

Enhanced Multi-Class Pulmonary Disorder Detection Using Hard Voting Ensemble of CNN Models on X-Ray Images

Marwa Elpeltagy¹, Ebeid Ali Ebeid², Farida Youness³

^{1,3}Systems and Computers Department, Faculty of Engineering,
Al-Azhar University, Egypt, Cairo

²Department of Mathematics (Computer Science), Faculty of Science,
Al-Azhar University, Egypt, Cairo

Article Info

Article history:

Received Apr 23, 2025

Revised Aug 27, 2025

Accepted Sep 13, 2025

Keyword:

Chest X-Ray
Efficientnetb0
Densenet-201
Resnet50
MobilenetV2
InceptionnetV3

ABSTRACT

Lung diseases represent a major public health concern, requiring timely and accurate diagnosis. Chest X-rays are widely used for initial screening, but manual interpretation is time-consuming and subject to variability among radiologists. To address these challenges, this study presents an automated deep learning-based framework for multi-class lung disease detection. The proposed approach integrates five convolutional neural network (CNN) architectures—EfficientNetB0, DenseNet201, ResNet50, MobileNetV2, and InceptionV3—within a hard-voting ensemble classifier to improve diagnostic performance. Transfer learning is applied to extract deep features from chest X-ray (CXR) images, and the ensemble strategy enhances overall accuracy compared to individual models. The system was evaluated into six categories, including normal, COVID-19, tuberculosis, opacity, bacterial pneumonia, and viral pneumonia. Results demonstrate that the ensemble achieves approximately 97% accuracy, outperforming current state-of-the-art methods. Furthermore, the model shows strong capability in differentiating bacteria from viral pneumonia, underscoring its potential as a reliable tool for automated lung disease diagnosis in clinical practice.

Copyright © 2025 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Marwa Elpeltagy
Department of Systems and Computers, Faculty of Engineering
Al-Azhar University
Egypt, Cairo
Email: marwaelpeltagy812.el@azhar.edu.eg

1. INTRODUCTION

Lung disease is one of the leading causes of death worldwide, including pneumothorax, tuberculosis, and pneumonia. Pneumothorax is an illness that damages the air sacs in the lungs [1]. The indications include cough, fever, loss of appetite, shortness of breath, stabbing chest pain, fatigue, confusion, and vomiting, especially in older adults. Pneumothorax is a different illness that may cause the lung to collapse partially or completely [1]. Intense activities such as climbing mountains, traveling, or diving for fun have a chance to cause lung collapse. These intense activities may result in substantial fluctuations in air pressure. Symptoms of a pneumothorax include discomfort in the chest, feeling short of breath, dryness, difficulty breathing, blue skin, a hacking cough, and tiredness. Tuberculosis is a medical condition caused by a contagious bacterial infection of the bronchial tubes [1].

Coughing for three weeks in a row, fever, chills, appetite loss, and night sweats that result in inadvertent weight loss are all signs of tuberculosis. Furthermore, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was the trigger of the global pandemic known as COVID-19, which was initially discovered in Wuhan, China (December 2019). Since then, the virus has propagated quickly throughout the world, resulting in over (6 million) deaths and about (481 million) confirmed infections, based on the World Health Organization. Antigen tests and reverse transcriptase-polymerase chain reaction (RT-

PCR) are the main diagnostic methods for identifying COVID-19 infection. Nevertheless, RT-PCR has drawbacks, especially regarding sensitivity, which results in a significant number of false negatives. Additionally, because RT-PCR is a laboratory-based method, it takes several hours to produce diagnostic outcomes. On the contrary, antigen tests are less sensitive than RT-PCR but provide faster results [2].

A quick and precise diagnostic test that can help isolate infected people early, reduce virus transmission, and allow for the prompt administration of targeted treatments is desperately needed, considering these difficulties. As a dependable and accurate method of early-stage COVID-19 detection, radiographic imaging has gained significant traction in response to the limitations of PCR-based testing. Another technique for identifying COVID-19 infections is diagnostic imaging, specifically computed tomography (CT). In addition to revealing lesions that may indicate the severity of the illness, X-ray images can detect lung abnormalities linked to COVID-19.

However, using CT imaging to diagnose COVID-19 necessitates analyzing several chest slices, which can be difficult and time-consuming. Furthermore, it is still difficult to differentiate COVID-19 from other types of viral pneumonia by using CT results. Even though COVID-19's distinctive CT findings are widely known [3], Conventional techniques for diagnosing pneumonia from CXR images take a lot of time and call for specific medical knowledge, which can result in incorrect diagnoses and therapy delays [4]. Due to these difficulties, radiologists require computerized diagnostic instruments that can help them rapidly and precisely diagnose COVID-19 through the examination of CT scans and X-ray images. Lung disease identification and therapy research has attracted a lot of interest worldwide.

Deep learning (DL) methods, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), have been used to automate the detection of lung disease from CXR to get around these challenges. The application of DL techniques is an intriguing answer to this diagnostic issue [5]. When it comes to medical imaging, DL has gained rapid recognition, particularly in the analysis of various medical image modalities (e.g., [6], [7], and [8]), such as CT (e.g., [9], and [10]). Diagnostic imaging using DL has shown promise in several fields, including pathology and radiology.

The effectiveness of these techniques in spotting anomalies in medical images has prompted more research into deep CNNs for the categorization of lung conditions in CT and X-ray scans. There are two different kinds of DL models: pretrained and non-pretrained models. While pretrained models were previously trained on publicly accessible image datasets, non-pretrained models must be trained from scratch and require a sizable image dataset. In numerous domains, pretrained models that extract higher-order semantic features tend to perform better than other image classification methods.

Several DL-based methods have been put forth to evaluate chest images and find lung infections linked to pulmonary disease. For instance, Silva et al. [11] presented a technique based on the EfficientNet [12] model that analyzes 2D chest CT scan images using transfer learning (TL) and a voting-based classification system

Hemdan et al. [13] Several DL-based methods have been put forth to evaluate chest images and find lung infections linked to pulmonary disease. For instance, Silva et al. [11] presented a technique based on the EfficientNet [12] model that analyzes 2D chest CT scan images using transfer learning (TL) and a voting-based classification system. Hemdan et al. [13] investigated several DL algorithms for the classification of COVID-19 images, and DenseNet showed the best accuracy at 90%.

ECOVNet is an ensemble framework based on EfficientNet variants, developed for COVID-19 detection from chest X-ray images. By leveraging transfer learning and combining multiple EfficientNet models, ECOVNet achieved higher robustness and accuracy compared to single CNN architectures. The ensemble strategy helped reduce overfitting and improved generalization across different datasets. This demonstrated the effectiveness of EfficientNet-based ensembles for medical image classification during the COVID-19 era [14].

Chowdhury et al. [15] accomplished an optimal accuracy of 97.94% by applying AlexNet, DenseNet201, ResNet18, and SqueezeNet to a dataset that included 423 COVID-19 positive X-ray pictures, 1485 viral pneumonia, and 1579 normal X-ray images. Additional investigation into DL models for COVID-19 detection [16] showed that DenseNet121, EfficientNet, VGG16 and NASNet, and Xception, conducted differently, with EfficientNet outperforming the others with an accuracy of 93.48%. Ensemble strategies are machine learning techniques that combine several base models, sometimes known as "weak" models, to produce a more robust and effective model. In terms of predictive accuracy and robustness, an ensemble approach typically performs better than its base models when certain requirements are met, such as guaranteeing diversity among the base models.

The study [17] suggests a method for tackling the COVID-19 detection problem that makes use of 3D Inception CNNs. After developing and comparing the Inception-V1 and Inception-V3 models, performance was improved by using soft-voting ensemble classifiers. According to the findings, Inception-V1 performed better than Inception-V3 across several performance indicators. Furthermore, the ensemble

models produced the best results and greatly improved performance, especially when soft voting was used on Inception-V1.

Two methods for classifying COVID-19 patients using chest X-rays are examined in the study in [18]: DL and conventional machine learning approaches. The first method combines the choices of three DL classifiers using two different decision fusion techniques: Bayes optimal and majority voting. The second method combines feature fusion and decision fusion to enhance classification performance.

The study in [19] makes use of a Kaggle dataset source that covers a range of lung pathologies, such as bacterial lung conditions, COVID-19 infections, tuberculosis, and normal scans. A hybrid model that combined CNN and Long Short-Term Memory (LSTM) networks was used to achieve the best performance, yielding a 96.2% accuracy rate. By efficiently capturing the temporal and spatial characteristics of lung disorders, this integrated framework improves the detection system's performance and accuracy. The suggested model exhibits the capacity to quickly identify a variety of illnesses, including TB and COVID-19, enabling prompt medical attention. Its potential to standardize lung disease detection and transform the diagnostic process by spotting important patterns in medical images early on accounts for its clinical relevance.

An automated method for identifying several lung conditions in CT and X-ray scans is presented in a study [20]. The suggested system uses a novel image enhancement model for image classification in addition to two pre-trained DL models and a customized CNN. Pre-processing and deep learning-based classification are the two primary phases of the lung disease detection process. An algorithm for improving images is created during the pre-processing phase. Two CNN models that have already been trained—VGG16Net and AlexNet—are used in the classification stage. The outcomes highlight how well the image enhancement model works as a crucial part of the processing pipeline.

The EfficientNetB4 architecture was refined, verified, and tested by the authors in [21]. They used TL as a multilabel approach to identify zone-specific and image-specific signs of active pulmonary tuberculosis in CXR images. Several metrics were employed to assess the model's performance. Furthermore, they used CNNs' visualization capabilities to examine the model's decision-making process and display the anomalies linked to tuberculosis. Both the intramural test set and an external test set from a geographically diverse region were used to evaluate the EfficientNetB4 model, and the results showed outstanding performance.

