Received: 29 October 2024 | Revised: 25 November 2025 | Accepted: 19 January 2026 | Published online: 11 March 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewJDSIS62024673

Multi-network–Based COVID-19 Detector Using Chest X-ray Images

Shubham[1] and Brijesh Kumar Chaurasia[2,*]

[1]*Department of Computer Science, Indian Institute of Information Technology, India*

[2]*Department of Computer Science & Engineering, Pranveer Singh Institute of Technology, India*

Abstract: The global outbreak of COVID-19 has emerged as the most catastrophic public health crisis in over five decades, exhibiting profound respiratory implications that primarily compromise pulmonary function. Timely and accurate detection of infection-induced anomalies within the lungs remains critical to curbing the spread and mitigating clinical severity. Among diagnostic imaging modalities, computed tomography (CT) and chest radiography (X-ray) have proven indispensable—offering varying degrees of sensitivity, resolution, and accessibility. To address these limitations, recent advancements in deep learning have introduced data-driven frameworks capable of autonomously learning discriminative features from medical imagery. In this study, we introduce a robust multi-network ensemble architecture tailored for automated multi-class pulmonary infection classification. The proposed framework integrates multiple heterogeneous convolutional neural networks (CNNs), namely, VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0, each initialized with pre-trained weights derived from large-scale image corpora via transfer learning. Experimental findings indicate that the ensemble model consistently outperforms its constituent networks, attaining a peak classification accuracy of 98.37%. This superior performance underscores the efficacy of multi-network integration in enhancing diagnostic reliability and enabling fine-grained discrimination across complex pathological classes.

Keywords: COVID-19, chest X-rays, CT scans, transfer learning, convolutional neural network

1. Introduction

The COVID-19 pandemic exposed significant gaps in global diagnostic infrastructure, particularly in the early stages when rapid, scalable testing solutions were critically needed [1]. Although molecular testing methods such as reverse transcription polymerase chain reaction (RT-PCR) are regarded as the clinical benchmark due to their specificity, they are limited by long processing times, specialized equipment requirements, and sensitivity to viral load, especially in asymptomatic or early-phase patients [2, 3]. These practical limitations motivated the healthcare community to explore auxiliary diagnostic tools that are faster and more accessible.

Chest radiography (CXR) and computed tomography (CT) scans emerged as useful imaging-based approaches for detecting COVID-19-induced pulmonary abnormalities [4]. Yet, the interpretation of these scans traditionally depends on trained radiologists, making the process time-consuming and susceptible to human error or variability, especially in high-pressure settings. To counteract these challenges, artificial intelligence (AI) techniques—particularly deep learning—have been introduced to automate medical image analysis. These models can detect subtle patterns in imaging data that may be overlooked by the human eye and have shown encouraging results in the detection of COVID-related lung infections [5–7]. However, a recurring limitation in most existing AI-based diagnostic systems is their narrow focus on binary classification—typically distinguishing between COVID-19 and non-COVID cases. Such approaches oversimplify clinical scenarios where multiple respiratory conditions can exhibit overlapping radiographic features. A more nuanced classification system is essential for differentiating COVID-19 from other lung pathologies like bacterial pneumonia, viral pneumonia, and other non-COVID abnormalities [8].

This study proposes a deep learning-based ensemble model that performs five-class classification of chest X-ray images, distinguishing among COVID-19, bacterial pneumonia, viral pneumonia, normal lungs, and nonspecific abnormal cases. The design aims to improve the diagnostic granularity needed for effective triage and clinical decision-making.

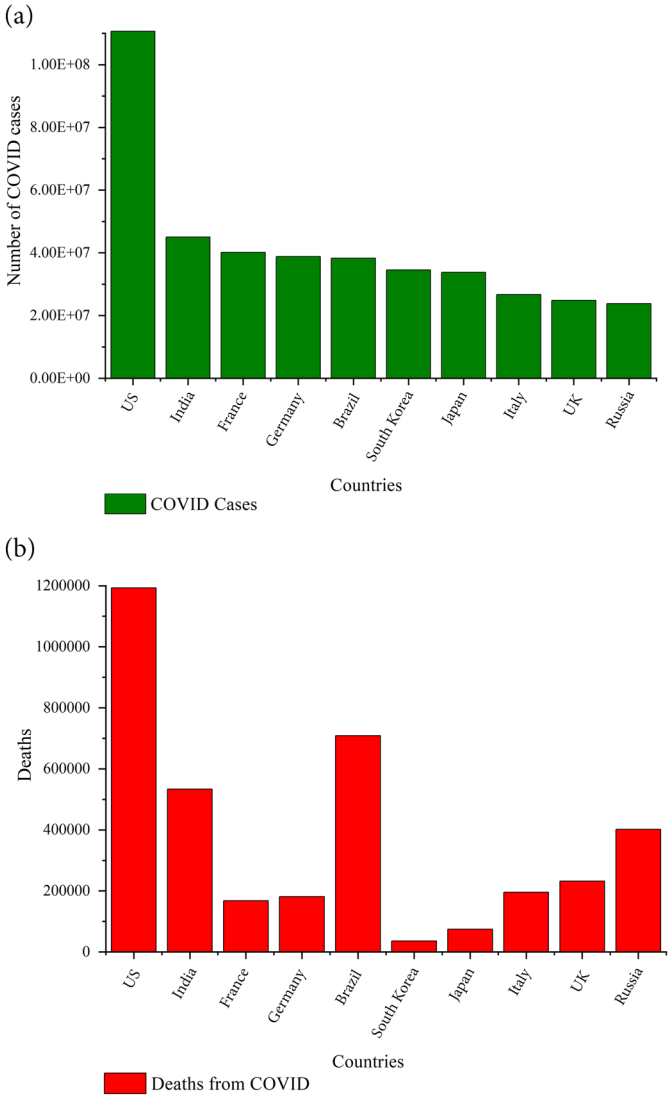
To accomplish this, we utilize a combination of five pre-trained convolutional neural networks (CNNs)—VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0—each adapted using transfer learning (TL). These models were trained on publicly available medical imaging datasets [9–11], and their predictions were fused using an ensemble approach. This architecture takes advantage of complementary feature extraction capabilities across models, resulting in improved classification accuracy. The proposed system achieves 98.37% accuracy, outperforming each individual CNN model.

In support of the problem's urgency, Figure 1(a)-(b) highlights the global COVID-19 burden, presenting country-wise infection and death statistics from March 2022. This epidemiological context illustrates the scale of the diagnostic demand and the need for AI-driven solutions in pandemic response efforts.

Several robust deep learning models based on recurrent neural networks (RNNs), such as long short-term memory (LSTM) networks and gated recurrent units, have been explored in medical imaging tasks—particularly for modeling temporal dependencies in image sequences (e.g., CT scan slices, video frames, or patient progression over time). In the context of chest X-rays, however, the input data typically consists of static two-dimensional images, where the primary challenge is spatial feature extraction rather than temporal pattern recognition.

*Corresponding author: Brijesh Kumar Chaurasia, Department of Computer Science & Engineering, Pranveer Singh Institute of Technology, India. Email: brijesh.chaurasia@psit.ac.in

Figure 1
(a) Number of COVID-19 cases in top 10 affected countries; (b) number of deaths due to COVID-19 in top 10 affected countries



While RNNs have been used in conjunction with CNNs in hybrid models for tasks like report generation or disease progression prediction, their application is less suitable for standalone image classification [12]. In contrast, CNNs are specifically designed to capture spatial hierarchies in visual data, making them the preferred architecture for static image analysis. Given that this study focuses on classifying individual CXR images into five disease categories, CNNs offer a more direct and effective approach [13].

The proposed method leverages a CNN-based ensemble to enhance feature diversity and classification performance, which would be difficult to replicate using RNNs alone in this setting.

