceptionResNetV2 with Adam achieved 93.89% accuracy, DenseNet201 with SGD achieved 95.19% accuracy, and InceptionV3 with the RMSProp optimizer achieved 92.39% accuracy. Gradient-Weighted Class Activation Mapping was applied to the best-performing model, DenseNet201, to effectively visualize the chest X-ray regions that contribute to prediction and classification.

Contribution/Originality: This study contributes to the existing literature by demonstrating the added benefits of optimizer tuning, transfer learning, and visualization techniques for multi-class chest X-ray classification. The primary contribution is the empirical evaluation of these strategies to improve model performance and interpretability.

1. INTRODUCTION

The lungs are a major part of the human respiratory system, which is a network of tissues and organs that helps us breathe. Lung infections can occur when microorganisms such as bacteria, fungi, and viruses enter the lungs and cause inflammation. Polluted air, smoking, and allergens can also contribute to lung diseases. Infections can range from mild to severe, and medical attention is needed if the condition worsens. Common diseases that affect the lungs include pneumonia, tuberculosis, lung cancer, and COVID-19.

A lung infection can be diagnosed through spirometry, blood cultures, and imaging methods such as CT scans and chest X-rays. The outcomes of various tests depend on the quality of the samples used for evaluation and may therefore sometimes yield false results. Chest X-rays are imaging methods that use lower doses of radiation to produce images of organs, tissues, and bones associated with the respiratory system. In addition, fluid, abnormalities, and lesions within the lungs can be detected effectively using chest X-rays. This cost-effective and highly efficient diagnostic modality for lung diseases is preferred by physicians. Automated detection approaches are needed to support clinical practice and enable the early diagnosis of chest diseases.

Although remarkable research has been conducted over the past decades in computer-aided technology, much more remains to be done in the medical imaging sector to improve existing solutions for detecting and managing dangerous lung diseases. As lung diseases such as tuberculosis and pneumonia can be life-threatening, greater attention is needed to detect them effectively. Extracting appropriate features to distinguish between different lung disorders is a complex problem, and conventional machine-learning approaches in computer-aided diagnosis systems may lack sufficient performance. This has necessitated the automation of feature extraction, which offers promising solutions for addressing many real-world problems, particularly in the medical sector.

Deep learning techniques for feature extraction have advanced drastically, and convolutional neural networks (CNNs) have demonstrated an exceptional capability to learn discriminative features automatically from chest X-ray images. This capability has made CNNs the backbone of most recent automated diagnostic systems for thoracic diseases using chest X-ray images. However, training CNNs from scratch is often limited by the scarcity of labelled datasets. Therefore, transfer learning has emerged as an effective strategy for leveraging knowledge from large-scale datasets such as ImageNet.

Despite their success, the efficacy of CNNs is highly dependent on the choice of architecture, optimizer, and training strategy. Most existing studies evaluate only one or two architectures with a single optimizer, preventing an objective comparison. The common practice of fine-tuning the entire network on limited medical data leads to overfitting and poor generalizability. Additionally, clinical explainability is not adequately addressed. To overcome these gaps, this study proposes an optimized transfer-learning approach for multiclass chest X-ray classification, incorporating Grad-CAM visualization. The specific objectives of this study are:

1. To implement a multiclass chest X-ray image classification system using five pretrained deep learning models and three optimizers.
2. To use the publicly accessible COVID-QU-Ex dataset, comprising 33,920 chest X-ray images, to classify COVID-19, pneumonia, and healthy chest X-rays.
3. To enhance clinical trust by integrating Grad-CAM-based visual explainability. It is applied to identify the localized feature regions selected by the best-performing models to make correct predictions.

The main contributions of this paper are as follows: a systematic evaluation of 15 pretrained model-optimizer combinations for multiclass chest X-ray image classification; specification of model and optimizer hyperparameters to improve reproducibility; and the incorporation of visual explainability using Grad-CAM to support the clinical validation of automated predictions. The remainder of this paper is organized as follows: Section 2 discusses the related literature on chest X-ray classification. Section 3 describes the dataset, proposed methodology, and pretrained models employed in this study. Section 4 presents the experimental results and performance analysis of the proposed deep learning models. Finally, Section 5 concludes the paper and outlines future work.

2. LITERATURE REVIEW

Deep learning techniques provide strong predictive models without prior knowledge of the underlying features in medical images. There is a risk of disease misinterpretation in medical images; therefore, extracting appropriate discriminative features is essential to reduce false predictions. Alshmrani, et al. [1] utilized VGG19 augmented with three-layer convolutional neural network (CNN) blocks for feature extraction and a fully connected classification stage. The method achieved 96.48% accuracy, enabling healthcare specialists to diagnose chest diseases quickly. The lightweight Chest X-Ray6 model [2] was designed and trained on over 21,000 chest X-rays to classify images of patients with COVID-19, cardiomegaly, pneumonia, pleural opacity, and lung opacity.

An optimized DenseNet201 model, termed LDDNet [3], was proposed, and its efficiency was evaluated for a three-class problem involving the classification of healthy, pneumonia, and COVID-19-affected chest X-rays. The model's performance was evaluated using the Adam, SGD, and Nadam optimizers. The best performance was achieved with the Nadam optimizer. LDDNet achieved 96.52% classification accuracy on a highly imbalanced dataset of 5,935 images and 97.07% accuracy on a balanced dataset of 5,002 images. Researchers also developed a fine-tuned VGG16 network, named LungNet22 [4], to classify lung diseases into nine categories. Using 80,000 images compiled from chest X-rays in public repositories, the system achieved 98.89% accuracy.

It was observed that the performance of various optimizers used in disease classification highly depends on the characteristics of the input images used for training. The RMSProp optimizer, along with LeakyReLU activation, was best suited for a modified Xception network [5], while Adamax with a cross-entropy loss function worked well for a fine-tuned DenseNet121 model [6]. The authors demonstrated the power of Bayesian optimization [7] for chest X-ray image-based COVID-19 diagnosis.