To distinguish between three different pulmonary illnesses, COVID-19, bacterial pneumonia, and viral pneumonia DCNN-based image detection model that is optimized through image augmentation techniques is proposed in [22]. A dataset with four categories—healthy, COVID-19, bacterial pneumonia, and viral pneumonia—was used to train and assess the model. The three pulmonary disease classes were successfully classified by the model. When contrasted with conventional texture descriptor-based techniques for pulmonary disease identification using X-ray and CT scan images, these results demonstrate the efficacy of the suggested model.

The author [23] suggested a method for identifying lung conditions from X-ray pictures. VGG Data STN with CNN (VDSNet) is the name given to this methodology. The model is tested using both the full and sample versions of the NIH Chest X-ray image dataset, which is obtained from the Kaggle repository. The findings show that VDSNet performs better than some of the current techniques in several performance metrics, such as validation accuracy, precision, recall, and F0.5 score. Although it has a marginally lower validation accuracy, VDSNet shows a significant reduction in training time when applied to the sample dataset. Many investigators have investigated the application of DL techniques to address the difficulties in identifying COVID-19 infections since the first outbreak. This review of the literature systematically groups previous research according to the different DL approaches that were used and the experimental datasets. The review attempts to give a thorough overview of the methods and data sources that have been most frequently used in the detection of COVID-19 by structuring the research in this way.

The author [24] suggests two CNN-based models for COVID-19 detection. Whereas the second model incorporates both CT and X-ray images, the first model uses either CT or X-ray images independently. Metrics, including confusion matrix, accuracy, recall, precision, and F1 score, were used to assess both models' performance. The multi-model CNN outperformed other models, achieving an accuracy of 93.83% for two classes and 99% for three classes. The suggested models, especially the second one, have a lot of promise to help radiologists predict COVID-19.

In [25] a change to the ResNet50 architecture is suggested for determining whether a patient has COVID-19. The addition of three layers—a "Conv" layer, a "Batch_Normalization" layer, and a "Activation_RelU" layer—is intended to increase the accuracy of feature extraction and classification. Numerous tests using COVID-19 CXR and CT scan images show that the modified model outperforms other current techniques, achieving a diagnostic accuracy of 97.7% on CT images and 97.1% on X-ray images.

In [26] a solution utilizing cutting-edge machine learning techniques for COVID-19 detection from CXR and lung CT (LCT) images is presented. The system uses three TL algorithms to improve training efficiency and makes use of two cutting-edge deep neural network models, EfficientNet and MixNet.

The method seeks to develop an AI-based diagnostic system that is adaptable to various image sources. Five-fold cross-validation is used to validate the method on four real-world datasets, ranging from 15,000 to 411,528 images. The results demonstrate strong prediction performance, with accuracy consistently above 95% across all datasets.

A thorough analysis of DL methods for COVID-19 medical imaging diagnosis is provided in Paper [27]. Within Computer-Aided Diagnosis Systems (CADs), it addresses important subjects like segmentation, classification, explainable AI (XAI), and predictive research. The role of rehabilitation technologies, such as the Internet of Medical Things (IoMT), in COVID-19 care is also covered in the paper. It also discusses difficulties and potential lines of inquiry, emphasizing uncertainty quantification (UQ) in DL models for COVID-19 diagnosis. A discussion and synopsis of the main conclusions round out the review.

A system for using deep TL models to predict and classify a variety of respiratory diseases is presented in [28]. It examines 19,488 pulmonary images (CT scans and CXRs), including those of healthy lungs, lung cancer, COVID-19, pulmonary embolism, and pneumoconiosis. A hybrid model (EfficientNetB6 + ResNet101V2) yields the best results when a variety of models, such as EfficientNet, DenseNet, and ResNet, are used. Histogram equalization and contrast enhancement are used as preprocessing methods for the images, and contour and thresholding techniques are used to extract features. Enhanced predictions for respiratory disease classification are provided by the hybrid model, which achieves 99.77% accuracy, 1.00 precision, 0.99 recall, and 1.00 F1 score with a loss of 0.001.

The study [29] investigates the rapid and precise identification of COVID-19 patients using DL in conjunction with X-ray images. It optimizes several TL models, attaining high detection accuracy—100% for EfficientNetB4 and 99.55% for Xception, among others. Using a CXR dataset, the optimized EfficientNetB4 model also exhibits outstanding performance in identifying lung diseases, achieving 99.17% accuracy, 99.13% precision, and a 99.14% F1-score. The findings highlight the potential of optimized TL models for quick and precise COVID-19 diagnosis, providing radiologists and other medical professionals with invaluable assistance.

Using a Kaggle CXR dataset, the Vision Transformer (ViT) architecture was proposed in [30] for the detection of pneumonia. To capture spatial relationships and global context in the images, the ViT model makes use of self-attention mechanisms. According to experimental results, the suggested ViT-based framework detects pneumonia with 97.61% accuracy, 95% sensitivity, and 98% specificity. The ViT model is a strong option for pneumonia detection since it performs better than conventional CNNs by handling different image resolutions and offering superior context comprehension.

Using features taken from CT scans, cough sound images (CSI), and CXR, a study [31] suggested four new CNN models for categorizing nine chest diseases. The model integrates novel methods such as multiple-way data generation (MWDG), rank-based average pooling (RBAP), dropout, max-pooling, and batch normalization (BANL). The dataset, which consists of data from 24 publicly accessible datasets on chest disease, is balanced using the SMOTE technique before training. The suggested model outperforms seven baseline models (such as VGG-19, ResNet, and DenseNet) and cutting-edge classifiers with an accuracy of 99.01%. These findings show how the model can help medical professionals, including radiologists, diagnose conditions affecting the chest.

This paper, which depends on a modified version of the Xception model, presents a simplified and effective TL model for COVID-19 detection in [32]. To boost performance, the modification adds dense layers and average pooling. Both test and validation datasets are used to compare the suggested model to other TL models. The outcomes show that the model can make data-driven decisions for COVID-19 diagnosis by achieving high diagnostic accuracy on a sizable and difficult dataset of CT images.

In [33], authors employ physiological data, such as heart rate and step counts, from an existing dataset to predict the likelihood of COVID-19 infection before symptoms manifest. The performance of three machine learning models—Gradient Boosting (GB), CatBoost, and TabNet—is trained and contrasted. The performance of three machine learning models—Gradient Boosting (GB), CatBoost, and TabNet—is trained and contrasted. To better explain its predictions, an interpretability layer is added to CatBoost, the model with the best performance. Additionally, a private dataset from Fitbit devices was developed to guarantee accuracy and lessen bias. Outperforming earlier research, the CatBoost model obtained 81% accuracy on the private dataset and 85% accuracy on the public dataset.

An ensemble model for the early detection of COVID-19 infection using CXR images was proposed by Kumar et al. [34]. This method integrates multiple pre-trained models, such as EfficientNet, GoogleNet, and XceptionNet, to incorporate TL. With improved generalization on both binary and multiclass COVID-19 datasets, the ensemble model divides patients into different groups: COVID-19, pneumonia, TB, or healthy.

A comparative analysis was conducted on two well-known datasets, confirming that the model's performance was better than the state-of-the-art techniques currently in use. Alghamdi et al. (2025) proposed an ensemble of DenseNet169, MobileNetV2, and Vision Transformer (ViT) for pneumonia detection, achieving 93.91% accuracy and 93.88% F1-score. Their approach demonstrated strong performance through transfer learning and regularization but was limited by manual hyperparameter tuning and potential variance in ensemble predictions [35].

A CNN-LSTM ensemble was proposed for multi-class lung disease classification, covering six conditions, including pneumonia and COVID-19. Using preprocessing and SMOTE to handle imbalance, the model achieved 89.31% accuracy and outperformed several CNN ensembles. Despite strong results, issues of overfitting, limited temporal modeling, and dataset variability were noted [36].

Patel and Shah (2024) introduced a three-level deep ensemble model combining pretrained CNNs (VGG16, InceptionV3, and MobileNetV2) with machine learning classifiers for classifying four lung conditions. The framework achieved 94% accuracy, though its generalizability to broader lung diseases and the added complexity of multi-level ensembles remain challenges [37].

A study applied deep learning models to chest X-rays for diagnosing COVID-19, pneumonia, and normal cases, performing both binary (normal vs. COVID-19) and three-class (normal, COVID-19, viral pneumonia) classification. The ensemble model outperformed a single CNN, achieving **95% accuracy** in multiclass and **94%** in binary classification. However, the work was limited to a small set of classes and internet-sourced datasets, raising concerns about generalizability to broader clinical settings [38].

A recent study (2025) proposed a deep ensemble framework combining ResNet-152, VGG19, EfficientNet, and Vision Transformer (ViT) with a meta-model for cardiac and respiratory disease diagnosis on the NIH Chest X-ray dataset, covering 14 disease classes plus "No Finding." While individual models achieved ~88%, the ensemble achieved superior performance with lower Hamming Loss and higher ROC-AUC, underscoring the effectiveness of diverse ensembles for multi-label chest disease classification [39]. Table 1 presents a summary of the related study.

Despite significant progress in the application of DL algorithms in medical imaging, diagnosing multi-class lung disease remains challenging. One significant issue is the visual similarity between different lung conditions on chest X-rays, which occasionally display overlapping radiographic features that can be confusing to even experienced medical professionals. Additionally, because different institutions use different imaging equipment, different acquisition methods, and various patient positions, there is a lot of variance in image quality. In addition, as labeled data for certain lung disorders is often limited, it may be difficult for models to learn representative characteristics for underrepresented classes. These challenges demonstrate that thoughtful architecture and effective training methods are needed to improve the accuracy and generalizability of deep learning models in real-world clinical settings. Our study represents a breakthrough in healthcare diagnostics and COVID-19 detection techniques. This study's primary contribution is as follows:

1. A novel model based on the Deep Convolutional Neural Network (DCNN) architecture has been proposed; it is intended to classify X-ray images into multiple classes.
2. A sizable collection of chest X-ray (CXR) images for five distinct pulmonary conditions is gathered from various websites, which are current and widely accessible. Normal pulmonary diseases, COVID-19, TB, opacity, bacterial pneumonia, and viral bacterial lungs are the six categories into which pulmonary diseases are divided.
3. The following five transfer learning (TL) models are used to extract deep features from input chest X-ray (CXR) images: DenseNet201, ResNet50, MobileNetV2, EfficientNetB0, and InceptionV3. After that, an ensemble approach is used to detect lung disorders using these models. Through a hard-voting strategy, this ensemble classifier improves the individual models' performance by 4%.
4. The accuracy, precision, sensitivity, and F1-score are among the performance metrics used to assess the suggested system's performance, along with a confusion matrix.
5. Existing systems are contrasted with the suggested model in the literature. According to the findings, the proposed model performs better than alternative approaches and offers a more precise and ideal generalization of the findings.
6. Despite the substantial similarities between bacterial and viral pneumonia, the suggested method shows a high degree of accuracy in differentiating between the two illnesses.