1.1. Motivation

The outbreak of the COVID-19 pandemic underscored the critical need for fast diagnostic methods, particularly in the early stages of infection. Although RT-PCR remains the gold standard due to its high specificity, its limitations are significant. The process is not only time-consuming but also requires a highly skilled workforce, making it a challenging and resource-intensive procedure. Moreover, the sensitivity of RT-PCR can vary, leading to concerns over false negatives. Compounding this issue is the global shortage of trained radiologists, making it increasingly difficult for healthcare facilities to consistently provide expert-level diagnostic assessments. Consequently, there is a growing demand for alternative diagnostic solutions that can complement or expedite traditional methods, ensuring quicker and more widespread detection of COVID-19. Therefore, artificial intelligence models can help detect the virus at an early stage quickly and accurately. In our study, we proposed using deep neural networks for the detection of COVID-19 and similar diseases. Most studies focus on COVID-positive or COVID-negative cases. We proposed classifying chest X-ray images into five categories: COVID-19 positive/COVID-19, bacterial pneumonia, viral pneumonia, COVID-19 negative/normal, and not normal (neither pneumonia nor COVID-19), which impact the lungs. We have classified bacterial and viral pneumonia separately. Not normal chest X-rays are caused by some irregularities in the chest, which are not due to COVID-19 or pneumonia. To work on these different diseases, we combined three datasets.

1.2. Contributions

The main contributions of this paper are briefly described as follows:

1) An ensemble CNN framework for multi-class classification of CXR images into five clinically relevant categories
2) Evaluation using real-world CXR datasets from multiple sources, demonstrating strong performance across all major diagnostic metrics
3) Transfer learning to improve generalization and reduce training complexity, even with limited and imbalanced data [14, 15]
4) Comprehensive evaluation using accuracy, precision, recall, F1 score, and receiver operating characteristic (ROC)-area under the curve (AUC) on three public datasets, demonstrating significant improvement over baseline models

1.3. Structure of the paper

The structure of this article is organized as follows: Section 2 provides a comprehensive review of recent studies conducted by researchers on the diagnosis of COVID-19 infection, focusing on various machine learning and deep learning techniques. Section 3 outlines the proposed methodology, detailing the framework and approach used in this work. In Section 4, we present an in-depth discussion of the convolutional neural network (CNN) models used, along with the experimental results, including classification accuracy and other relevant performance metrics. Finally, Section 5 concludes the work by summarizing the key findings and implications of this study.

2. Related Work

After the outbreak of COVID-19, researchers have contributed a lot to the early detection of this infection by applying various state-of-the-art machine learning and deep learning techniques. Deep learning (DL) models are more popular as they provide more accurate results. Mohan et al. [16] developed the COVID-Net framework, which is one of the earliest research projects that applied CNN for COVID-19 infection detection. Multi-class classification is performed on chest X-ray images with classes, namely, COVID-19, non-COVID-19 infection (e.g., viral and bacterial), and no infection, and reported an accuracy of 96%.

Al-qaness et al. [17] claimed that the chest X-ray images are first pre-processed and then pre-trained weights of VGG16 are used to train the model. This is again used for multi-class classification and achieved an accuracy score of 92.3%. Islam and Tarique [18] based their study on transfer learning and applied two CNN models, including VGG16, for classification. It reported accuracy, precision, and recall scores of 100%,

100%, and 100%, respectively. COVID-19 detection by optimizers is presented in the study of Mzoughi et al. [19]. They have achieved up to 98% accuracy; however, datasets and complexity are issues. Dual deep learning techniques for COVID-19 detection are discussed [20]. It used CNN for feature extraction, and then LSTM was applied for classification. It reported an accuracy, AUC, specificity, sensitivity, and F1 score of 99.9%, 99.2%, 99.3%, and 98.9%, respectively. Khan et al. [21] introduced the CoroNet framework, which used the pre-trained weights of the Xception model. The study reported the multi-class classification results by considering different combinations of classes, like firstly normal vs. bacterial pneumonia vs. viral pneumonia vs. COVID-19 pneumonia with an accuracy score of 89.6% and secondly normal vs. COVID-19 pneumonia vs. non-COVID-19 pneumonia with an accuracy score of 95.0%. In order to classify chest X-ray images, Malik et al. [22] used the CDC Net that incorporates residual network thoughts and dilated convolution. It considered six classes, namely, COVID-19, pneumothorax, pneumonia, lung cancer, tuberculosis, and normal. It is also observed that the CDC Net model achieved accuracy up to 99.39%. Hasija et al. [23] presented a COVID-19 detection model for multi-class classification of chest X-ray images. In multi-class classification, it considered three classes, i.e., normal, COVID, and pneumonia, with an accuracy of 91–98%.

Minaee et al. [24] applied transfer learning and used four pre-trained models: DenseNet121, SqueezeNet, ResNet18, and ResNet50 for binary classification. It considered two classes, COVID and non-COVID, and obtained a sensitivity rate of 98%. The hybrid strategy described in the study of Ezzat et al. [25] used pre-trained DenseNet121 weights and a gravitational search optimization algorithm to find the optimal DenseNet121 hyperparameter values. Marques et al. [26] used the EfficientNet architecture for both binary and multi-class classification. The binary classification model achieved 99.6% accuracy in distinguishing COVID-19 from normal cases, while the multi-class model reached 96.7% accuracy in differentiating COVID-19, pneumonia, and normal cases. Das et al. [27] applied Truncated Inception Net to six datasets for binary classification with an accuracy score of 99.96%. Ayalew et al. [28] and Babukarthik et al. [29] detected the COVID-19 infection by training the deep learning model from scratch for binary classification with classes COVID and normal. A framework that can function in situations of data deficit was proposed by Bhosale and Patnaik [30]. With an F1 score of 85%, this system combines transfer learning and self-supervised learning for binary classification.

Few studies have considered the use of a 3D model for diagnosing COVID-19 infection. Kuzinkovas and Clement [31] presented a deep learning framework for early flood detection using satellite imagery. Their approach integrates a CNN-based feature extractor with a bidirectional LSTM to capture both spatial and temporal patterns, enabling more accurate identification of flood-prone regions. The model outperforms traditional machine learning baselines by leveraging multitemporal remote sensing data. Wang et al. [32] proposed a weakly supervised deep learning model for COVID-19 classification. It first segments the three-dimensional (3D) lung region with a pre-trained UNet, and then this segmented region is fed into the 3D neural network for classification. An attention-based 3D model [33] is proposed for classification into three classes: common pneumonia, non-pneumonia, and COVID-19. This proposed framework generates the 3D features with 3D convolution on the 3D CT scans. These generated 3D features, along with the bag model, are combined for a better classification model. The research article used in this paper is represented in Figure 2.

3. Proposed Methodology

TL [15] is a widely used DL technique trained on large-scale datasets. The learning capabilities gained during training can be transferred to other models for solving similar kinds of tasks [14].

Figure 2
Research paper inclusion year-wise

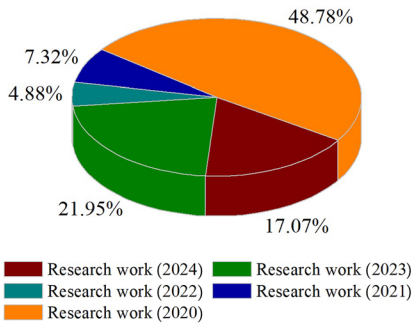
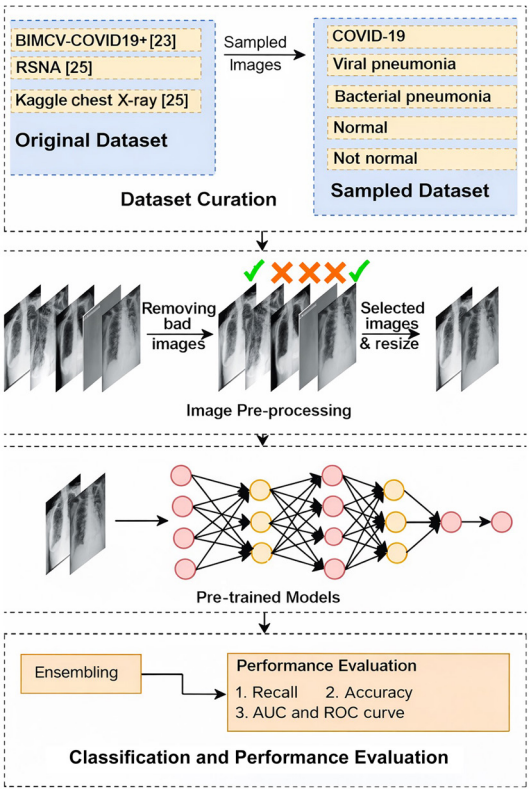


Figure 3 shows a block diagram of the proposed transfer learning model to detect COVID-19. Transfer learning provides pre-trained weights and hyperparameters that can be directly used for a new model with minimal parameter fine-tuning. There are various latest pre-trained models like ResNet, VGGNet, InceptionV3, Xception, GoogLeNet, AlexNet, and MobileNet trained on the ImageNet dataset. TL has manifold advantages [14, 34].