An automated system for classifying four diseases against healthy chest X-ray images was proposed [8]. The performance of eight pre-trained convolutional neural network models, including Darknet 53, AlexNet, GoogLeNet, DenseNet 201, Darknet 19, InceptionResNetV2, MobileNetV2, and ResNet 18, was evaluated on a dataset comprising 3,500 images. Cross-validation with different K values was performed, and the system achieved its highest accuracy of 97.49% with DenseNet 201 at K = 10. A weighted-average ensemble meta-classifier model using Local Interpretable Model-Agnostic Explanations (LIME) to highlight affected regions was designed [9]. In addition, a fine-tuned Mobile LungNetV2 [10] was demonstrated for chest disease classification.

Before being used as input to deep learning models, medical images undergo necessary preprocessing and data augmentation steps. Researchers usually apply resizing, normalization, noise cancellation, and contrast enhancement techniques to improve image quality and thus make the images compatible with the chosen deep network models. A Deep Convolutional Spiking Neural Network classifier for multiclass lung disease classification was implemented [11]. The system utilized an Anisotropic Diffusion Filter for noise removal, followed by an empirical wavelet transform for feature extraction, before being fed to the classifier. The detection accuracy of the system was significantly improved by using an Arithmetic Optimization Algorithm.

A convolutional neural network model incorporating residual network functions and dilated convolution was designed and named CDC-Net [12] for classifying chest X-rays for lung diseases such as pneumothorax, COVID-19, lung cancer, tuberculosis, and pneumonia. Vision transformers have recently gained more attention in medical image classification because of their enhanced performance compared with convolutional neural networks. The Input-Enhanced Vision Transformer (IEViT) was proposed for chest disease classification [13]. This method showed improved performance over all variants of vision transformers for four sets of chest X-ray images across the tuberculosis, COVID-19, and pneumonia categories.

A deep-learning architecture incorporating a VGG16-based feature-extraction module and classifiers such as neural networks, K-nearest neighbour, support vector machines, Naive Bayes, and random forest was demonstrated for the effective diagnosis of lung diseases [14]. An ensemble model comprising ResNet50, VGG19, and VGG16 achieved 98.34% accuracy for binary classification and 94.68% accuracy in distinguishing among COVID-19, non-

COVID pneumonia, and healthy chest X-rays [15]. The performance of ResNet50, SqueezeNet, and EfficientNet-B0 was evaluated using the Adam, RMSProp, and SGDM optimizers to achieve the best possible classification accuracy for chest X-rays [16]. COVID-DenseNet, which combines multi-scale fusion and an attention mechanism, achieved an accuracy of 96.89% [17].

Medical experts utilize computer-based disease diagnosis systems to better interpret lung diseases using chest X-ray images. Such systems incorporate enhancement techniques to reduce noise and improve image quality, thereby increasing the efficiency of lung disorder identification. Although deep learning techniques have boosted system efficiency, analyses reveal that outcomes are highly dependent on several factors. Deep learning techniques require a sufficient amount of data to train the system and develop fully efficient models. Data imbalance in available datasets has become a concern in developing deep learning algorithms. The relative number of positive disease images is comparatively lower in most datasets, which exacerbates overfitting issues and causes the developed model to become biased. The limited availability of datasets has forced researchers to use data augmentation techniques, which allow multiple data samples to be created from existing ones through rotation, flipping, scaling, translation, and other methods. The discussed literature is summarized in Table 1.

Table 1. Summary of existing deep learning models for CXR classification with reported accuracy.

References	Model + Optimizer	Accuracy (%)
Alshmrani, et al. [1]	VGG19 + 3 CNN blocks	96.48
Nahiduzzaman, et al. [2]	Chest X-Ray6	97.60
Podder, et al. [3]	LDDNet: DenseNet201 + Nadam	97.07
Shamrat, et al. [4]	LungNet22: VGG16	98.89
AbdElhamid, et al. [5]	Modified Xception + RMSProp	99.3
Chauhan, et al. [6]	DenseNet121 + Adamax	98.32
Loey, et al. [7]	Bayesian Optimized Model	96
Karaddi and Sharma [8]	DenseNet201	97.2
Jangam, et al. [9]	Ensemble + LIME	98
Shamrat, et al. [10]	Mobile LungNetV2	96.97
Malik, et al. [12]	CDCNet: Res + Dilated Conv	99.39
Okolo, et al. [13]	IEViT: Vision Transformer	98.59
Sharma and Guleria [14]	VGG16 + ML Classifiers	95.4
Kuzinkovas and Clement [15]	Ensemble of ResNet50, VGG16, VGG19	94.68
Ruan and Zhang [17]	COVID-DenseNet	96.89

3. MATERIAL AND METHODS

The medical industry requires more human resources to examine imaging data, such as chest radiographs and CT scans. Researchers have demonstrated that deep learning methods can be used to predict various lung diseases based on chest imaging. This method works well when such models are trained with a huge number of medical images. Deep learning employs numerous hidden-layer neurons that are well trained to achieve high performance, especially in challenging tasks. It has the advantage of automating the feature-extraction step and thus improves resource and time efficiency. Therefore, it has a significant impact on the field of medical science. Available medical data are multi-dimensional and noisy and may cause difficulty in extracting distinct features. Therefore, processing the source data is essential before feeding it to deep learning models. The depth of the network measures the number of trainable parameters. Developing a new deep learning model is time-consuming, so researchers focus on transfer-learning techniques to implement a disease-prediction model. In transfer learning, network models trained for a dedicated task using millions of images can be reused by researchers for a newly defined task. These networks are capable of efficiently extracting dominant features from a given dataset.

3.1. Dataset

Researchers from Qatar University collated the COVID-QU-Ex dataset [18, 19] comprising 33,920 chest X-ray images: 11,263 non-COVID cases (bacterial and viral pneumonia), 11,956 COVID-19 cases, and 10,701 normal (healthy) cases. The dataset is publicly available under the Creative Commons Attribution 4.0 International License (CC BY 4.0). It includes PA/AP frontal chest X-ray images with a resolution of 256×256 pixels in PNG format. All images in the dataset were included in this study. The dataset shows a minor class imbalance across categories. As patient-level metadata were not available, an image-level split was performed, with 27,134 images used for training and 6,786 for testing, while preserving class balance to avoid bias. Table 2 depicts the distribution of images used for training and testing in the proposed work.