The rest of this paper is structured as follows: Section 2 explains the algorithm of the proposed method. Section 3 discusses experimental results and is followed by Section 4, which represents the conclusion.

Table 1. Literature review in tabular form

Reference/year	Number of Classes	Model	Accuracy
[17], 2022	COVID-19 and Normal	3D Inception CNNs	—
[19], 2024	Normal, COVID-19 infections, tuberculosis, and bacterial lung conditions	Hybrid CNN and LSTM	96.2%
[20], 2023	collection of lung diseases	CNN models—AlexNet and VGG16Net	—
[21], 2023	Tuberculosis and Normal	EfficientNetB4	—
[22], 2024	healthy, COVID-19, Bacterial and viral pneumonia	DCNN	—
[23], 2020	lung diseases(NIH Chest X-ray image dataset)	VGG Data STN with CNN (VDSNet).	—
[24], 2021	COVID-19 and Normal	multi-model CNN	Two classes (93.83%) Three classes (99%)
[25], 2021	COVID-19 and Normal	Modified ResNet50	97.7% for CT 97.1% for X-ray
[26], 2022	COVID-19 and Normal	EfficientNet and MixNet	95%
[28], 2023	lung cancer, COVID-19, pulmonary embolism, pneumoconiosis, and healthy lungs	(EfficientNetB6 + ResNet101V2)	99.77%
[29], 2023	COVID-19 and Normal	EfficientNetB4	99.17%
[30], 2024	pneumonia using (Kaggle CXR dataset)	Vision Transformer (ViT)	97.61%
[31], 2024	nine chest diseases	Four novel CNN models	99.01%
[32], 2023	COVID-19 and Normal	modified version of the Xception model	—
[33], 2024	COVID-19 and Normal	Gradient Boosting (GB), CatBoost, and TabNet	85%
[34], 2021	COVID-19, pneumonia, tuberculosis, or healthy	ensemble model integrating several pre-trained models, including EfficientNet, GoogLeNet, and XceptionNet	—
[35], 2023	2 (Normal, Pneumonia)	Ensemble of DenseNet169, MobileNetV2, Vision Transformer (ViT)	93.91%
[36], 2025	6(Tuberculosis (TB), Normal, pneumothorax, pneumonia, COVID-19, and cardiomegaly)	CNN-LSTM Ensemble	89.31%
[37], 2024	COVID-19, lung opacity, Normal, and viral pneumonia.	Ensemble of VGG16, InceptionV3, MobileNetV2 with ML classifiers (SVM, RF, DT, KNN)	94%
[38], 2024	3 (Normal, COVID-19, viral pneumonia)	CNN, ViT+ Ensemble of (ResNet50+VGG-19)	95%
[39], 2025	14 different thoracic disease labels, in addition to a "No Finding" class	ensemble of ResNet-152, VGG19, EfficientNet, and Vision Transformer (ViT).	-

2. METHOD

The detection of lung disorders from CXR images is the foundation of our suggested ensemble model. Because it combines the prediction values of the five pre-trained DL models—EfficientNetB0, DenseNet201, ResNet50, Mobilenetv2, and InceptionV3net—its final prediction is higher than the predictions of any one model alone. The architecture was selected due to its demonstrated performance in a range of computer vision tasks and its capacity to manage radiological image complexity. These structures have been optimized for the detection of lung diseases. The structure of the proposed model is displayed as a block diagram in Figure 1, and Algorithm 1 contains Python-style QA pseudocode. The subsequent subsections provide a discussion of the suggested approach.

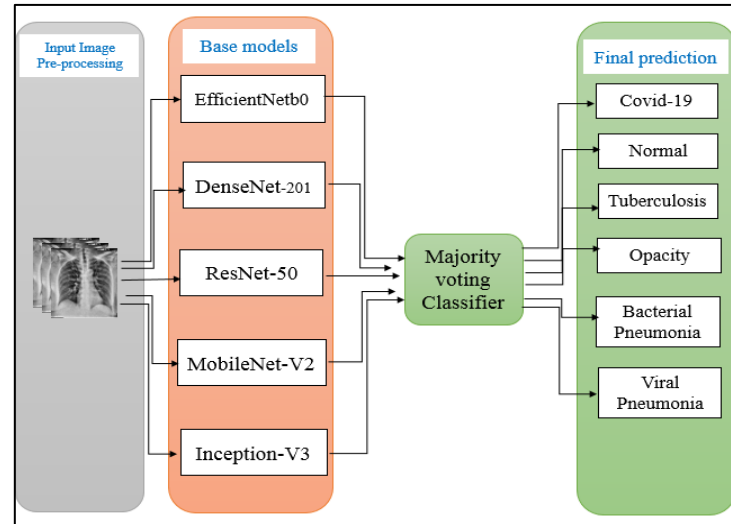


Figure 1. The proposed method framework.

Algorithm1 Algorithm for the proposed model

```

Input    X-ray images dataset
Output   Predicted lung disease label (normal, COVID-19, tuberculosis, opacity, bacterial pneumonia,
         viral pneumonia)

Step 1:  Initialize Parameters
         Lung_Disease_Types = ["normal", "COVID-19", "tuberculosis", "opacity", "Bacterial
         Pneumonia", "Viral Pneumonia"]
         Base_Models = [EfficientNetB0, DenseNet-201, ResNet50, MobileNetV2, InceptionV3Net]

Step 2:  Load and Preprocess X-ray Images
         Dataset = Load_Xray_Images("path ")
         Preprocessed_Dataset = Preprocess(Dataset)

Step 3:  Split Data into Training (80%) and Testing (20%) sets
         Training_Data, Testing_Data = Split_Data(Preprocessed_Dataset, train_ratio=0.8)

Step 4:  Initialize an empty list for storing model predictions
         Predicted_Labels_List = []

Step 5:  Training Each Model and Make Predictions
         for each model in Base_Models:
             Model(i) = trainNetwork(Training_Data, model, parameters)
             Predicted_Labels(i) = classify(Model(i), Testing_Data)
             Store Predicted_Labels(i) in Predicted_Labels_List

Step 6:  Ensemble Learning - Majority Voting
         Final_Predicted_Labels = Majority_Voting(Predicted_Labels_List)

Step 7:  Evaluate Model Performance
         Accuracy = sum(Final_Predicted_Labels == Testing_Data.Labels) / len(Testing_Data) * 100

Step 8:  Output Results
         print "Final Predicted Labels:", Final_Predicted_Labels
         print "Final Model Accuracy: ", Accuracy, "%"
  
```

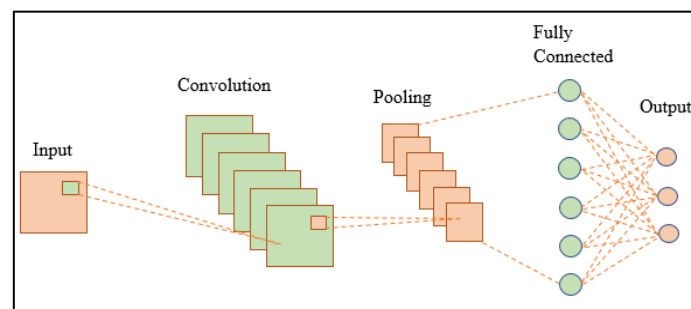


Figure 2. General architecture of CNN

2.1. Deep Transfer Learning (Individual Models)

CNNs are a type of deep neural network that is widely used in image analysis. A CNN is essentially made up of several modules that carry out convolutions, add nonlinearity to the framework, downsample capabilities, and finally classify images. A CNN starts with a convolutional layer and then moves on to ReLU, pooling, and fully connected layers. Convolutional filters are applied to the input feature map by the convolutional layer, which is the central part of a CNN. This process creates an output feature map. A CNN's general structure is depicted in Figure 2.

2.1.1. DenseNet-201

In [40], a DCNN was first presented. This network can classify images into 1,000 different categories and was trained on more than one million images from the ImageNet database [41]. The pretrained model has 201 layers, and the network's input image is 128 by 128 pixels in size. Batch normalization, the ReLU (Rectified Linear Unit) activation function, and a 3x3 filter-based convolution operation are all included in each layer. A method for reducing overfitting during training, batch normalization also improves the training process's stability and efficiency.

When a nonlinear transformation is applied, the ReLU function sets all negative values to zero. A pooling layer is used to shrink the feature map's size after the last dense layer. Following the pooling process, the likelihood that each image belongs to a specific class is ascertained using a LogSoftmax classifier.

2.1.2. Inception Network

The ability of inception networks to lessen overfitting is well known, especially when dealing with smaller datasets. Additionally, these networks' size enabled the trainable 3D model to be fitted to the device memory based on various configurations. The development of CNNs advanced significantly with the introduction of Inception networks. Increasing the depth of convolutional layers was the main goal of CNN architecture prior to the development of Inception. This strategy did have a few drawbacks, though. There were more parameters in the network as the number of layers increased, possibly going over the memory limit.

Moreover, particularly when trained on relatively small data sets, very deep architectures are more prone to overfitting. The Inception module was developed because of Inception networks switching from fully connected to sparsely connected network architectures to address these issues. A pooling path and convolutional filters of different sizes (such as 1, 3x3, and 5x5) that process the input image make up the basic Inception module.