Figure 3
Block diagram for the proposed framework



Firstly, it offers the most efficient substitute for training a model; it eliminates the requirement to train the model from the beginning, saving a significant amount of time.

Secondly, the model requires less parameter tweaking for reliable results because it is trained using pre-trained weights.

Ensemble learning (EL) [34] is a technique that combines multiple models to solve a particular task with a more accurate result than can typically be obtained by a standalone model. It improves the performance of the single model, which reduces the chances of poor

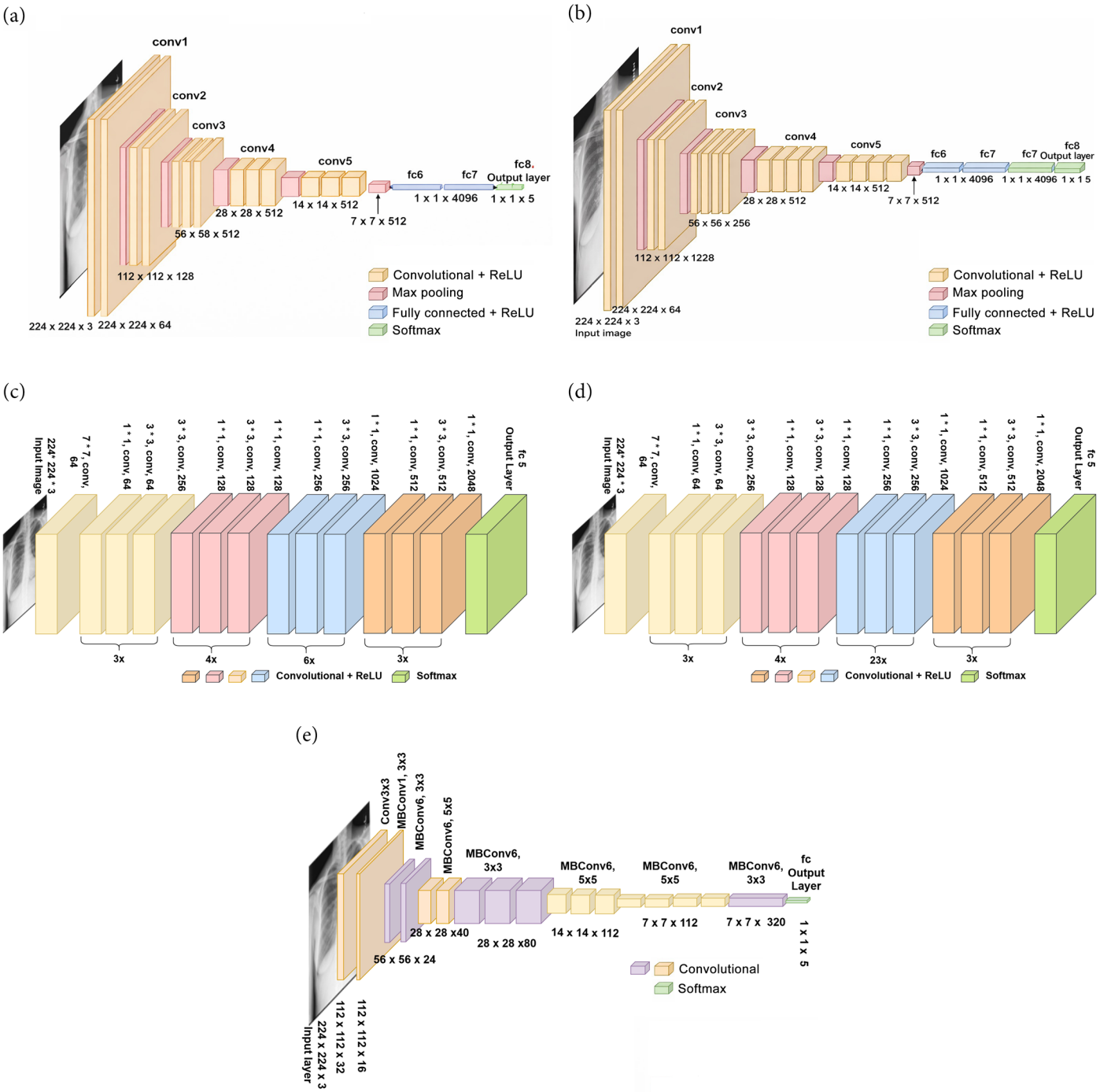
decisions. In the COVID-19-related research, ensemble learning has the state-of-the-art technique to get the best out of the CT scan and chest X-ray image with various deep learning models [35].

In this study, we use TL to fine-tune and train the model on CT scans and chest X-ray images for the task of multi-class classification. To address the detection of COVID-19 infection, we utilize five state-of-the-art CNN architectures, specifically VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0. These models have been selected based on their proven effectiveness in image recognition tasks. The following sections provide a detailed description of each of these models, outlining their unique architectures and characteristics:

1) VGG16—It [36, 37] is a convolutional neural network architecture by Visual Geometry Group from Oxford. It has a total of 21 layers: 13 convolutional layers, 5 max pooling layers, and 3 dense layers, but out of these 21 layers, only 16 are weight layers. It was trained on the ImageNet dataset, which has more than a million images and comprises 1000 plus classes (Figure 4(a)).
2) VGG19—It [37] is an advanced version of its predecessor, VGG16, with additional layers that enhance feature extraction capability. It comprises 16 convolutional layers, 3 dense layers, 5 max pooling layers, and 1 softmax layer, amounting to more than 144 million parameters (Figure 4(b)). The dense architecture includes 2 fully connected layers with 4096 neurons each activated by

Figure 4

(a) VGG16 architecture; (b) VGG19 architecture; (c) ResNet50 architecture; (d) ResNet101 architecture; (e) EfficientNetB0 architecture



ReLU, followed by a final dense layer with softmax activation for classification.

3) ResNet50—It [37] stands for residual network consisting of 50 layers. It includes 5 stages, each having a convolution and an identity block, and both blocks contain 3 convolutional layers each. This architecture has more than 23 million trainable parameters (Figure 4(c)). After the convolutional base, a Global Average Pooling layer is applied, followed by a dense layer of 256 neurons with ReLU activation, a dropout layer (rate = 0.5), and a final softmax layer. This dense configuration was kept consistent across models for a fair comparison.
4) ResNet101—It [37, 38] is an extended version of ResNet50, designed with deeper residual blocks to improve representational depth. It consists of 101 layers and approximately 45 million trainable parameters (Figure 4(d)). The model utilizes skip connections to overcome the vanishing gradient problem effectively. Similar to ResNet50, the convolutional feature extractor is followed by a Global Average Pooling layer, a dense layer of 256 neurons with ReLU activation, a dropout layer (rate = 0.5), and a final dense layer with softmax activation for output classification.
5) EfficientNetB0—It [38] is designed to optimize network performance by uniformly scaling depth, width, and resolution. It consists of 1 convolution layer, 1 MBConv1 3 × 3, 6 MBConv6 3 × 3, 9 MBConv6 5 × 5, and 5 MMConv1 layers, each incorporating depthwise convolution, batch normalization, and activation functions (Figure 4(e)). After the convolutional feature extraction, a Global Average Pooling layer is applied, followed by a dense layer of 256 neurons with ReLU activation, a dropout layer (rate = 0.5), and a softmax layer for classification. The dense configuration remains consistent with that of the other architectures.