3.2. Deep Learning Models

Various deep learning models analyzed in this work include ResNet50, DenseNet201, InceptionV3, InceptionResNetV2, and NASNetMobile. All the pre-trained models were trained on the ImageNet database, which includes millions of images split among 1,000 categories. A transfer learning approach was used to modify the existing models so that they could be effectively utilized for new prediction tasks.

Table 2. Distribution of CXR images in the dataset.

Dataset	COVID-QU-Ex
Total images	33920
COVID-19	11956
Pneumonia	11263
Normal	10701
Images used for training	27134 (9564 COVID-19, 9010 Pneumonia, and 8560 Normal)
Images used for testing	6786 (2392 COVID-19, 2253 Pneumonia and 2141 Normal)

ResNet is a convolutional neural network (CNN) model that stands for Residual Network. The network is built by stacking a number of residual blocks together. ResNet-50 [20] represents a 50-layer convolutional neural network architecture with 48 convolutional layers, one max-pooling layer, and one average-pooling layer. It follows a bottleneck design for its blocks, which reduces the number of trainable parameters and computations, thus making training faster. DenseNet-201 [21] is a convolutional neural network model that uses densely arranged blocks to connect each layer to every other layer. The first level of convolutional and pooling layers is followed by dense blocks and transition layers. A dense block includes batch normalization, ReLU activation, and convolution. The accuracy of the model depends on the density of connections between layers. As each layer shares the learned features with subsequent layers, this model strengthens feature propagation across the layers and thus reduces the vanishing gradient problem.

Table 3. Characteristics of deep learning models.

Network	Convolutional layer depth	Parameters (in million)
ResNet50	50	25.6
DenseNet201	201	20.0
InceptionV3	48	23.9
InceptionResNetV2	164	55.9
NASNetMobile	153	5.3

InceptionV3 [22] is a CNN model that improves upon the error rate of its predecessors. Large filter sizes make convolutions computationally expensive. InceptionV3 addresses this issue by factorizing larger convolutions into smaller ones and further into asymmetric convolutions. To alleviate vanishing-gradient problems, the network incorporates auxiliary classifiers that also act as regularizers. It has been shown that introducing residual connections

into Inception models (Inception-ResNet) [23] speeds up training and enhances system performance. Another compact CNN model with high performance in ImageNet data classification is NASNet Mobile (Neural Architecture Search (NAS) Network) [24]. The network can be represented using normal and reduction cell formats. The characteristics of the deep learning models used in this work are depicted in Table 3.

3.3. Proposed Model Architecture

The proposed deep-learning framework for multiclass chest X-ray-based lung disease classification is shown in Figure 1. The framework comprises preprocessing, transfer learning using five pre-trained CNN models, and classification. This study systematically evaluates the performance of five pre-trained CNN models in combination with three optimizers: Adam, SGD, and RMSProp.

Initially, all images from the COVID-QU-EX dataset were preprocessed to make them suitable for deep learning models. Image denoising techniques were applied to remove noise associated with the images. Resizing and normalization were employed to enhance the generalization capability of the trained model when classifying unseen test data. Subsequently, the dataset was split into training and testing sets using an 80:20 ratio. A random seed of 42 was set in Python to ensure reproducibility across experiments. In this work, no separate validation dataset was used.

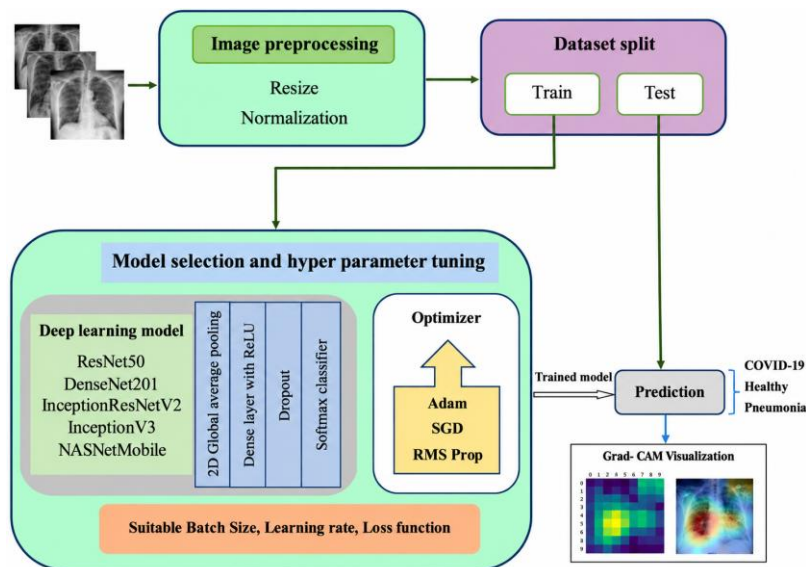


Figure 1. Proposed framework for Lung disease classification.

Algorithm 1: Pseudocode of the Proposed Transfer Learning Framework

Input: Dataset D , Model Set M , Optimizer Set O
Output: Best Model M_{best} and Performance Metrics

1. Load dataset D
2. Resize all images to 224×224 pixel resolution for ResNet50, DenseNet201, NASNetMobile and 299×299 pixel resolution for InceptionV3, Inception-ResNetV2
3. Normalize pixel values of images to $[0,1]$ and apply Denoising
4. Split dataset D into D_{train} and D_{test} with split ratio 80:20
5. **for** each model $m \in M$ **do**
6. Load m with ImageNet weights and freeze base layers
7. Add new classification head : 2D Global Average Pooling $\rightarrow$ Dense layer (512) with ReLU activation $\rightarrow$ Dropout(0.4) $\rightarrow$ Dense(3, Softmax classifier)
8. **for** each optimizer $o \in O$ **do**
9. Train m on D_{train} for 25 epochs
10. Evaluate performance of trained model on D_{test} and calculate the performance metrics to determine the best model $M_{best} \leftarrow m$
11. **end for**
12. **end for**
13. Report results of M_{best} with O_{best} on D_{test}

Figure 2. Pseudocode for the proposed architecture.