The input for the next layer is a single vector created by concatenating the outputs of these operations. Large convolutions applied on top of layers with a lot of filters, however, can get computationally costly. The Inception module was improved to address this problem by adding 1x1 convolutions, which reduce dimensionality before the more computationally demanding 3x3. The so-called Inception-V1 network [42], which has 27 layers, was constructed using this Inception module.

The same study produced two subsequent optimizations, Inception-V2 and Inception-V3 [43]. The main goal of Inception-V2's architectural changes was to lower computing expenses in comparison to Inception-V1's predecessor. To increase the computational efficiency of convolutions, these optimizations were accomplished by utilizing sophisticated factorization techniques. Inception-V2 created two 5x5 convolutions by factorizing 3x3 convolution filters. This change had significant effects on accuracy and computation time.

While smaller kernel sizes, such as 3x3, are frequently chosen because of their lessened impact on network accuracy, which may be hampered by the representational bottleneck associated with larger kernel sizes, it is important to note that a 5x5 convolution is 2.78 times more expensive to compute than a 3x3 convolution. Furthermore, a 33% reduction in computational cost was achieved by factorizing nxn convolutions into a stack of 1xn and nx1 convolutions instead of using nxn convolutions. All the improvements from Inception-V2 are included in Inception-V3, along with additional features like (i) the RMSprop optimizer, (ii) label smoothing regularization, (iii) 7x7 factorized convolutions, and (iv) batch normalization in the fully connected layer. According to the experimental findings in the original paper, Inception-V3 outperformed both Inception-V1 and Inception-V2, achieving superior performance.

2.1.3. ResNet50

This pre-trained CNN, which has 50 layers overall, is a variation of the ResNet architecture. Of these layers, 48 are convolutional layers, and the other two are used for max pooling and average pooling, accordingly [44].

2.1.4. Efficient Net

It is a technique for CNN architecture and scaling that uses a compound coefficient to scale the network's depth, width, and resolution dimensions consistently [12]. Two fundamental ideas underpin the EfficientNet family: the baseline architecture's optimization and the efficacy of its scaling approach. The 2D Depth-wise convolution blocks that make up the foundational network, EfficientNet B0, have been shown to be extremely efficient in terms of both computational time and cost. More than a million photos from the ImageNet dataset were used to train EfficientNet-B0 [45].

Enhancing image classification accuracy has traditionally involved taking an existing network and adding more layers to increase its depth, adding more filters to each layer to increase its width, and then providing higher-resolution images with more detailed information. To avoid any one component becoming a performance-limiting factor, EfficientNet uses a combined scaling approach, increasing depth, width, and resolution all at once in a balanced way.

2.1.5. MobileNetV2

This CNN architecture was created with mobile devices in mind. MobileNetV2, in contrast to other CNN architectures, uses a unique structure where the connections are mostly between the bottleneck layers. Furthermore, the intermediate expansion layer efficiently filters nonlinear features by using depth-wise separable convolutions. An initial set of 32 convolutional layers and 19 bottleneck layers make up the MobileNetV2 architecture. In this work, we suggest a MobileNetV2 modification for the X-ray image-based pulmonary disease identification task.

In comparison to other DL models, its design offers several advantages. With small datasets, training can be challenging, and visual classification tasks are particularly prone to overfitting. By utilizing design elements that help avoid overfitting and provide a quick and effective architecture that maximizes memory usage with the least amount of error, MobileNetV2 reduces this risk. Additionally, as [46] shows, it allows for quick execution, which makes it a perfect model for experimentation and parameter optimization.

Table 2 summarizes the five CNN architectures used in the ensemble model, along with details on their parameter counts, structural features, and selection criteria in relation to the detection of lung disease.

Table 2. A Summary of CNN Architectures Used in the Ensemble Models.

Model	Features	Layers	Parameters	Justification for Lung Disease Detection
DenseNet201	Dense connections to reduce disappearing gradients and reuse features	201	~20M	Efficient in obtaining fine-grained characteristics and compatible with a small number of medical datasets
Inception Networks	Initialization modules for the extraction of multi-scale features	159	~23M	Handles a variety of visual patterns in lung disorders and records features at various spatial resolutions.
ResNet50	Deep architecture without degradation is made possible via residual connections.	50	~25.6M	Dependable in identifying small patterns; demonstrated in numerous medical imaging tasks
Efficient Net	Compound resolution, width, and depth scaling	237	~5.3M	Performance and efficiency are balanced by high accuracy with fewer parameters, making it perfect for thorough chest X-rays.
MobileNetV2	Using depth-wise separable convolutions in a lightweight design	88	~3.4M	Extremely effective; appropriate for deployment in real-time or mobile environments with limited resources in healthcare

These days, CNNs are a standard in medical imaging, particularly in the use of CT scans and CXRs to detect lung diseases. The CNN architecture choice has a significant impact on the model's efficiency, accuracy, and performance. Different CNN architectures, such as DenseNet, EfficientNet, MobileNet, ResNet, and InceptionNet, each have specific benefits and drawbacks based on their design and intended application. High-resolution scans may use a lot of memory, despite EfficientNet's reputation for state-of-the-art accuracy and computational efficiency. Although DenseNet can be computationally expensive, it performs well on smaller datasets and is excellent at feature propagation. ResNet's deep design and residual connections help it avoid vanishing gradients and capture fine details, but it's expensive. Despite its speed and portability, MobileNet might not be precise enough for mobile or real-time applications. Despite its intricate architecture, InceptionNet's ability to efficiently collect multi-

scale data makes it ideal for large datasets. The choice of architecture is influenced by the objective, size of the dataset, and computational capabilities; for classification tasks, deeper models such as ResNet and EfficientNet are most appropriate, whereas lighter models like MobileNet are better suited for edge devices.

2.2. The Fine-Tuning Process

The CXR images in the dataset were first converted from RGB to grayscale, and then their sizes were adjusted to match the input dimensions of each model. Scaling was required to guarantee compatibility with the neural network designs used in this investigation because the image dimensions varied throughout the dataset. All input images were reduced to a standard dimension of $224 \times 224 \times 3$ during the fine-tuning phase, except for the Inception model, which required an input size of $299 \times 299 \times 3$ to satisfy its architectural requirements. The dataset's images are all in JPEG (.jpg) format. Every transfer learning (TL) model that is being studied goes through several steps after this preprocessing stage:

- **Layer Modification:** To accommodate the new lung disease classification task, we substituted new layers for the previous three.

- **Fine-Tuning Strategy:** We used a fine-tuning strategy to modify the pre-trained models to fit our task. The entire model had to be unfrozen and retrained with a very low learning rate of $1e-4$. This method's justification is to progressively modify the pre-trained features to better suit the distinct qualities of our recently acquired image dataset.

By fine-tuning, we sought to enhance the models' performance in detecting pulmonary diseases while leveraging the benefits of TL. We expected significant gains in model performance for this task by utilizing pre-trained features and customizing them to our data set.

2.3. Blending Policies for the Voting Classifier

Machine learning solutions known as ensemble methods are designed to start with a collection of base models, or "base" or "weak" models, and work their way up to a more robust and effective model. When specific conditions are fulfilled, such as requiring diversity, an ensemble usually outperforms its base models. By combining several classifiers using a hard voting process, ensemble approaches improve a model's overall accuracy (ACC), as shown in Figure 3. Keep in mind that the technology used to create the base models is not generally constrained, so homogeneity is not necessary.

When it comes to voting methods, there are two popular schemes. The two primary blending techniques recognized by the study's community are voting and stacking. The voting indicates that the two primary voting policies—hard and soft voting—are viable. The voting ensemble strategy, which offers models with a high capacity for generalization, is the most favored ensemble approach when using the combination method. Several base classifier outputs must be combined to produce the actual classification given an instance that needs to be classified. The following is a schematic illustration of this idea. An illustration of the hard voting process is provided in Figure 3.

One of the following labels is applied to each base classifier output: opacity, bacterial pneumonia, TB, COVID-19, normal, and viral bacterial lungs. Because majority voting is in place, the class with the most votes from the base models is chosen. The class with the most votes is predicted by hard voting using base models. The voting policy produces the class named Bacterial Pneumonia, which is predicted by three of the five models, as illustrated in the schematic representation example in Figure 3. EfficientNetB0, DenseNet201, ResNet50, Mobilenetv2, and InceptionV3net are the five pre-trained models whose features were fused using the ensemble model. With this method, each classifier's output is evaluated, and the class with the most votes is chosen.

Using just one classifier might not be sufficient because each one operates independently and may extract different information from the data. Ensemble methods increase overall accuracy and lessen the drawbacks of using a single classifier by combining the output from several classifiers and basing the final decision on the majority vote.

Hard voting was preferred over stacking and soft voting for both technical and practical reasons. Compared with soft voting, which aggregates expected probability, hard voting is less prone to calibration discrepancies across heterogeneous architectures and is simpler to implement consistently. Moreover, while stacking works well, it introduces complexity by requiring an additional meta-learning layer, which may be overfit in situations where training data is a limited issue that frequently occurs in medical imaging. Particularly in a classification with multiple classes context like lung disease detection, the hard voting ensemble improves accuracy, resilience, and generalization by maintaining model independence and combining predictions employing a straightforward majority rule.

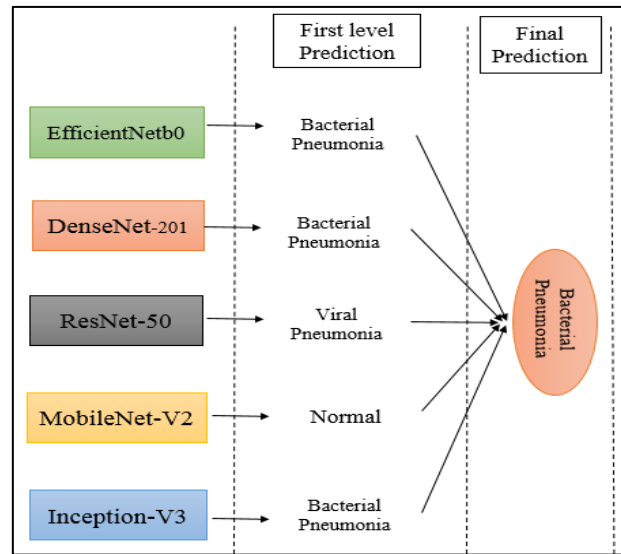


Figure 3. Hard voting classifier example.