All ensemble models were implemented using Keras with TensorFlow as the backend. Input images were resized to 224 × 224 × 3, and each model was initialized with ImageNet pre-trained weights. The original classification head was replaced with dense layers on top of the models and softmax activation to support five-class output.

Only the top layers were trained, while the convolutional base remained frozen. Models were compiled using the Adam optimizer with sparse categorical cross-entropy as the loss function and accuracy as the evaluation metric.

Training was performed for up to 50 epochs, using early stopping based on validation accuracy with a patience of 5. A batch size of 32 was used, with data shuffled before each epoch. Stratified fivefold cross-validation was applied to ensure balanced evaluation. Hyperparameters were selected based on common deep learning practices and adjusted empirically for stable convergence and performance.

3.1. Dataset

We collected data from multiple sources, as listed in Table 1. The dataset includes 3,616 COVID-19 images, 11,821 images of not normal lungs, 8,851 normal lung X-rays, 1,345 viral pneumonia images, and 2,530 bacterial pneumonia images.

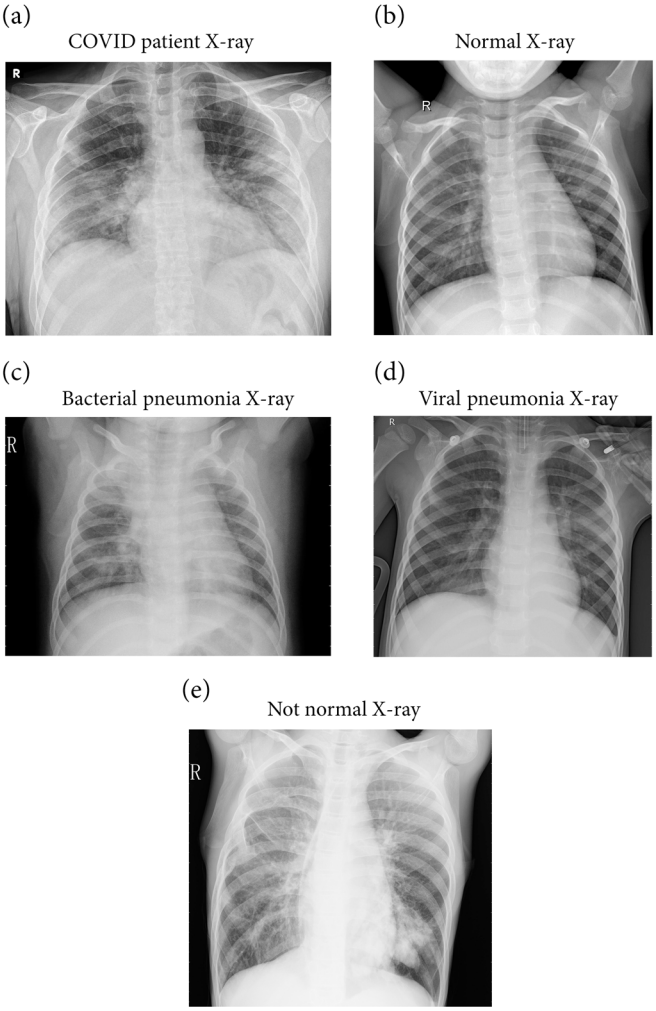
Table 1
Interpretation of the mean scale for belief, concern, and practice

S. No.	Dataset	Sources
1	COVID-19	BIMCV-COVID19+ [9]
2	Normal and not normal	RSNA [10]
3	Viral pneumonia and bacterial pneumonia	Kaggle chest X-ray images (pneumonia) [11]

3.2. Data processing

Different pre-processing techniques are applied on these datasets. We remove many bad images containing wires and arrows manually because they can lead to misclassification. To balance all the classes, we consider 2,000 COVID-19 images, 2,000 not normal images, 2,000 normal images, 1,345 viral pneumonia, and 2,000 bacterial images. Figure 5 shows the sample of used images in proposed TL model.

Figure 5
Good images from X-ray samples



We discard bad images from data in the pre-processing step because any machine learning model can learn undesired features from them, leading to misclassification. Figure 6 shows the bad images. Some examples of the rejected images for this study are listed below.

1) Some of the images have wire over the chest.
2) Some of the images have electrodes connected to the chest.
3) Some images have some electronic device connected to the chest.
4) Some images are pure solid grey images; very hard to recognize the lung in the chest X-ray.
5) Some images contained geometrical shapes, e.g. circles and arrows—name, patient ID, and date mentioned in very large font size.

Table 2 shows the distribution and the number of images in each class. During the augmentation process, we resized the images to 224 × 224 pixels. After resizing, the dataset was split into 80% for training, 10% for validation, and 10% for testing.

Figure 6
Bad images from X-ray samples

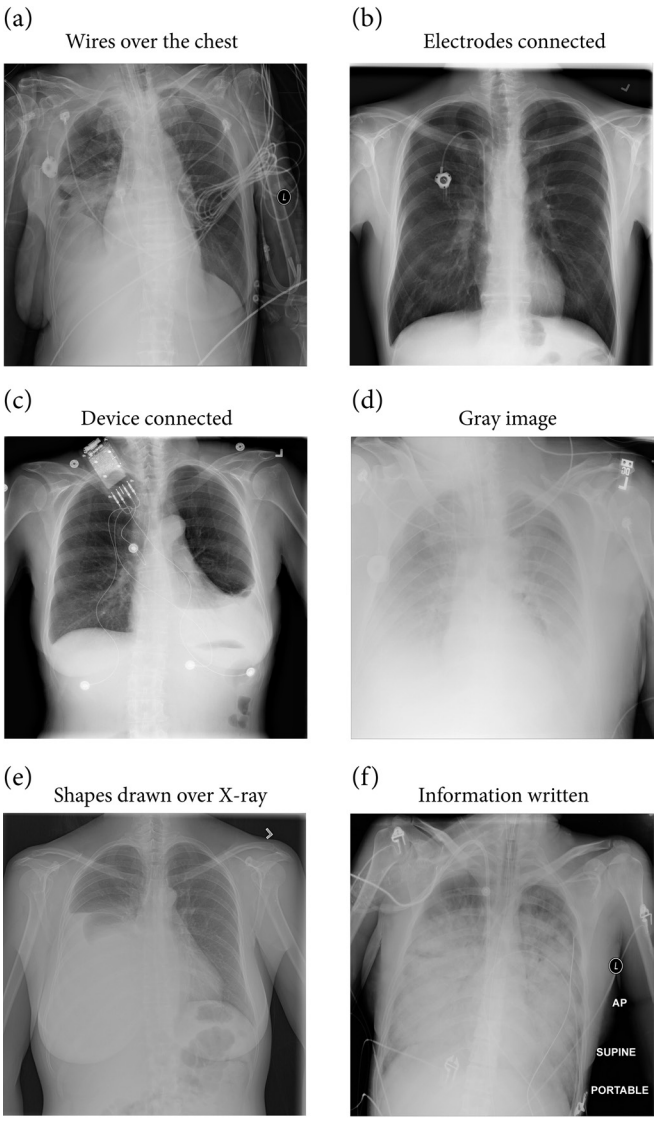


Table 2
Comparative evaluation of the dataset in suggested model

Classes	Train	Validation	Test
COVID-19	1,600	200	200
Viral pneumonia	976	121	121
Bacterial pneumonia	1,600	200	200
Normal	1,600	200	200
Not normal	1,600	200	200
Total	7,376	921	921

3.3. Method

In this work, we extracted features from the images using TL. Five distinct deep CNN models—VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0—are used to extract features. The goal of the technique known as TL is to acquire information and skills while addressing a single problem and then use those skills to address other related problems. TL-based CNN models provide the following advantages:

1) Large datasets are not required.
2) The model's training time complexity is reduced.
3) Knowledge can be leveraged to address new challenges.
4) Learning occurs at a faster rate.
5) Fewer resources are needed compared to traditional training.