Five pre-trained deep network models, namely ResNet50, DenseNet201, InceptionResNetV2, InceptionV3, and NASNetMobile, were used to evaluate the impact of well-known optimizers on lung disease prediction from chest X-ray images through feature extraction and classification. Fine-tuning was performed on the pre-trained network models, which were then retrained to effectively learn new features from the dataset.

Transfer learning was initialized using ImageNet weights. The initial convolutional layer features of the base models were frozen, while the final layer characteristics were fine-tuned to learn the specific attributes of a different dataset and assigned task. A 2D global average pooling layer was included in the classification head of all base models to reduce overfitting by lowering the number of trainable parameters. A dense layer (512 units) with ReLU activation, followed by a dropout layer, was used to minimize overall system loss and thus enhance the generalization capability of the models. A final dense layer with softmax activation was used for the 3-class problem. Suitable hyperparameters, such as batch size, learning rate, loss function, and optimizers, were selected. The pseudocode of the proposed transfer learning framework is presented in Figure 2.

Optimizers are algorithms used to mould and shape deep learning networks into their most accurate form by adjusting model parameters, such as weights and learning rates. Thus, they enable models to converge faster with minimum loss and maximum accuracy and improve the generalization ability of models for prediction on unseen data. Selecting suitable optimizers based on the application is a critical task in deep learning. Hyperparameter tuning and optimization strongly influence the training speed and generalization ability of deep learning models. Optimizers that perform best in one task may yield poor results in other tasks. Thus, no single optimizer dominates its competitors in all defined tasks. Various optimizers, such as Adam, SGD, and RMSProp, were used to train the proposed method. The hyperparameter and implementation settings of the proposed framework are presented in Table 4.

Table 4. Hyperparameter and implementation settings of the proposed framework.

Parameter	Value
Input Image Size	224×224 , 299×299
Data Split	80% Train, 20% Test.
Batch Size	16
Max. Epochs	50
Loss Function	Sparse Categorical Cross-Entropy
Learning Rate	0.0001 for Adam, RMSProp. 0.01 for SGD
Adam Optimizer	$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-6}$
RMSProp Optimizer	$\rho = 0.9$, $\epsilon = 1 \times 10^{-6}$
SGD Optimizer	momentum = 0.9
Dropout Rate	0.4
Random Seed	42
Shuffle	True for training
Validation Strategy	Test Set only. No validation set used
Hardware and Software Versions	NVIDIA GTX 1660Ti 6GB GPU, CUDA 12.1, Python 3.10

4. RESULTS AND DISCUSSION

The performance of the proposed pretrained models, incorporating various optimizers such as Adam, SGD, and RMSProp, was evaluated using a publicly accessible dataset comprising chest X-ray images, namely COVID Qu-Ex. All input images were resized according to the requirements of the pretrained models, i.e., 224×224 for ResNet50, DenseNet201, and NASNetMobile, and 299×299 for InceptionV3 and InceptionResNetV2, to make them suitable for the deep learning models. Images were normalized to ensure that all image features were on a common scale.

The concept of transfer learning facilitates the repurposing of pretrained networks for novel classification problems. In the proposed approach, the early convolutional layers of each base model were frozen, preserving the ImageNet-pretrained weights without fine-tuning. A global average pooling 2D layer was then applied to downsample the feature maps, thereby reducing model complexity and the risk of overfitting. To build the classification head, a dense layer with ReLU activation and a dropout layer were appended. This modification improves learning efficiency and minimizes training loss. The dataset used in this work included images that were

relatively equally distributed across three categories: pneumonia, COVID-19, and healthy images. Therefore, data augmentation was not performed during training. The training-to-test image ratio was set at 80:20. Of the 33,920 images in total, the training dataset contained 9,564 COVID-19, 9,010 pneumonia, and 8,560 normal chest X-ray images. The test dataset comprised 2,392 COVID-19, 2,253 pneumonia, and 2,141 healthy chest X-ray images.

The loss function used was sparse categorical cross-entropy. A batch size of 16 was used for training over 50 epochs. The pre-trained models were trained using three different optimizers, i.e., Adam, SGD, and RMSProp. The performance of the models trained with these optimizers was evaluated using the test dataset. An Intel Core i7 processor with a NVIDIA GTX 1660Ti 6GB GPU was used for the experiments in Keras with a TensorFlow backend and Python. Accuracy, recall, precision, and F1-score, as described in Equations 1-4 were the evaluation metrics used to measure the effectiveness of the proposed method.

$$Acc = \frac{t_P + t_N}{f_P + f_N + t_P + t_N} \quad (1)$$

$$Rec = \frac{t_P}{f_N + t_P} \quad (2)$$

$$Prec = \frac{t_P}{f_P + t_P} \quad (3)$$

$$F1Score = \frac{2(Precision * Recall)}{Precision + Recall} \quad (4)$$

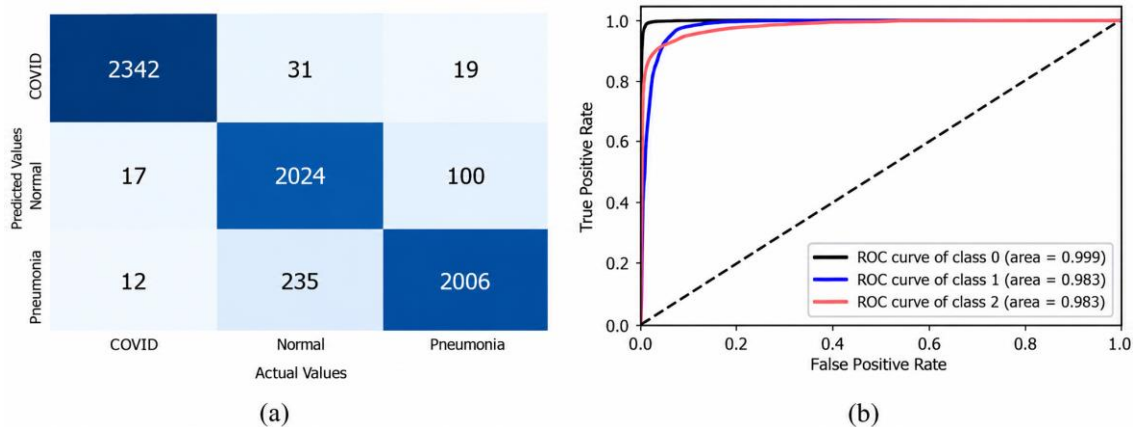


Figure 3. Confusion matrix and ROC curves obtained for the InceptionResNetV2 model with the Adam optimizer.