3. RESULT AND DISCUSSION

This section examines the suggested method's robustness, considering the experiments that were carried out. The following describes the used dataset and the analysis of the results.

3.1. Dataset

Finding a suitable dataset is one of the requirements for image recognition and computer vision research. One of the most important prerequisites for deep learning is a vast quantity of training images. The dataset of CXR images used in our study was collected from four publicly accessible data repositories that are accessible on the most recent "Mendeley" website. The data set used in the recommended model is shown in Table 3 below. Figure 4 depicts a sample of this dataset, which is explained in the following points.

1. Tuberculosis CXR Image Dataset [47]: This dataset features 514 normal CXR images and 2494 CXR images of TB patients gathered from a nearby Pakistani hospital. To maintain uniformity and make additional analysis easier, every image has been processed and resized to a consistent size.
2. COVID-19 Curated Dataset for X-Rays of the Posterior-Anterior Chest [48]: The current dataset includes 3270 normal X-rays, 1656 viral pneumonia X-rays, 3001 bacterial pneumonia X-rays, and 1281 COVID-19 X-rays.
3. CXR Image (Lung Opacity) [49]: It provides 4,350 excellent images over four different classes: Normal X-ray images of healthy lungs are included in this category (class distribution: 1200 images) and are used as a benchmark for comparing them to lungs that are diseased or infected. Class Distribution: Lung Opacity: These pictures show cases where there are abnormalities or lung opacities that need more investigation and diagnosis. COVID (Class Distribution: 1050 Images): This class helps with early disease detection and monitoring by providing X-ray images of patients who have confirmed or suspected COVID-19 cases. The X-ray images in this category are linked to cases of viral pneumonia, which helps with the identification and comprehension of this kind of lung infection (Class Distribution: 1000 Images).

To ensure image quality and relevancy, the dataset was extensively filtered using the curation process outlined in [48]. Noise, compression artifacts, pixelation, washed-out visuals, the presence of medical implants, side views, aspect ratio distortion, cropped or magnified images, CT (sliced) images, rotated scans, and images with annotations were among the issues we identified and removed. Duplicate photos were also eliminated to preserve data integrity and prevent model bias. This refinement process yields a higher-quality, cleaner dataset that is suitable for deep learning-based classification tasks in medical imaging.

Table 3. Distribution of chest X-ray images across various categories

Category	Number of Images
COVID-19	1281
Lung_Opacity	1200
Normal	3270
Pneumonia-Bacterial	3001
Pneumonia-Viral	1656
TB Chest X-rays	2494

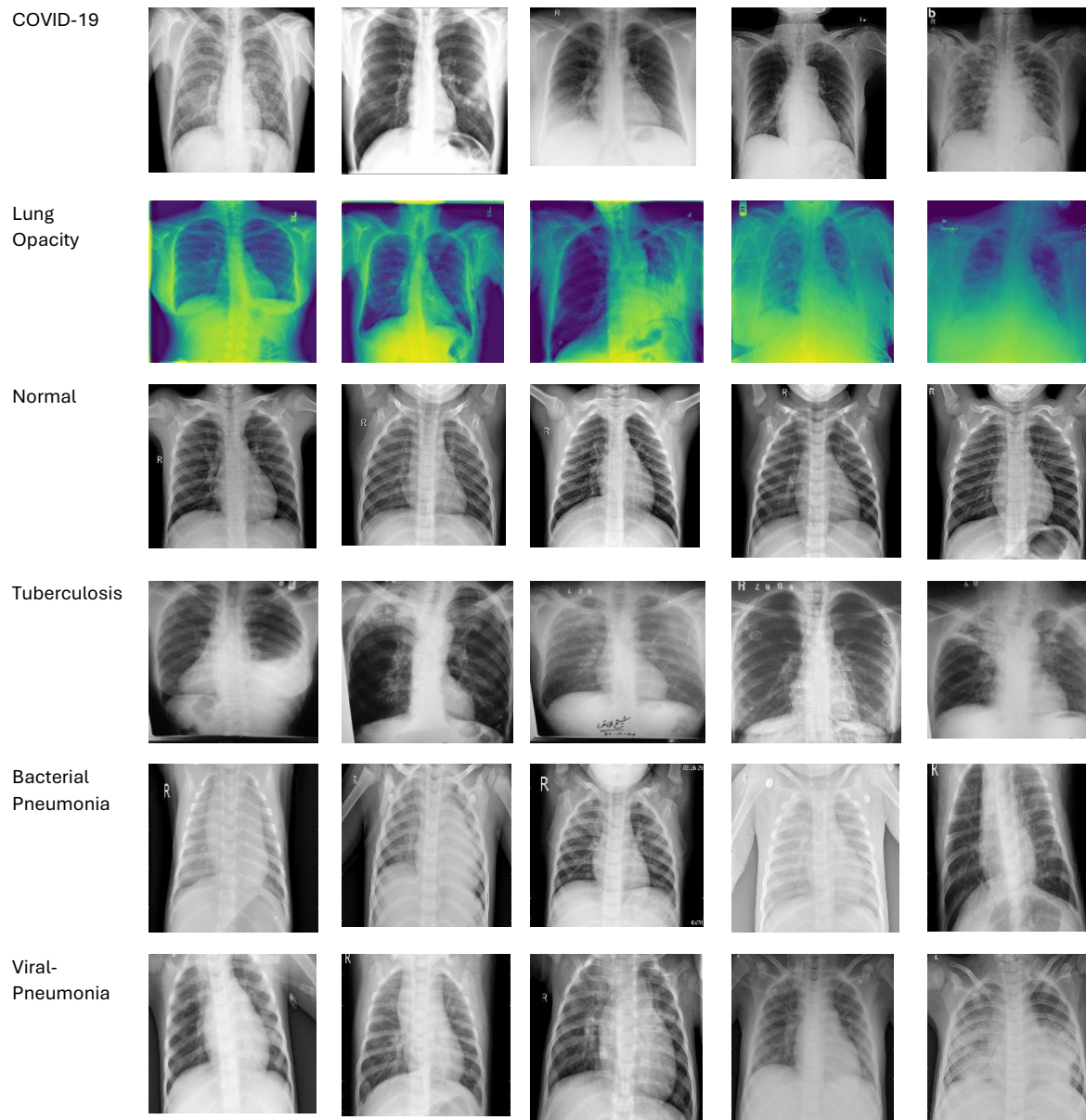


Figure 4. Examples of lung diseases CXR images

3.2. Data Splitting and Measuring Metrics

Based on the experimental findings, we assess the suggested method's efficacy in this section. The previously described datasets were used in the experiments. The datasets were divided into training and testing sets using a standard data-splitting methodology, following the distribution percentages listed below:

- Training Set: The DL model was trained using 80% of the dataset. This section served to speed up the learning process and make it possible for the model to identify trends in the data.
- Testing Set: The testing subset was the final 20% of the dataset. The model's final performance was assessed solely using this subset, which revealed information about the model's capacity for generalization and prediction accuracy on data that had not yet been seen.

We sought to strike the ideal balance between model training, parameter tuning, and a thorough assessment of the model's performance on unseen data by splitting the dataset into these two separate subsets with the designated proportions. Table 4 below provides a summary of the technical specifications and implementation specifics:

Table 4. Technical specifications and implementation.

System Description	Details
Programming Language	MATLAB R2021a
Operating System	Windows 10
Processor	Intel Core i5
RAM	6 GB
Optimizer	Adam
Learning Rate	1e-4
Number of Epochs	5
Minibatch Size	20

To assess the usefulness of the suggested approach, we utilized popular evaluation metrics: accuracy, recall, precision, and F1-score are computed according to the following equations (1-4).

$$\text{accuracy} = \frac{\text{number of TN} + \text{number of TP}}{\text{number of TN} + \text{number of TP} + \text{number of FP} + \text{number of FN}} \quad (1)$$

$$\text{recall} = \frac{\text{number of TP}}{\text{number of FN} + \text{number of TP}} \quad (2)$$

$$\text{precision} = \frac{\text{number of TP}}{\text{number of FP} + \text{number of TP}} \quad (3)$$

$$\text{F1 - score} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (4)$$

In this case, the correctly identified positive class is denoted by TP (True Positive), the correctly expected negative class by TN (True Negative), the incorrectly estimated positive class by FP (False Positive), and the incorrectly estimated negative class by FN (False Negative). As a broad measure of the model's performance, accuracy is the percentage of accurate predictions—both true positives and true negatives—out of all the predictions. By calculating the percentage of predicted positives that are accurate, precision—also referred to as positive predictive value—measures the accuracy of positive predictions. Recall, also known as sensitivity or the true positive rate, measures how well a model can identify every positive instance in a class and shows how many of the real positives were identified. The F1-score is the harmonic means of precision and recall, offering a balanced evaluation of both metrics. For each class, the F1-score is calculated using its respective precision and recall and can be averaged across all classes using macro-averaging, which computes the simple average of per-class F1-scores. Similarly, accuracy, recall, and precision are calculated for each class individually, and their averages across all classes are subsequently computed, as the following equations (5-8)

$$\text{PREC} = \frac{\sum_{k=1}^k \text{PREC}_k}{k} \quad (5)$$

$$\text{SENS} = \frac{\sum_{k=1}^k \text{REC}_k}{k} \quad (6)$$

$$\text{ACC} = \frac{\sum_{k=1}^k \text{ACC}_k}{k} \quad (7)$$

$$\text{F1} = 2 \cdot \frac{\text{PREC} \cdot \text{REC}}{\text{PREC} + \text{REC}} \quad (8)$$

Where PREC_k is the precision for the k -th class, REC_k is the recall for the k -th class, ACC_k is the balanced accuracy for the k -th class.