Softmax is utilized in all models as an activation function in the last layer's fully linked five neurons for multi-class categorization. The softmax formula is represented by Equation (1), and the input vector to the softmax function is denoted by the symbol ($\overrightarrow{z_l}$). Any real value, whether positive, zero, or negative, can be entered. e^{z_i} represents the standard exponential function applied to each input vector element. As the formula's numerator, the normalization term makes sure that each function output value adds up to one and is between 0 and 1, which indicates the number of classes and a legitimate probability distribution. For training, Adam optimizer [39] was used. Since the picture class labels are all one integer and the classes are mutually exclusive, we have used sparse categorical cross-entropy as our loss function. Sparse categorical cross-entropy has the advantage of using less memory and processing time because it only needs one integer per class as opposed to a whole vector.

$$\sigma\left(\overrightarrow{z_l}\right) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \qquad \text{for } i = 1, 2, \ldots, K \tag{1}$$

Equation (2) represents the sparse categorical cross-entropy, where represents the classes, for $c = 1, 2, 3, \ldots, M$ is the number of classes, y_i is the actual probability of a class, and $\hat{y}_l$ is the predicted probability of that particular class. It computes the average difference between the actual and predicted probability distributions across all classes in the problem.

$$Loss = -\sum_{c=1}^{M} y_i \log\left(\hat{y}_l\right) \tag{2}$$

Since we have mutually exclusive classes, only the observed class has an output of one, and the outputs for all other classes are zero. The simplified loss function is

$$Loss = -\log\left(\hat{y}_l\right) \tag{3}$$

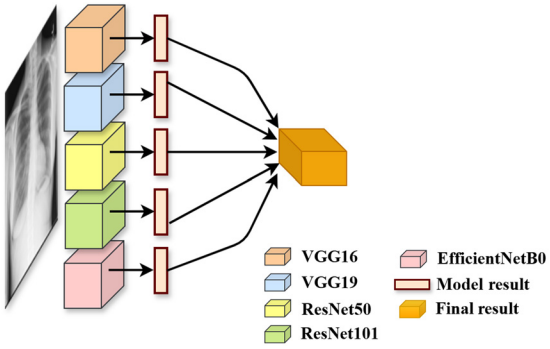
We have not included the top layers for all these models and added dense layers on top of these models. On top of VGG16 and VGG19, we added two hidden layers and one output layer with 4096 and 5 neurons, respectively. These hidden layers are fully connected to the pre-trained model.

Furthermore, we froze all pre-trained model layers and trained only newly added layers. On the other hand, only one output layer is added and trained for Resnet50, ResNet101, and EfficientNetB0. The architecture of the models is shown in Figure 4. A maximum of 50 epochs is set for the training. However, we used validation accuracy as an early stopping criterion to avoid overtraining. It allows us to specify the number of training epochs and stop training once the model performance stops improving.

We used ensemble technique to improve the results. Figure 7 represents the architecture of ensemble technique in which five models (VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0) are combined.

The inference has been obtained on the test images that are passed to the trained model. We used a voting approach to assign the final label. The procedural steps are described in Algorithm 1. This algorithm includes model initialization, training of additional dense layers, and final prediction using majority voting. The function argmax refers to selecting the class label with the highest probability score from

Figure 7
Architecture of ensemble model



Algorithm 1
Procedural steps for ensemble-based classification of chest X-ray images using five CNN models (VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0)

Input: Chest X-ray images, ground truth labels.
Output: Predicted labels.

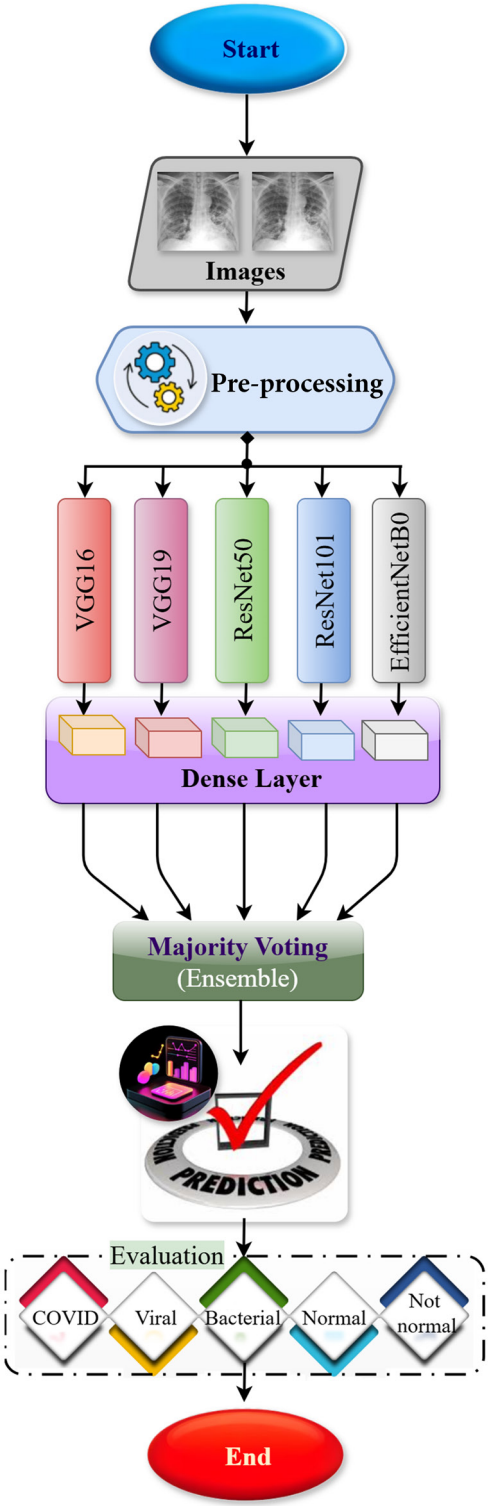
for each model ∈ models *do*
  Add fully connected nodes at the end of the pre-trained model.
  for c = 0 to 4 *do*
    $n \leftarrow$ number of the images in c class.
    for $train_i$ = 0 to n *do*
      Hyper-tune the parameter
      Train the newly added layers and the learning for the rest layers is disabled.
    end for
  end for
end for
$Predicted_c \leftarrow [\,]$
$m \leftarrow$ number of the images for test.
for $test_j$ = 0 to m *do*
  $temp \leftarrow [0, 0, 0, 0]$
  for i = 0 to 4 *do*
    $temp = temp[model_i\,(test_j)] + 1$
  end for
  $Predicted_c \leftarrow$
  $Predicted_c.append[argmax(temp)]$
end for

the softmax output layer of each model. The temp refers to a temporary variable that stores the predicted class probabilities or labels generated by each model in the ensemble. It is used to aggregate predictions before applying the ensemble decision rule (e.g., majority voting).

The flowchart in Figure 8 outlines the complete workflow of the proposed system. The process begins with raw chest X-ray image acquisition, followed by pre-processing operations such as resizing and normalization.

The processed images are then fed in parallel into five pre-trained CNN models—VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0—each adapted to the five-class classification task by replacing their top layers. During training, only the added dense layers are fine-tuned, while the convolutional bases are frozen. After inference, each model outputs a class prediction, which is aggregated through a majority voting mechanism to produce the final decision. The final prediction is evaluated using standard performance metrics including accuracy, precision, recall, F1 score, and additional error-based indicators.