Table 5. Comparative analysis of evaluation metrics for proposed architectures with Adam optimizer.

Network	Class	TP	TN	FN	FP	Precision	Recall	F1 Score	Acc. (%)
ResNet50	COVID	2197	4344	50	195	91.85%	97.77%	94.72%	89.71
	Normal	1928	4295	350	213	90.05%	84.64%	87.26%	
	Pneumonia	1963	4235	298	290	87.13%	86.82%	86.97%	
InceptionV3	COVID	2334	4374	20	58	97.58%	99.15%	98.36%	93.3
	Normal	1911	4457	188	230	89.26%	91.04%	90.14%	
	Pneumonia	2087	4287	166	246	89.46%	92.63%	91.02%	
InceptionResNetV2	COVID	2342	4365	29	50	97.91%	98.78%	98.34%	93.89
	Normal	2024	4379	266	117	94.54%	88.38%	91.36%	
	Pneumonia	2006	4414	119	247	89.04%	94.40%	91.64%	
DenseNet201	COVID	2365	4321	73	27	98.87%	97.01%	97.93%	93.17
	Normal	1954	4427	218	187	91.27%	89.96%	90.61%	
	Pneumonia	2004	4361	172	249	88.95%	92.10%	90.49%	
NASNetMobile	COVID	2312	4359	35	80	96.66%	98.51%	97.57%	91.55
	Normal	1843	4430	215	298	86.08%	89.55%	87.78%	
	Pneumonia	2058	4210	323	195	91.34%	86.43%	88.82%	

Where t_n , t_p , f_n , and f_p denote true negatives, true positives, false negatives, and false positives, respectively. These quantities are derived directly from the confusion matrix. The confusion matrix provides a detailed comparison between actual ground-truth labels and model predictions and thus serves as a fundamental tool for evaluating classification performance. The left diagonal represents the total number of diseased and normal cases correctly predicted by the model, while the right diagonal shows the incorrect classifications. Receiver operating characteristic (ROC) curves were plotted for all models using different optimizers. The ROC curves were depicted in a one-versus-rest format, with one class considered positive and the remaining classes considered negative. As this work highlights a three-class classification task, Classes 0, 1, and 2 represent the COVID-19, Normal, and Pneumonia classes, respectively. For Class 0, COVID-19 was taken as the positive class, and the other predictions were considered negative.

With the Adam optimizer, the modified InceptionResNetV2 achieved the highest accuracy (93.89%), while ResNet50 gained an accuracy of 89.71%, InceptionV3 with 93.30%, DenseNet201 with 93.17% and NASNetMobile with 91.55%. For SGD optimizer, modified DenseNet201 outperformed with 95.19% accuracy while ResNet50 gained an accuracy of 91.74%, InceptionV3 with 94.7%, InceptionResNetV2 with 92.79% and NASNetMobile with 93.45%. For RMSProp optimizer, modified InceptionV3 outperformed with 92.39% accuracy while ResNet50 gained an accuracy of 89.19%, InceptionResNetV2 with 90.12%, DenseNet201 with 89.77% and NASNetMobile with 88.97%.

Table 6. Comparative analysis of evaluation metrics for proposed architectures with SGD optimizer.

Network	Class	TP	TN	FN	FP	Precision	Recall	F1 Score	Acc. (%)
ResNet50	COVID	2359	4245	149	33	98.62%	94.06%	96.29%	91.74
	Normal	1998	4334	311	143	93.32%	86.53%	89.80%	
	Pneumonia	1869	4433	100	384	82.96%	94.92%	88.54%	
InceptionV3	COVID	2366	4359	35	26	98.91%	98.54%	98.73%	94.7
	Normal	2008	4448	197	133	93.79%	91.07%	92.41%	
	Pneumonia	2053	4406	127	200	91.12%	94.17%	92.62%	
InceptionResNetV2	COVID	2291	4343	51	101	95.78%	97.82%	96.79%	92.79
	Normal	1963	4440	205	178	91.69%	90.54%	91.11%	
	Pneumonia	2043	4300	233	210	90.68%	89.76%	90.22%	
DenseNet201	COVID	2376	4371	23	16	99.33%	99.04%	99.19%	95.19
	Normal	2018	4460	185	123	94.26%	91.60%	92.91%	
	Pneumonia	2066	4415	118	187	91.70%	94.60%	93.13%	
NASNetMobile	COVID	2362	4337	57	30	98.75%	97.64%	98.19%	93.45
	Normal	1931	4453	192	210	90.19%	90.96%	90.57%	
	Pneumonia	2049	4338	195	204	90.95%	91.31%	91.13%	