3.3. Result Analysis

The effectiveness of the suggested system is assessed using accuracy, a well-known evaluation metric. The ResNet50 model outperforms other models like EfficientNetB0, DenseNet-201, Mobilenetv2, and InceptionV3net in identifying differences in lung abnormalities, as evidenced by its remarkable accuracy of 93.50% when compared to the base models. The accuracy rates of other models are also good: EfficientNetB0 achieves 92.8% accuracy, mobilenetv2 and inceptionV3 achieve 92.7% and 92.10%, respectively, and DenseNet-201 achieves a high accuracy of 93.02%. All these accuracy values highlight how well the models can produce accurate predictions and highlight how useful they could be for the efficient diagnosis of lung conditions.

Table 5 shows that the suggested model achieves roughly 97.59%, 96.20%, 96.89% accuracy, precision, recall, and F1-score measures using the given X-ray dataset. There are notable variations in accuracy, precision, recall, and F1 score between the performance of the suggested ensemble-based model and the corresponding monolithic model. The accuracy for different base model combinations is shown in Table 6. Additionally, Table 7 highlights how well the suggested method separates the two related classes of bacterial and viral pneumonia. The accuracy of these two frequently perplexing classes has significantly increased thanks to the ensemble-based approach.

Table 5. Accuracy, Precision, Recall, and F1-Score for base models and ensemble models.

#	Model	Accuracy	Precision	Recall	F1-Score
1	EfficientNetB0 [12]	92.8%	92.79%	92.65%	92.72%
2	DenseNet-201 [40]	93.02%	93.64%	91.91%	92.77%
3	ResNet50 [44]	93.50%	93.66%	93.01%	93.33%
4	Mobilenet-v2 [46]	92.70%	92.68%	92.68%	92.68%
5	Inception-v3 [43]	92.10%	92.08%	91.52%	91.80%
6	Proposed Voting-Based Method	96.90%	97.59%	96.20%	96.89%

Table 6. Different combinations of base models.

Model	Accuracy
EfficientNetB0+ DenseNet-201+ ResNet50 +mobilenetv2	95.1%
EfficientNetB0+ DenseNet-201+ ResNet50 +inceptionv3	95.23%
EfficientNetB0+ ResNet50 +mobilenet-v2 +inception-v3	96.5%
DenseNet-201+ ResNet50 +mobilenet-v2 +inception-v3	96.1%
DenseNet-201+ ResNet50 +mobilenet-v2 +inception-v3	95.2%
EfficientNetB0+ DenseNet-201+ ResNet50 +mobilenet-v2 +inception-v3 (proposed method)	96.90%

Table 7. Accuracy of Viral and Bacterial Pneumonia for base models and ensemble-based models.

Model	Viral Pneumonia Accuracy	Bacterial Pneumonia Accuracy
EfficientNetB0 [12], [50]	76.7%	85.6%
DenseNet-201 [40], [51]	84.3%	84.1%
ResNet50 [44]	80.4%	85.8%
Mobilenet-v2 [46], [52]	72.4%	88.1%
Inception-v3 [43]	74.7%	85.7%
Proposed voting-based method	96.3%	91%

To demonstrate the robustness of our proposed model, the dataset was split into five folds: Fold1, Fold2, Fold3, Fold4, and Fold5. We employed a five-fold cross-validation approach, conducting five separate experiments. In each experiment, four folds were used for training, while the remaining fold served as the test set. Table 8 presents the cross-validation results obtained from applying our model to the aforementioned lung disease datasets. Figure 5 illustrates the accuracy and loss curves for the proposed model, indicating that the ensemble classifier enhances the performance of the individual models by approximately 4% through a hard voting strategy.

Table 8. Results of 5-fold cross-validation of the proposed method on the lung datasets

Fold index	DensNet-201 Accuracy	Inception-v3 Accuracy	MobileNet-v2 Accuracy	EfficientNetB0 Accuracy	ResNet50 Accuracy	Proposed voting method Accuracy
Fold1	91.75%	92.87%	92.41%	91.79%	92.1%	95.2%
Fold2	92.8%	92.7%	92.3%	93.03%	92.95%	95.92%
Fold3	92.41%	93.96	93.22%	94.1%	93.49%	96.2%
Fold4	92.83%	93.37%	91.63%	92.75%	92.72%	95.9%
Fold5	92.64%	92.11%	92.50%	92.72%	91.5%	95.07%
average	92.49%	93.00%	92.43%	92.89%	92.55%	95.7%

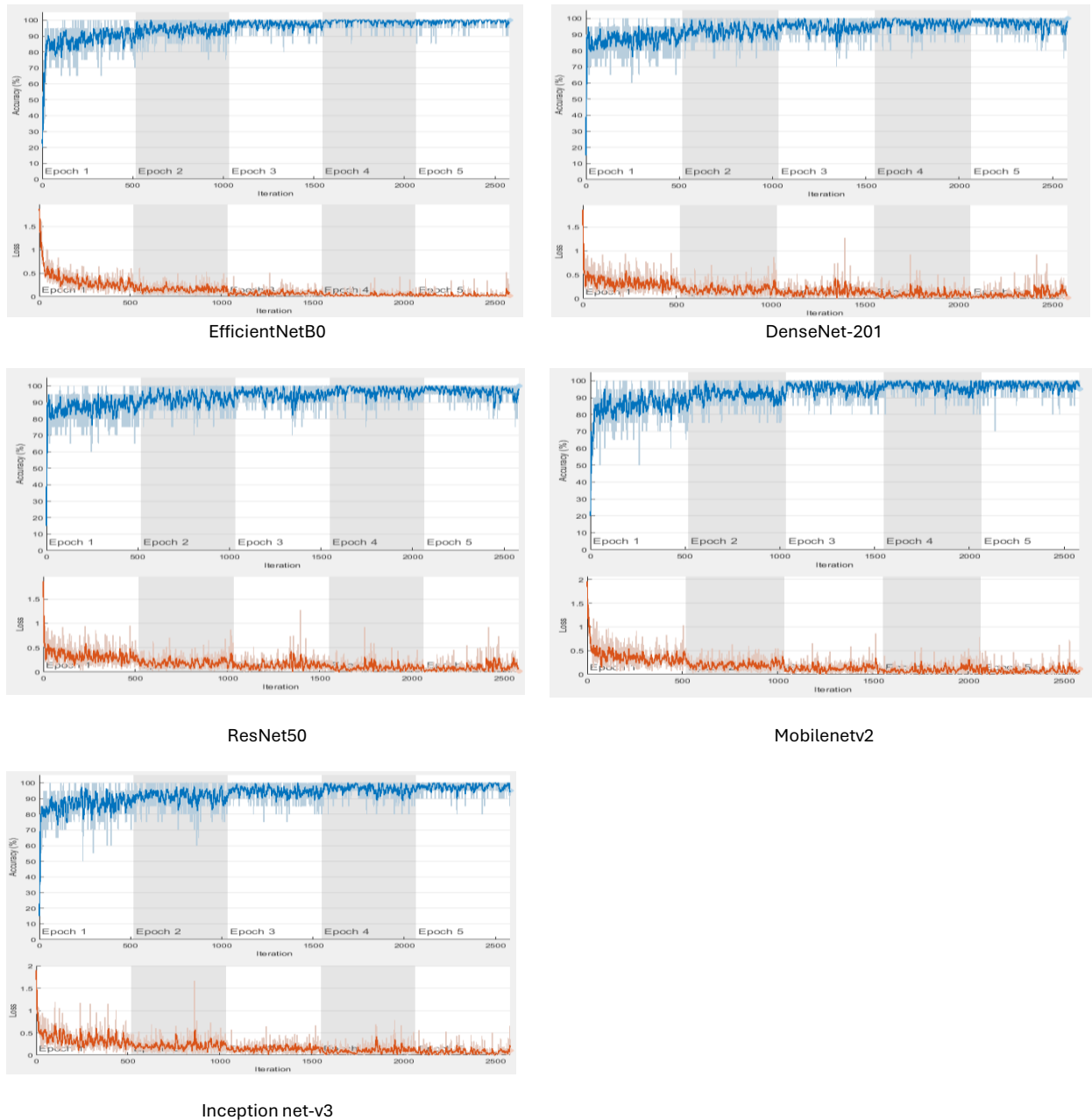


Figure 5. The Accuracy and loss curves for the proposed model.

3.4. Confusion Matrix Analysis

The confusion matrix for the base models using the X-ray dataset is shown in Figure 6. The ensemble model's confusion matrix is shown in Figure 7. The suggested methodology and its potential for in-depth classification are explained by the confusion matrix. The confusion matrices of the base model demonstrate that it is capable of accurately identifying opacity, COVID-19, tuberculosis, and normal diseases. Because of their similarities, bacterial and viral pneumonia are the least accurate. The suggested method can successfully differentiate between bacterial and viral pneumonia, as well as other illnesses like COVID-19, tuberculosis, and opacity, based on the ensemble-based model confusion matrix.