Figure 8
Block diagram of the proposed ensemble deep learning framework for chest X-ray classification



3.4. Computational complexity

The proposed ensemble framework comprises five pre-trained CNNs—VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0—each fine-tuned using transfer learning [40]. To reduce computational cost, only the top added layers of each model were retrained while the convolutional backbones were frozen.

Inference performance was benchmarked on a batch of 1,000 chest X-ray images using Google Colab with a Tesla T4 GPU (15 GB GPU memory) and 12 GB system RAM. The time taken by each model to complete predictions was as follows:

1) VGG16: 41.4 seconds
2) VGG19: 10.6 seconds
3) ResNet50: 20.8 seconds
4) ResNet101: 26.3 seconds
5) EfficientNetB0: 26.9 seconds

The total inference time of the ensemble model was approximately 126 seconds for 1,000 images, computed as the sum of individual model predictions followed by a negligible ensemble decision time (~0.072 seconds). As inference runs sequentially in a single-GPU setup, the total time reflects the combined cost of all five models.

4. Results

This section presents a comprehensive analysis of the performance of the proposed ensemble deep learning model for multi-class classification of CXR images. The model was evaluated against five well-established CNN architectures: VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0. The evaluation includes accuracy, recall, confusion matrix analysis, ROC-AUC, and comparative benchmarking. The goal is to assess whether integrating multiple models via ensemble learning provides meaningful gains in classification performance, particularly for differentiating between multiple lung conditions.

4.1. Cross-validation accuracy and performance trends

To ensure robust evaluation, all models were assessed using fivefold stratified cross-validation. As shown in Table 3, the proposed ensemble model consistently outperformed individual CNNs across all folds. It achieved an average classification accuracy of 98.37%, with only minor variations across splits. Among the individual models, ResNet50 and EfficientNetB0 performed relatively well, with average accuracies of 96.37% and 95.80%, respectively. However, the ensemble model surpassed all baselines, demonstrating the effectiveness of model aggregation in capturing diverse spatial and contextual patterns present in chest radiographs.

The accuracy improvements observed are not merely statistical; they reflect a reduction in misclassification errors across critical disease classes. For instance, while individual models occasionally misclassified viral pneumonia as normal or bacterial pneumonia due to overlapping imaging features, the ensemble's collective decision-making significantly reduced such confusions.

4.2. Confusion matrix interpretation

To further evaluate class-wise predictions, confusion matrices were generated for each model. Figure 9 displays confusion matrices for the five individual CNNs and the final ensemble. The ensemble model demonstrates a more diagonal dominant matrix, indicating a higher rate of correct classifications across all categories.

In particular, the model correctly classified 197 out of 200 COVID-19 cases, misclassifying only 3 instances. Similarly, bacterial and viral pneumonia classes saw correct predictions for 197 and 196 images, respectively. These results indicate that the model is not only accurate in aggregate but also maintains high sensitivity for each disease category which is a critical factor for clinical utility.

4.3. Recall and class-wise performance

Table 4 provides a breakdown of recall for each class across all models. The ensemble model achieved the highest recall in almost all categories, including:

1) COVID-19: 98%
2) Viral pneumonia: 96%
3) Bacterial pneumonia: 98%
4) Normal: 98%
5) Not normal (non-COVID, non-pneumonia): 99%

This class-wise balance is important because many prior models are skewed toward the COVID-19 class and underperform on other categories. The proposed method achieves a uniformly high recall, which indicates that the model generalizes well across diverse radiographic patterns and is not biased toward overrepresented classes.

This demonstrates that our model not only performs competitively but does so in a more challenging and clinically realistic setting, which includes differential diagnosis across multiple lung conditions. The enhanced granularity in classification is expected to reduce misdiagnoses and aid in more accurate triaging in healthcare systems.

4.4. Summary of key results

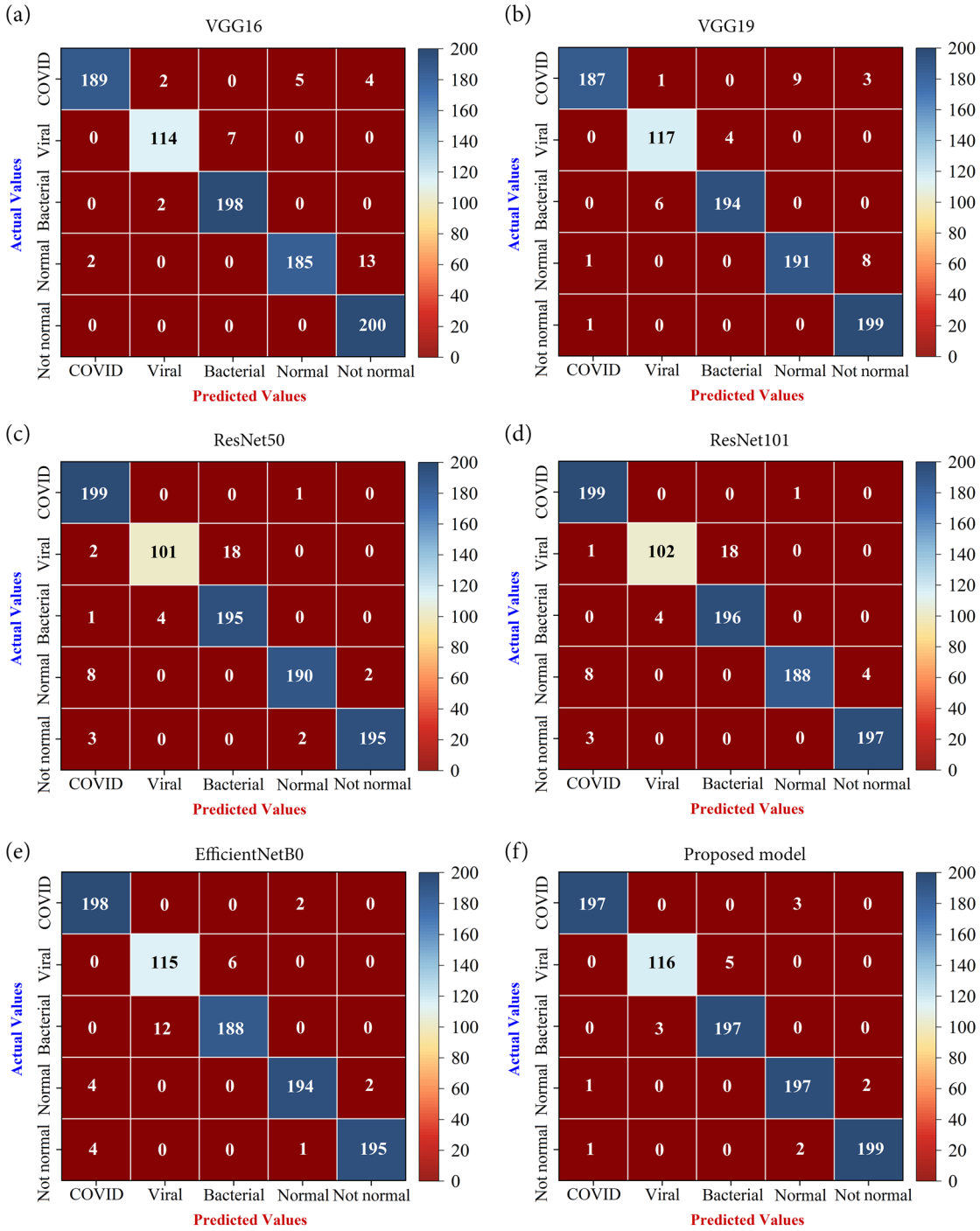
Below is the summary:

1) The ensemble model outperformed all individual CNNs in accuracy, recall, and AUC.
2) It demonstrated consistent performance across all five disease categories, reducing class imbalance bias.
3) Model stability was confirmed through validation curves and cross-validation.
4) Compared to existing literature, the model achieved superior accuracy while tackling a more complex classification problem.