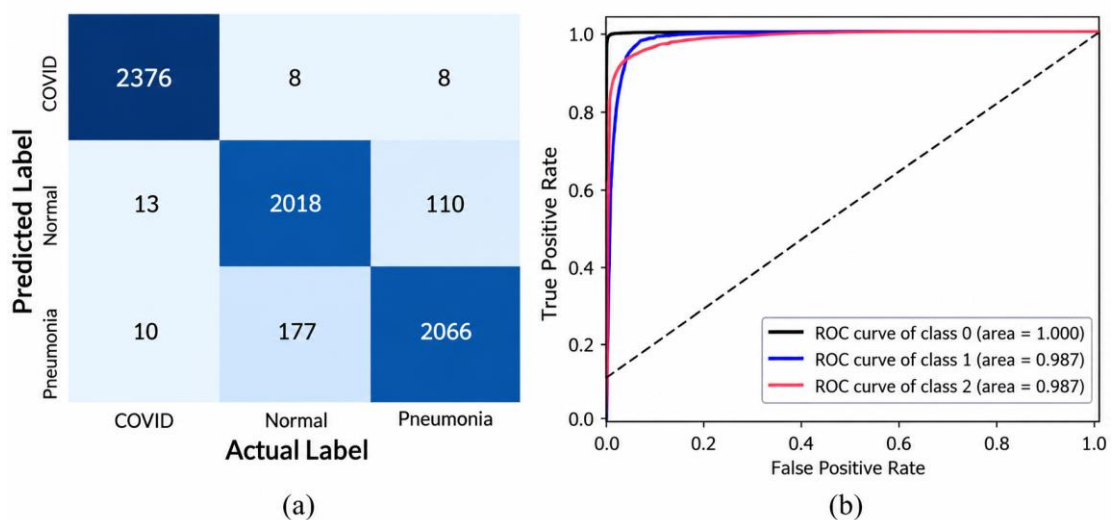
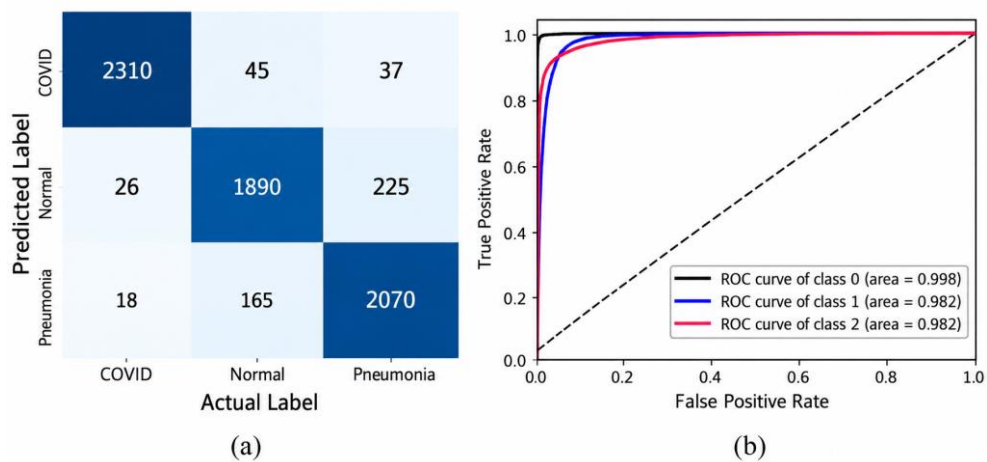


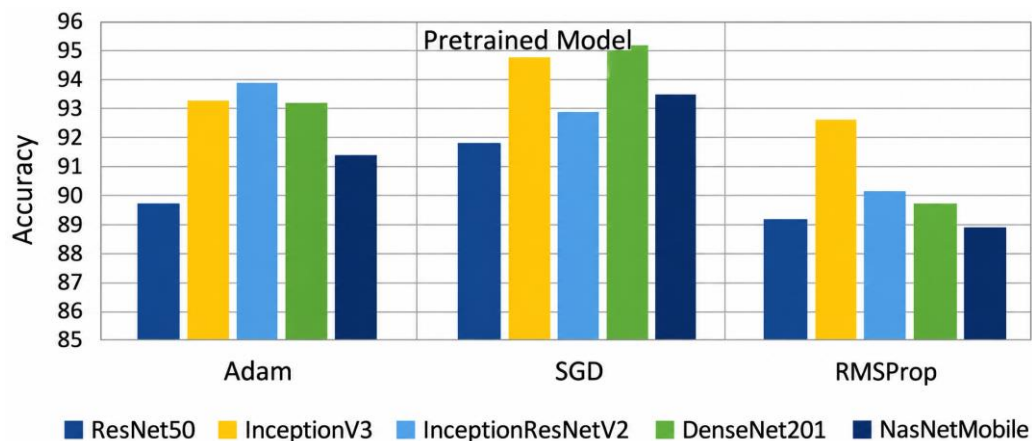
Figure 4. Confusion matrix and ROC curves obtained for the DenseNet201 model with the SGD optimizer.

Table 7. Comparative analysis of evaluation metrics for proposed architectures with the RMSProp optimizer.

Network	Class	TP	TN	FN	FP	Precision	Recall	F1 Score	Accuracy (%)
ResNet50	COVID	2290	4269	125	102	95.74%	94.82%	95.28%	89.19
	Normal	1787	4394	251	354	83.47%	87.68%	85.52%	
	Pneumonia	1976	4176	357	277	87.71%	84.70%	86.18%	
InceptionV3	COVID	2310	4350	44	82	96.57%	98.13%	97.35%	92.39
	Normal	1890	4435	210	251	88.28%	90.00%	89.13%	
	Pneumonia	2070	4271	262	183	91.88%	88.77%	90.29%	
InceptionResNetV2	COVID	2312	4307	87	80	96.66%	96.37%	96.51%	90.12
	Normal	1806	4399	246	335	84.35%	88.01%	86.14%	
	Pneumonia	1998	4196	337	255	88.68%	85.57%	87.10%	
DenseNet201	COVID	2388	4074	320	4	99.83%	88.18%	93.65%	89.77
	Normal	1856	4416	117	397	82.38%	94.07%	87.84%	
	Pneumonia	1848	4388	257	293	86.31%	87.79%	87.05%	
NASNetMobile	COVID	2271	4288	106	121	94.94%	95.54%	95.24%	88.97
	Normal	1790	4364	281	351	83.61%	86.43%	85.00%	
	Pneumonia	1977	4172	361	276	87.75%	84.56%	86.13%	

**Figure 5.** Confusion matrix and ROC curves obtained for the InceptionV3 model with the RMSProp optimizer.

All models achieved their highest accuracy on the test data using the SGD optimizer, except InceptionResNetV2, which achieved its best results using Adam. Table 5, 6, and 7 depict the evaluation metrics for the proposed architectures using different optimizers. Confusion matrices and receiver operating characteristic curves for the best-performing models using the three optimizers, namely Adam, SGD, and RMSProp, are shown in Figures 3-5. The comparison of the accuracy obtained by the proposed models using different optimizers is consolidated in Figure 6.