EfficientNetB0						
COVID-19	254 9.8%	0 0.0%	0 0.0%	2 0.1%	2 0.1%	1 0.0%
Lung _O capacity	0 0.0%	240 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
Normal	0 0.0%	0 0.0%	648 25.1%	11 0.4%	12 0.5%	0 0.0%
Pneumonia-Bacterial	0 0.0%	0 0.0%	2 0.1%	522 20.2%	86 3.3%	0 0.0%
Pneumonia-Viral	1 0.0%	0 0.0%	4 0.2%	65 2.5%	231 9.0%	0 0.0%
TB Chest X-rays	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	498 19.3%
DenseNet-201						
COVID-19	247 9.6%	0 0.0%	0 0.0%	1 0.0%	1 0.0%	1 0.0%
Lung _O capacity	1 0.0%	240 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
Normal	2 0.1%	0 0.0%	651 25.2%	7 0.3%	19 0.7%	0 0.0%
Pneumonia-Bacterial	1 0.0%	0 0.0%	1 0.0%	556 21.6%	102 4.0%	1 0.0%
Pneumonia-Viral	1 0.0%	0 0.0%	2 0.1%	36 1.4%	209 8.1%	0 0.0%
TB Chest X-rays	4 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	497 19.3%
ResNet50						
COVID-19	246 9.5%	0 0.0%	0 0.0%	0 0.0%	2 0.1%	2 0.1%
Lung _O capacity	2 0.1%	240 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
Normal	3 0.1%	0 0.0%	647 25.1%	10 0.4%	20 0.8%	0 0.0%
Pneumonia-Bacterial	1 0.0%	0 0.0%	1 0.0%	521 20.2%	85 3.3%	0 0.0%
Pneumonia-Viral	1 0.0%	0 0.0%	6 0.2%	69 2.7%	224 8.7%	0 0.0%
TB Chest X-rays	3 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	497 19.3%
Mobilenetv2						
COVID-19	252 9.8%	0 0.0%	0 0.0%	1 0.0%	1 0.0%	1 0.0%
Lung _O capacity	0 0.0%	240 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
Normal	0 0.0%	0 0.0%	652 25.3%	10 0.4%	9 0.3%	0 0.0%
Pneumonia-Bacterial	0 0.0%	0 0.0%	0 0.0%	497 19.3%	67 2.6%	0 0.0%
Pneumonia-Viral	2 0.1%	0 0.0%	2 0.1%	92 3.6%	254 9.8%	1 0.0%
TB Chest X-rays	2 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	497 19.3%

Inception net v3

Figure 6. Confusion matrix for the base models.

Confusion Matrix								
Output Class	COVID-19	255 9.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Lung_Opacity	0 0.0%	240 9.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Normal	0 0.0%	0 0.0%	653 25.3%	1 0.0%	10 0.4%	0 0.0%	98.3% 1.7%
	Pneumonia-Bacterial	0 0.0%	0 0.0%	0 0.0%	590 22.9%	58 2.2%	0 0.0%	91.0% 9.0%
	Pneumonia-Viral	0 0.0%	0 0.0%	1 0.0%	9 0.3%	263 10.2%	0 0.0%	96.3% 3.7%
	TB Chest X-rays	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	499 19.3%	99.8% 0.2%
	99.6% 0.4%	100% 0.0%	99.8% 0.2%	98.3% 1.7%	79.5% 20.5%	100% 0.0%	96.9% 3.1%	
	Target Class	COVID-19	Lung_Opacity	Normal	Pneumonia-Bacterial	Pneumonia-Viral	TB Chest X-rays	

Figure 7. Ensemble-based model confusion matrix.

3.5. Comparative Analysis

The experimental results were also compared with those published in related works found in the literature. Firstly, the modified ResNet50, as described in [25], is applied to the proposed dataset, and the results are reported in Figure 8. Moreover, the methods in [50], [51], [52], [53] were also applied to the aforementioned dataset, and the results were recorded. The comparison highlights that the proposed method demonstrated better performance.

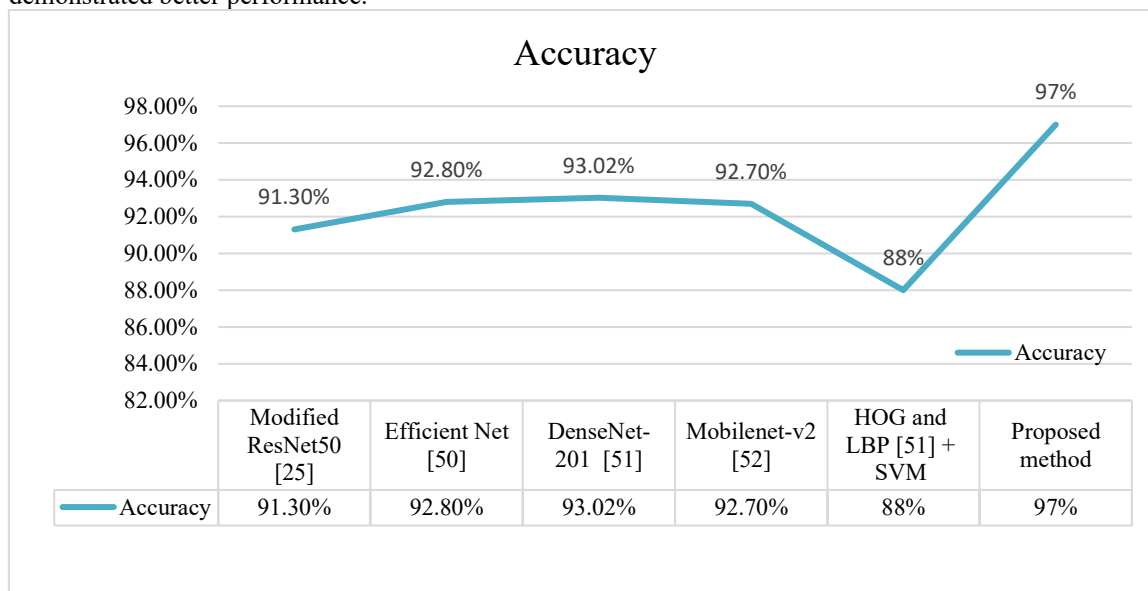


Figure 8. Performance comparison with the state of the artwork.

Our system's reported 97% accuracy in Figure 8 indicates that it has a lot of potential to help with the clinical diagnosis of lung diseases. While achieving high levels of accuracy is encouraging, it's crucial to examine the factors that lead to this success. Numerous recent studies routinely ignore important details like dataset size, class distribution, and the relative efficacy of different model designs. To close these gaps, this paper provides a comprehensive evaluation of several state-of-the-art CNN architectures, including MobileNetV2, InceptionV3, DenseNet201, ResNet50, and EfficientNetB0. Class imbalance, inter-class similarity, and picture acquisition variability are several issues that are highly relevant in real-world medical imaging scenarios, and the proposed method systematically examines how each architecture handles them. The main innovation in this research is the careful choice and comparative assessment of CNN models intended for multi-class lung disease detection. As a result, the benefits and drawbacks of the models are better understood, which aids in directing more thoughtful model deployment in clinical settings.

5. CONCLUSION

We developed a set of CNN models in this study to accurately predict different lung conditions from X-ray images. It incorporates the confidence scores of five popular DL architectures: MobileNetV2, InceptionV3, DenseNet201, ResNet50, and EfficientNetB0, which are fine-tuned to align with the study objectives. The Deep features extracted from each of these models were then fed into a hard-voting ensemble to boost classification accuracy. The method is designed to classify six categories: opacity, bacterial pneumonia, viral pneumonia, COVID-19, tuberculosis, and normal.

Extensive experiments demonstrated that the proposed ensemble outperformed current state-of-the-art methods, achieving 97%, 97.59%, 96.20%, and 96.89% in accuracy (ACC), precision (PREC), recall (REC), and F1-score (F1-S) values, respectively.

Comparative analyses demonstrated that the proposed method outperforms other popular deep transfer learning models, confirming its high diagnostic capabilities. The ensemble approach increases accuracy, generalization, and resilience by utilizing the many benefits of individual designs. This makes the proposed approach a more reliable and efficient solution for real-world medical imaging applications.

Data availability The authors confirm that the data supporting the findings are available from the corresponding author, upon reasonable request.

Conflict of interest The authors declare no conflict of interest.

REFERENCES

- [1] American Lung Association, "What Is Pneumonia & Who Is at Risk? | American Lung Association." 2025. [Online]. Available: <https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia/learn-about-pneumonia>
- [2] A. Brihn *et al.*, "Diagnostic Performance of an Antigen Test with RT-PCR for the Detection of SARS-CoV-2 in a Hospital Setting — Los Angeles County, California, June–August 2020," *MMWR. Morbidity and Mortality Weekly Report*, vol. 70, no. 19, pp. 702–706, 2021, doi: 10.15585/mmwr.mm7019a3.
- [3] R. Zhang *et al.*, "CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city," *European Radiology*, vol. 30, no. 8, pp. 4417–4426, 2020, doi: 10.1007/s00330-020-06854-1.
- [4] H. X. Bai *et al.*, "Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT," *Radiology*, vol. 296, no. 2, pp. E46–E54, 2020, doi: 10.1148/radiol.2020200823.
- [5] "A survey on deep learning in medical image analysis - ScienceDirect." [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1361841517301135>
- [6] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, 2019, doi: 10.1016/j.compbiomed.2019.103345.
- [7] M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Three-Class Mammogram Classification Based on Descriptive CNN Features," *BioMed Research International*, vol. 2017, 2017, doi: 10.1155/2017/3640901.
- [8] T. Rahman, M. E. H. Chowdhury, and A. Khandakar, "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using," *MDPI, J. app Sci.*, vol. 3233, pp. 1–17, 2020.
- [9] W. Alakwaa, M. Nassef, and A. Badr, "Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)," *International Journal of Biology and Biomedical Engineering*, vol. 11, no. 8, pp. 66–73, 2017, doi: 10.14569/ijacsa.2017.080853.
- [10] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 49–56, 2017, doi: 10.1016/j.cmpb.2016.10.007.
- [11] P. Silva *et al.*, "COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis," *Informatics in Medicine Unlocked*, vol. 20, 2020, doi: 10.1016/j.imu.2020.100427.
- [12] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.
- [13] and M. E. K. E. E. Hemdan, M. A. Shouman, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," *arXiv preprint arXiv:2003.11055*, 2020, doi: 10.48550/arXiv.2003.11055.
- [14] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and N. Rezoana, "ECOVNet: An Ensemble of Deep Convolutional Neural Networks Based on EfficientNet to Detect COVID-19 From Chest X-rays," 2020, doi: 10.7717/peerj-cs.551.
- [15] M. E. H. Chowdhury *et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [16] B. Nigam, A. Nigam, R. Jain, S. Dodia, N. Arora, and B. Annappa, "COVID-19: Automatic detection from X-ray images by utilizing deep learning methods," *Expert Systems with Applications*, vol. 176, no. July 2020, p. 114883, 2021, doi: 10.1016/j.eswa.2021.114883.
- [17] A. Manconi, G. Armano, M. Gnocchi, and L. Milanese, "A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19," *Applied Sciences (Switzerland)*, vol. 12, no. 15, 2022, doi: 10.3390/app12157554.
- [18] L. Menshawy, A. H. Eid, and R. F. Abdel-Kader, "Ensemble deep models for covid-19 pandemic classification using chest x-ray images via different fusion techniques," *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 1, pp. 51–65, 2023, doi: 10.26555/ijain.v9i1.922.
- [19] S. Sharma and K. Guleria, "Multiclass lung disease classification with ensembled CNN and LSTM modeling," *AIP Conference Proceedings*, vol. 3209, no. 1, 2024, doi: 10.1063/5.0229036.
- [20] M. H. Al-Sheikh, O. Al Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim, "Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images," *Scientific Reports*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.1038/s41598-023-46147-3.
- [21] J. Devasia, H. Goswami, S. Lakshminarayanan, M. Rajaram, and S. Adithan, "Deep learning classification of active tuberculosis lung zones wise manifestations using chest X-rays: a multi label approach," *Scientific Reports*, vol. 13, no. 1, pp. 1–15, 2023, doi: 10.1038/s41598-023-28079-0.
- [22] A. R. T. Abdulahi, R. O. Ogundokun, A. R. Adenike, M. A. Shah, and Y. K. Ahmed, "PulmoNet: a novel deep learning based pulmonary diseases detection model," *BMC Medical Imaging*, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12880-024-01227-2.
- [23] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100391, 2020, doi: 10.1016/j.imu.2020.100391.
- [24] Abdelwhab Ouahab, "Multimodal Convolutional Neural Networks for Detection of Covid-19 Using Chest X-Ray and CT Images," *Optical Memory and Neural Networks (Information Optics)*, vol. 30, no. 4, pp. 276–283, 2021, doi: 10.3103/S1060992X21040044.
- [25] M. Elpeltagy and H. Sallam, "Automatic prediction of COVID-19 from chest images using modified ResNet50," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26451–26463, 2021, doi: 10.1007/s11042-021-10783-6.
- [26] L. T. Duong, P. T. Nguyen, L. Iovino, and M. Flammini, "Automatic detection of Covid-19 from chest X-ray