Table 3
Stratified K-fold cross-validation

Class name	Fold-1 (%)	Fold-2 (%)	Fold-3 (%)	Fold-4 (%)	Fold-5 (%)	Mean (%)
VGG16	95.54	96.51	96.02	92.89	95.84	95.35
VGG19	96.08	95.96	95.66	95.56	94.15	95.48
ResNet50	95.66	96.69	96.44	95.90	97.17	96.37
ResNet101	93.91	95.30	96.74	95.45	96.02	95.49
EfficientNetB0	96.08	95.78	95.45	96.74	94.94	95.80
Proposed ensemble	98.10	98.34	98.28	98.49	98.66	98.37

Figure 9
Confusion matrix for respective models



These results validate the initial hypothesis that ensemble learning improves diagnostic performance over single-model approaches and confirm the feasibility of applying deep learning for multi-class disease detection from chest X-ray images.

4.5. Novel contributions and comparative advantage

This study introduces a multi-class classification framework that addresses a broader clinical spectrum than traditional binary or three-class models. By classifying chest X-ray images into five distinct categories—COVID-19, bacterial pneumonia, viral pneumonia, normal, and not normal—the proposed approach better reflects real-world diagnostic needs.

A key innovation lies in the use of an ensemble of five diverse CNN architectures (VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0), which enhances performance through model diversity. Each model contributes unique feature representations, and their combination through majority voting improves classification stability and accuracy.

Transfer learning further strengthens the framework by leveraging pre-trained weights, allowing efficient training on limited

Table 4
Recall evaluation

	Recall (%)					
Class name	**COVID**	**Viral**	**Bacterial**	**Not normal**	**Normal**	**Accuracy**
VGG16	96	96	93	97	99	96.20
VGG19	96	91	98	98	97	96.63
ResNet50	99	83	97	95	97	95.55
ResNet101	92	92	98	98	99	95.77
EfficientNetB0	99	95	94	97	97	96.63
Proposed ensemble	**98**	**96**	**98**	**98**	**99**	**98.37**

medical datasets. The ensemble model achieved 98.37% accuracy, outperforming all individual models and many existing studies, despite handling a more complex classification task.

In summary, the key contributions of this work are:

1) A multi-class diagnostic framework that distinguishes between five major lung conditions from chest X-rays.
2) An ensemble of heterogeneous CNNs that improves classification robustness and feature diversity.
3) Use of transfer learning and modular training, enabling the model to generalize well with limited data.
4) Empirical superiority over existing methods, especially in multi-class diagnostic settings.

These innovations collectively contribute to a more accurate, interpretable, and scalable deep learning solution for automated chest X-ray analysis in pandemic and post-pandemic healthcare environments.

The ResNet50 model, the baseline model, produced the most considerable increase in execution time, with a precision of 96.73% and an accuracy of 96.76% [35]. However, the model considers only three classes: normal, pneumonia, and COVID-19. However, the performance of the proposed model is 96.63% over a large dataset and multi-class (five classes).

5. Conclusion and Future Scope

We have collected datasets from three sources, BIMCV-COVID-19+[9], RSNA[10], and Kaggle chest X-ray images (pneumonia) [11], and created a dataset of five classes, namely, COVID-19, viral pneumonia, bacterial pneumonia, normal, and not normal. Following the initial data acquisition, rigorous pre-processing was conducted to eliminate low-quality or corrupted radiographic images, which could otherwise introduce noise and propagate erroneous learning signals during model training. The curated dataset comprised high-fidelity chest X-ray scans wherein pulmonary regions are distinctly visible, thereby enhancing the model's ability to localize and learn relevant pathological features. In line with transfer learning paradigms [41], we utilized pre-trained convolutional neural networks as backbone feature extractors and appended custom classification heads consisting of fully connected dense layers. The convolutional base of each network was frozen to preserve its pre-learned hierarchical representations, while only the newly integrated top layers were subjected to training on the domain-specific dataset. The ensemble comprised five distinct architectures—VGG16, VGG19, ResNet50, ResNet101, and EfficientNetB0—each selected for its unique architectural advantages. The individual performance metrics for these models were observed as follows: VGG16 (96.20%), VGG19 (96.63%), ResNet50 (95.55%), ResNet101 (96.19%), and EfficientNetB0 (96.63%). Upon integration into the ensemble framework, overall classification accuracy was significantly elevated, reaching 98.37%, which underscores the synergistic benefit of the multi-network fusion strategy. Unlike most existing studies that primarily address binary or three class COVID-19 classification problems, the proposed ensemble framework is designed to identify five distinct chest conditions associated with COVID-19, offering a more inclusive and practical diagnostic solution. In future work, this approach can be further validated on additional datasets and across diverse clinical settings to evaluate its generalizability and reliability. A detailed comparison with other ensemble learning strategies such as stacking, bagging, and boosting may also help position the model's performance within the broader research landscape. Furthermore, the framework could be extended to include emerging COVID-19 variants and adapted for federated learning environments, thereby enabling secure, collaborative, and scalable diagnostic model development without compromising patient data privacy.

Acknowledgement

The authors sincerely acknowledge the valuable guidance and professional support provided by Dr. Zaki Siddique and Dr. Om Shankar Chaurasiya of Maharani Laxmi Bai Medical College, Jhansi, as well as Dr. Sunil Kumar from King George's Medical University, Lucknow. Their expertise and constructive insights greatly contributed to clarifying key aspects of this research. The authors also express gratitude to the technical staff, particularly Mr. Ankit Chaurasiya, Dialysis Technician at Maharani Laxmi Bai Medical College, Jhansi, whose consistent assistance and practical support were instrumental in resolving technical challenges and facilitating the successful completion of this work.

Ethical Statement

Ethical approval exemption status for this dataset (free use). This study uses chest X-ray data that are publicly available for research use. All data are anonymized and do not include any personal or identifiable information. No experiments were conducted on human participants or animals by the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in BIMCV COVID-19+ at https://arxiv.org/pdf/2006.01174. The data that support the findings of this study are openly available in Radiological Society of North America at https://www.kaggle.com/c/rsna-pneumonia-detection-challenge. The data that support the findings of this study are openly available in Mendeley Data at https://data.mendeley.com/datasets/rscbjbr9sj/2.

Author Contribution Statement

Shubham: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Brijesh Kumar Chaurasia:** Conceptualization, Validation, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Panjeta, M., Reddy, A., Shah, R., & Shah, J. (2024). Artificial intelligence enabled COVID-19 detection: Techniques, challenges and use cases. *Multimedia Tools and Applications*, *83*(2), 4639–4666. https://doi.org/10.1007/s11042-023-15247-7

[2] Singh, A. K., Kumar, A., Kumar, V., & Prakash, S. (2024). COVID-19 detection using adopted convolutional neural networks and high-performance computing. *Multimedia Tools and Applications*, *83*(1), 593–608. https://doi.org/10.1007/s11042-023-15640-2

[3] Hu, J. C., & Sethi, S. (2024). New methods to detect bacterial or viral infections in patients with chronic obstructive pulmonary disease. *Expert Review of Respiratory Medicine*, *18*(9), 693–707. https://doi.org/10.1080/17476348.2024.2396413

[4] Shukla, M. M., Tripathi, B. K., Nagle, M., & Chaurasia, B. K. (2022). COVID-19 & lung disease detection using deep learning. In *2022 14th International Conference on Computational Intelligence and Communication Networks*, 430–434. https://doi.org/10.1109/CICN56167.2022.10008269

[5] Tan, S. L., Selvachandran, G., Paramesran, R., & Ding, W. (2025). Lung cancer detection systems applied to medical images: A state-of-the-art survey. *Archives of Computational Methods in Engineering*, *32*(1), 343–380. https://doi.org/10.1007/s11831-024-10141-3

[6] Aboshosha, A. (2025). AI based medical imagery diagnosis for COVID-19 disease examination and remedy. *Scientific Reports*, *15*(1), 1607. https://doi.org/10.1038/s41598-024-84644-1