**Figure 6.** Comparison of accuracy for proposed models with various optimizers.

4.1. Grad CAM Visualization

Gradient-weighted Class Activation Mapping [25] uses the gradients of the target class at the final convolutional layer of a deep learning model to produce a localized region map that highlights the key regions of an image that contribute to the classification prediction. This method is used to perform visual validation of a designed model by precisely measuring the patterns responsible for the prediction process. The output produced by the Grad-CAM algorithm is an activation heatmap for any predicted label. Figure 7 shows the Grad-CAM visualization of COVID-19 test images using the DenseNet-201 model with the SGD optimizer. Column 1 shows the ground-truth CXR images with COVID-19 infection. According to study Tahir, et al. [18] chest X-ray images were localized for COVID-19 infections with 83.05% IoU (Intersection over Union) and 88.21% DSC (Dice Similarity Coefficient). The infection masks for ground-truth COVID-19-infected chest X-rays were supplied with the dataset. Because radiologist annotations were unavailable, we used these infection masks to verify heatmap localization in our work. Columns 2 and 3 depict the activation heatmap for the last convolutional layer, conv5-block32-2, of the DenseNet-201 model and the corresponding intensity levels, respectively. High-intensity values (yellow in the heatmap) represent areas with a high degree of infection, whereas low-intensity values represent areas with a lower degree of infection. Because the key regions in the heatmaps differ across CXRs, the model is not biased toward the given dataset. In comparison with the infection masks, the model can highlight regions and patterns of interest to predict a particular class label. The activation heatmap is superimposed on the original image and shown in Column 4.

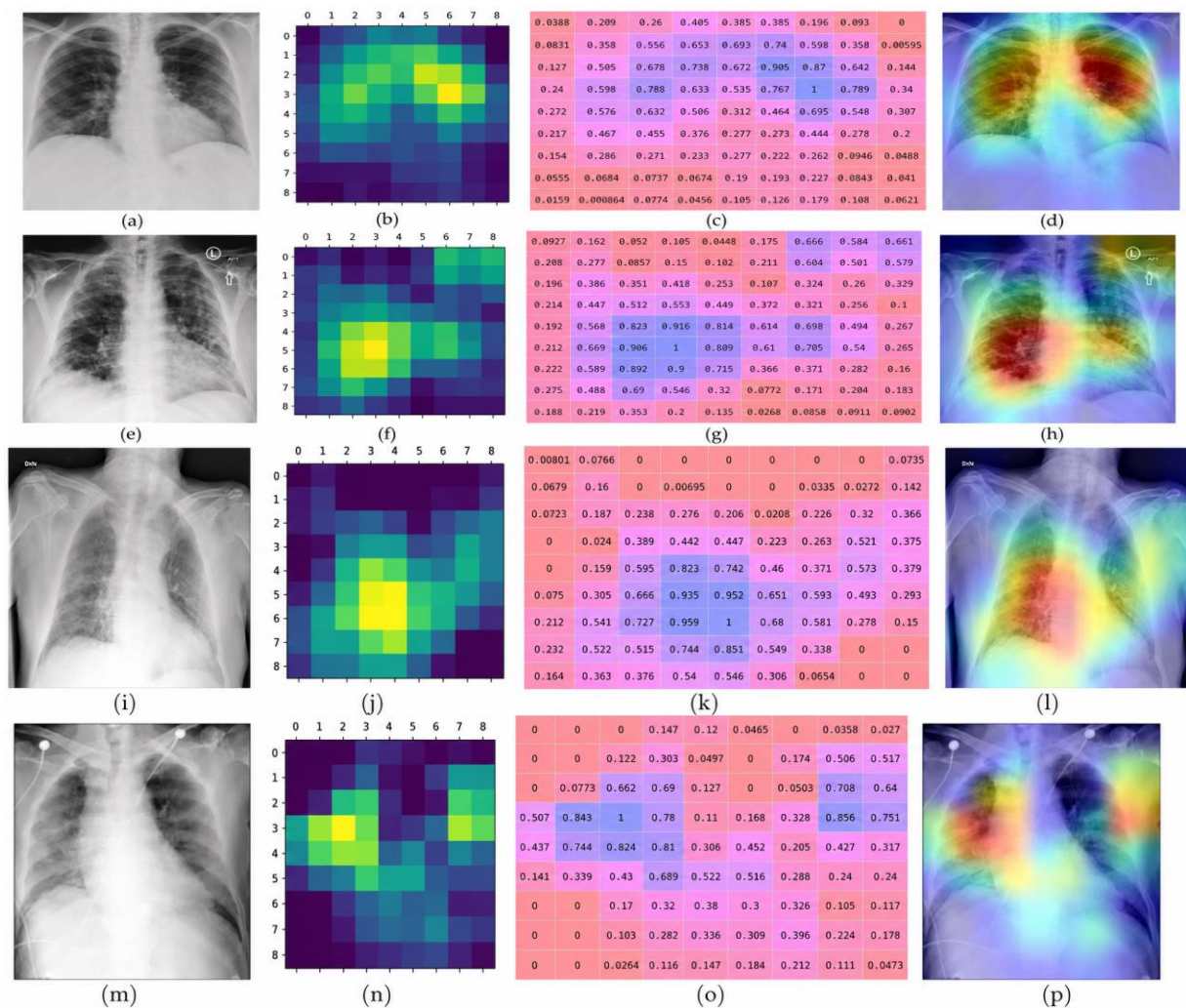


Figure 7. Grad CAM visualization of test images (COVID-19) for DenseNet 201 with SGD optimizer. Column 1- Original Chest X-Ray images with COVID-19 infection, Column 2- Colour-coded activation heatmap by Grad- CAM algorithm, Column 3- Intensity values corresponding to the activation heatmap, Column 4- Activation heatmap superimposed on original image.

A comparison of the proposed method on the COVID-QU-Ex three-class dataset is presented in Table 8. The superior performance of SGD over adaptive optimizers such as Adam can be attributed to its regularization properties. SGD with momentum tends to converge to flatter minima, leading to better generalization on medical datasets such as COVID-QU-Ex. Regarding computational efficiency, the proposed DenseNet201 model contains 20.01 million trainable parameters and requires 4.29 GFLOPs per inference. Compared with ResNet50, this represents a 22% reduction in parameters while using a deeper architecture. When evaluated on an NVIDIA GTX 1660Ti 6GB GPU, the model achieved an average inference latency of 22.4 ± 0.9 ms per CXR image, equivalent to a throughput of 45 images per second. This efficiency enables deployment of the model in real-time clinical workflows with limited computational resources.