- and lung computed tomography images using deep neural networks and transfer learning,” *Applied Soft Computing*, vol. 132, p. 109851, 2023, doi: 10.1016/j.asoc.2022.109851.
- [27] A. Shoeibi *et al.*, “Automated detection and forecasting of COVID-19 using deep learning techniques: A review,” *Neurocomputing*, vol. 577, 2024, doi: 10.1016/j.neucom.2024.127317.
- [28] A. Koul, R. K. Bawa, and Y. Kumar, “An Analysis of Deep Transfer Learning-Based Approaches for Prediction and Prognosis of Multiple Respiratory Diseases Using Pulmonary Images,” *Archives of Computational Methods in Engineering*, vol. 31, no. 2, pp. 1023–1049, 2024, doi: 10.1007/s11831-023-10006-1.
- [29] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, “Empowering COVID-19 detection: Optimizing performance through fine-tuned EfficientNet deep learning architecture,” *Computers in Biology and Medicine*, vol. 168, no. August 2023, p. 107789, 2024, doi: 10.1016/j.compbiomed.2023.107789.
- [30] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, “Efficient pneumonia detection using Vision Transformers on chest X-rays,” *Scientific Reports*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-52703-2.
- [31] H. Malik and T. Anees, *Multi-modal deep learning methods for classification of chest diseases using different medical imaging and cough sounds*, vol. 19, no. 3 March. 2024, doi: 10.1371/journal.pone.0296352.
- [32] K. Morani, E. K. Ayana, and D. Unay, “Covid-19 detection using modified xception transfer learning approach from computed tomography images,” *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 3, pp. 524–536, 2023, doi: 10.26555/ijain.v9i3.1432.
- [33] M. A. Talib, Y. Afadar, Q. Nasir, A. B. Nassif, H. Hijazi, and A. Hasasneh, “A tree-based explainable AI model for early detection of Covid-19 using physiological data,” *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12911-024-02576-2.
- [34] N. Kumar, M. Gupta, D. Gupta, and S. Tiwari, “Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 1, pp. 469–478, 2023, doi: 10.1007/s12652-021-03306-6.
- [35] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, “Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks,” *Applied Sciences (Switzerland)*, vol. 12, no. 13, pp. 1–10, 2022, doi: 10.3390/app12136448.
- [36] R. R. Nair and T. Singh, “Exploring Ensemble Architectures for Lung X-Ray Multi-Class Image Classification using CNN-LSTM,” *Procedia Computer Science*, vol. 258, pp. 852–861, 2025, doi: 10.1016/j.procs.2025.04.317.
- [37] M. Patel, M. Shah, and P. & Shah, “Deep Ensemble Learning Model for Diagnosis of Lung Diseases from Chest X-Ray Images,” *Indian Journal of Science and Technology*, vol. 17, no. 8, pp. 702–712, 2024, [Online]. Available: <https://www.indjst.org/>
- [38] M. Pal, S. Parija, and G. Panda, “An effective ensemble approach for classification of chest X-ray images having symptoms of COVID: A precautionary measure for the COVID-19 subvariants,” *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, no. March, p. 100547, 2024, doi: 10.1016/j.prime.2024.100547.
- [39] H. Nakrani, E. Q. Shahra, S. Basurra, R. Mohammad, E. Vakaj, and W. A. Jabbar, “Advanced Diagnosis of Cardiac and Respiratory Diseases from Chest X-Ray Imagery Using Deep Learning Ensembles,” *Journal of Sensor and Actuator Networks*, vol. 14, no. 2, pp. 1–17, 2025, doi: 10.3390/jsan14020044.
- [40] G. Huang, Z. Liu, and L. van der Maaten, “Densely Connected Convolutional Networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 39, no. 9, pp. 1442–1446, 1978.
- [41] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 248–255, 2009, doi: 10.1109/CVPR.2009.5206848.
- [42] D. A. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, “Going deeper with convolutions | IEEE Conference Publication | IEEE Xplore.” doi: 10.1109/CVPR.2015.7298594.
- [43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.
- [44] A. Sai Bharadwaj Reddy and D. Sujitha Juliet, “Transfer learning with RESNET-50 for malaria cell-image classification,” *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, pp. 945–949, 2019, doi: 10.1109/ICCSP.2019.8697909.
- [45] M. Tan *et al.*, “Mnasnet: Platform-aware neural architecture search for mobile,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 2815–2823, 2019, doi: 10.1109/CVPR.2019.00293.
- [46] H. Huo, Y. L. Yu, and Z. H. Liu, “Facial expression recognition based on improved depthwise separable convolutional network,” *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18635–18652, 2023, doi: 10.1007/s11042-022-14066-6.
- [47] S. Kiran and D. I. Jabeen, “Dataset of Tuberculosis Chest X-rays Images,” *Mendeley Data*, vol. V2, 2024, doi: 10.17632/8j2g3csprk.2.
- [48] L. K. P. S. B. R. K. T. S. S. B. K. Sait U, “Curated dataset for COVID-19 posterior-anterior chest radiography images (X-Rays),” *Mendeley Data.*, no. September, pp. 9–10, 2020, doi: 10.17632/9xkhgts2s6.3.
- [49] M. A. Talukder, “Chest X-Ray Image - Mendeley Data,” *V1*, 2023, doi: 10.17632/m4s2jn3csb.1.

- [50] R. Raza *et al.*, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Engineering Applications of Artificial Intelligence*, vol. 126. 2023. doi: 10.1016/j.engappai.2023.106902.
- [51] M. Bunde and G. M. Danciu, "Pneumonia Image Classification Using DenseNet Architecture," *Information (Switzerland)*, vol. 15, no. 10, 2024, doi: 10.3390/info15100611.
- [52] N. Das R. Ghosh, A. Ahamed, B. Sadhukhan, "Lung nodule classification using MobileNet Transfer Learning," *In 2023 9th International conference on smart computing and communications (ICSCC). IEEE*, pp. 290–295, 2023, doi: 10.1109/ICSCC59169.2023.10335043
- [53] S. Kaur *et al.*, "High-accuracy lung disease classification via logistic regression and advanced feature extraction techniques," *Egyptian Informatics Journal*, vol. 29, no. October 2024, pp. 1–11, 2025, doi: 10.1016/j.eij.2024.100596.

BIOGRAPHY OF AUTHORS



Marwa Elpeltagy    received B.Sc. and M. Sc. Degrees from Azhar University, Egypt, in 2010 and 2015, respectively, all in Systems and Computers Engineering. She received her Ph.D. degree from the Egypt-Japan University of Science and Technology, Electronics and Communication Engineering, Egypt, in August. 2018. She joined Kyushu University, Fukuoka, Japan, as a special research student as a part of her Ph.D. program funded by the Ministry of Higher Education in Egypt (MOHE) in September 2017.



Ebeid Ali Ebeid    obtained his B.Sc. degree in computer science from the Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt, in 2002, and his M.Sc. degree from the same faculty in 2015. He was involved as a teaching assistant in the Department of Mathematics, Faculty of Science, Al-Azhar University in 2012, and then as a lecturer assistant in 2016. He received his PhD degree in 2020, and he is currently an assistant professor of computer science at the same faculty. His research interests include biometrics, pattern recognition, computer vision, machine learning, and AI.



Farida Youness    received the B.Sc., M.Sc., and Ph.D. degrees from the Faculty of Engineering, Al-Azhar University, Cairo, Egypt. She is currently a lecturer with the Faculty of Engineering, Al-Azhar University, and previously was a Research Assistant with the National Research Institute. Her research interests include deep learning, machine learning, text summarization, natural language processing, artificial intelligence, and informatics.