Table 8. Comparison of the proposed method on COVID-QU-Ex 3-class dataset.

Reference+ Model	Accuracy (%)
Ensemble of ResNet50, VGG16,VGG19 [15]	94.68
InceptionV3+ RMSProp	92.39
InceptionResNetV2+Adam	93.89
Proposed DenseNet201 + SGD	95.19

5. CONCLUSIONS

This study investigates the impact of widely used optimizers, such as Adam, SGD, and RMSProp, on transfer learning-based chest X-ray classification. Experiments were conducted on the COVID QU-Ex dataset for three classes: COVID-19, pneumonia, and normal. Five pretrained deep-learning models InceptionV3, ResNet50, InceptionResNetV2, DenseNet201 and NASNetMobile were fine-tuned and evaluated. The results demonstrate that DenseNet201 with SGD achieved the highest classification accuracy of 95.19%. Grad-CAM visualization revealed that the convolutional layers of the best-performing model effectively highlighted localized lung regions and pathological patterns relevant to predicting the correct class label. The findings underscore the significance of optimizers in transfer learning and visualization techniques for clinical explainability. This study utilized a publicly accessible dataset without patient metadata, and no external validation was performed by radiologists. Future research aims to scale the current dataset to a larger, multi-institutional dataset with additional classes. The efficacy of the network can be further improved by comparing model predictions with chest X-ray images annotated by radiologists.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Engineering Journal*, vol. 64, pp. 923-935, 2023. <https://doi.org/10.1016/j.aej.2022.10.053>
- [2] M. Nahiduzzaman, M. R. Islam, and R. Hassan, "ChestX-Ray6: Prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network," *Expert Systems with Applications*, vol. 211, p. 118576, 2023. <https://doi.org/10.1016/j.eswa.2022.118576>
- [3] P. Podder *et al.*, "LDDNet: A deep learning framework for the diagnosis of infectious lung diseases," *Sensors*, vol. 23, no. 1, p. 480, 2023. <https://doi.org/10.3390/s23010480>

- [4] F. M. J. M. Shamrat *et al.*, "LungNet22: A fine-tuned model for multiclass classification and prediction of lung disease using X-ray images," *Journal of Personalized Medicine*, vol. 12, no. 5, p. 680, 2022. <https://doi.org/10.3390/jpm12050680>
- [5] A. A. AbdElhamid, E. AbdElhalim, M. A. Mohamed, and F. Khalifa, "Multi-classification of chest X-rays for COVID-19 diagnosis using deep learning algorithms," *Applied Sciences*, vol. 12, no. 4, p. 2080, 2022. <https://doi.org/10.3390/app12042080>
- [6] T. Chauhan, H. Palivela, and S. Tiwari, "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging," *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100020, 2021. <https://doi.org/10.1016/j.ijime.2021.100020>
- [7] M. Loey, S. El-Sappagh, and S. Mirjalili, "Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data," *Computers in Biology and Medicine*, vol. 142, p. 105213, 2022. <https://doi.org/10.1016/j.combiomed.2022.105213>
- [8] S. H. Karaddi and L. D. Sharma, "Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks," *Expert Systems with Applications*, vol. 211, p. 118650, 2023. <https://doi.org/10.1016/j.eswa.2022.118650>
- [9] E. Jangam, C. S. R. Annavarapu, and A. A. D. Barreto, "A multi-class classification framework for disease screening and disease diagnosis of COVID-19 from chest X-ray images," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 14367-14401, 2023. <https://doi.org/10.1007/s11042-022-13710-5>
- [10] F. M. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Computers in Biology and Medicine*, vol. 155, p. 106646, 2023. <https://doi.org/10.1016/j.combiomed.2023.106646>
- [11] R. Rajagopal, R. Karthick, P. Meenalochini, and T. Kalaichelvi, "Deep convolutional spiking neural network optimized with arithmetic optimization algorithm for lung disease detection using chest X-ray images," *Biomedical Signal Processing and Control*, vol. 79, no. Part 2, p. 104197, 2023. <https://doi.org/10.1016/j.bspc.2022.104197>
- [12] H. Malik, T. Anees, M. Din, and A. Naeem, "CDC_Net: Multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 13855-13880, 2023. <https://doi.org/10.1007/s11042-022-13843-7>
- [13] G. I. Okolo, S. Katsigiannis, and N. Ramzan, "IEViT: An enhanced vision transformer architecture for chest X-ray image classification," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107141, 2022. <https://doi.org/10.1016/j.cmpb.2022.107141>
- [14] S. Sharma and K. Guleria, "A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks," *Procedia Computer Science*, vol. 218, pp. 357-366, 2023. <https://doi.org/10.1016/j.procs.2023.01.018>
- [15] D. Kuzinkovas and S. Clement, "The detection of COVID-19 in chest X-rays using ensemble CNN techniques," *Information*, vol. 14, no. 7, p. 370, 2023. <https://doi.org/10.3390/info14070370>
- [16] G. D. Deepak and S. K. Bhat, "Detection of COVID-19 & non-COVID diseases from chest-X-rays using deep learning-based CNNs," *Soft Computing*, vol. 30, no. 1, pp. 449-471, 2026. <https://doi.org/10.1007/s00500-025-10905-4>
- [17] L. Ruan and C. Zhang, "Covid-19 diagnosis from X-ray images using Deep Learning," in *Proceedings 2023 IEEE 9th International Conference Cloud Computing and Intelligent Systems (CCIS), Dali, China*, 2023, pp. 213-216.
- [18] A. M. Tahir *et al.*, "COVID-QU-Ex dataset, Kaggle, dataset, 2021," Retrieved: <https://www.kaggle.com/datasets/anasmohammedtahir/covidqu>. 2021.
- [19] M. E. H. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665-132676, 2020. <https://doi.org/10.1109/ACCESS.2020.3010287>
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

- [21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.
- [22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
- [23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proceedings 31st AAAI Conference Artificial Intelligence (AAAI)*, 2017, pp. 4278–4284.
- [24] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8697–8710.
- [25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings IEEE International Conference Computer Vision (ICCV)*, 2017, pp. 618–626.

Views and opinions expressed in this article are the views and opinions of the author(s). Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.