[7] Singh, S., Tripathi, B. K., & Rawat, S. S. (2023). Deep quaternion convolutional neural networks for breast cancer classification. *Multimedia Tools and Applications*, *82*(20), 31285–31308. https://doi.org/10.1007/s11042-023-14688-4

[8] Mallick, D., Singh, A., Ng, E. Y.-K, & Arora, V. (2025). Classifying chest X-rays for COVID-19 through transfer learning: A systematic review. *Multimedia Tools and Applications*, *84*(2), 689–748. https://doi.org/10.1007/s11042-024-18924-3

[9] Chowdhury, D., Banerjee, S., Sannigrahi, M., Chakraborty, A., Das, A., Dey, A., & Dwivedi, A. D. (2023). Federated learning based COVID-19 detection. *Expert Systems*, *40*(5), e13173. https://doi.org/10.1111/exsy.13173

[10] Kumar, R., Arora, R., Bansal, V., Sahayasheela, V. J., Buckchash, H., Imran, J., …, & Raman, B. (2022). Classification of COVID-19 from chest X-ray images using deep features and correlation coefficient. *Multimedia Tools and Applications*, *81*(19), 27631–27655. https://doi.org/10.1007/s11042-022-12500-3

[11] Srivastava, S., & Tripathi, B. K. (2024). Development of deep intelligent system in complex domain for human recognition. *International Journal of Advanced Intelligence Paradigms*, *28*(1–2), 53–73. https://doi.org/10.1504/IJAIP.2024.139949

[12] Moreh, F., Hasan, Y., Rizvi, Z. H., Tomforde, S., & Wuttke, F. (2025). Hybrid neural network method for damage localization in structural health monitoring. *Scientific Reports*, *15*(1), 7991. https://doi.org/10.1038/s41598-025-92396-9

[13] Shaik, A., Dutta, S. S., Sawant, I. M., Kumar, S., Balasundaram, A., & De, K. (2025). An attention based hybrid approach using CNN and BiLSTM for improved skin lesion classification. *Scientific Reports*, *15*(1), 15680. https://doi.org/10.1038/s41598-025-00025-2

[14] Chaurasia, B. K., Raj, H., Rathour, S. S., & Singh, P. B. (2023). Transfer learning-driven ensemble model for detection of diabetic retinopathy disease. *Medical & Biological Engineering & Computing*, *61*(8), 2033–2049. https://doi.org/10.1007/s11517-023-02863-6

[15] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data*, *9*(1), 102. https://doi.org/10.1186/s40537-022-00652-w

[16] Mohan, G., Subashini, M. M., Balan, S., & Singh, S. (2024). A multiclass deep learning algorithm for healthy lung, Covid-19 and pneumonia disease detection from chest X-ray images. *Discover Artificial Intelligence*, *4*(1), 20. https://doi.org/10.1007/s44163-024-00110-x

[17] Al-qaness, M. A. A., Zhu, J., AL-Alimi, D., Dahou, A., Alsamhi, S. H., Abd Elaziz, M., & Ewees, A. A. (2024). Chest X-ray images for lung disease detection using deep learning techniques: A comprehensive survey. *Archives of Computational Methods in Engineering*, *31*(6), 3267–3301. https://doi.org/10.1007/s11831-024-10081-y

[18] Islam, R., & Tarique, M. (2022). Chest X-ray images to differentiate COVID-19 from pneumonia with artificial intelligence techniques. *International Journal of Biomedical Imaging*, *2022*(1), 5318447. https://doi.org/10.1155/2022/5318447

[19] Mzoughi, H., Njeh, I., Slima, M. B., & BenHamida, A. (2023). Deep Efficient-Nets with transfer learning assisted detection of COVID-19 using chest X-ray radiology imaging. *Multimedia Tools and Applications*, *82*(25), 39303–39325. https://doi.org/10.1007/s11042-023-15097-3

[20] Islam, M. Z., Islam, M. M., & Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, *20*, 100412. https://doi.org/10.1016/j.imu.2020.100412

[21] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Computer Methods and Programs in Biomedicine*, *196*, 105581. https://doi.org/10.1016/j.cmpb.2020.105581

[22] Malik, H., Anees, T., Din, M., & Naeem, A. (2023). 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*, *82*(9), 13855–13880. https://doi.org/10.1007/s11042-022-13843-7

[23] Hasija, S., Akash, P., Hemanth, M. B., Kumar, A., & Sharma, S. (2022). A novel approach for detection of COVID-19 and pneumonia using only binary classification from chest CT-scans. *Neuroscience Informatics*, *2*(4), 100069. https://doi.org/10.1016/j.neuri.2022.100069

[24] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical Image Analysis*, *65*, 101794. https://doi.org/10.1016/j.media.2020.101794

[25] Ezzat, D., Hassanien, A. E., & Ella, H. A. (2021). An optimized deep learning architecture for the diagnosis of COVID-19

disease based on gravitational search optimization. *Applied Soft Computing*, *98*, 106742. https://doi.org/10.1016/j.asoc.2020.106742

[26] Marques, G., Agarwal, D., & de la Torre Díez, I. (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Applied Soft Computing*, *96*, 106691. https://doi.org/10.1016/j.asoc.2020.106691

[27] Das, D., Santosh, K. C., & Pal, U. (2020). Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Physical and Engineering Sciences in Medicine*, *43*(3), 915–925. https://doi.org/10.1007/s13246-020-00888-x

[28] Ayalew, A. M., Salau, A. O., Tamyalew, Y., Abeje, B. T., & Woreta, N. (2023). X-ray image-based COVID-19 detection using deep learning. *Multimedia Tools and Applications*, *82*(28), 44507–44525. https://doi.org/10.1007/s11042-023-15389-8

[29] Babukarthik, R. G., Adiga, V. A. K., Sambasivam, G., Chandramohan, D., & Amudhavel, J. (2020). Prediction of COVID-19 using Genetic Deep Learning Convolutional Neural Network (GDCNN). *IEEE Access*, *8*, 177647–177666. https://doi.org/10.1109/ACCESS.2020.3025164

[30] Bhosale, Y. H., & Patnaik, K. S. (2023). Application of deep learning techniques in diagnosis of Covid-19 (coronavirus): A systematic review. *Neural Processing Letters*, *55*(3), 3551–3603. https://doi.org/10.1007/s11063-022-11023-0

[31] Kuzinkovas, D., & Clement, S. (2023). The detection of COVID-19 in chest X-rays using ensemble CNN techniques. *Information*, *14*(7), 370. https://doi.org/10.3390/info14070370

[32] Wang, X., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H, …, & Zheng, C. (2020). A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Transactions on Medical Imaging*, *39*(8), 2615–2625. https://doi.org/10.1109/TMI.2020.2995965

[33] Liu, W., Ni, Z., Chen, Q., & Ni, L. (2023). Attention-guided partial domain adaptation for automated pneumonia diagnosis from chest X-ray images. *IEEE Journal of Biomedical and Health Informatics*, *27*(12), 5848–5859. https://doi.org/10.1109/JBHI.2023.3313886

[34] Yang, Y., Lv, H., & Chen, N. (2023). A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review*, *56*(6), 5545–5589. https://doi.org/10.1007/s10462-022-10283-5

[35] Kumar, A., & Chaurasia, B. K. (2024). Detection of SARS-CoV-2 virus using lightweight convolutional neural networks. *Wireless Personal Communications*, *135*(2), 941–965. https://doi.org/10.1007/s11277-024-11097-0

[36] Puspita, R., Rahayu, C. (2024). Pneumonia prediction on chest X-ray images using deep learning approach. *IAES International Journal of Artificial Intelligence*, *13*(1), 467–474. http://doi.org/10.11591/ijai.v13.i1.pp467-474

[37] Kumar, S., & Mallik, A. (2023). COVID-19 detection from chest X-rays using trained output based transfer learning approach. *Neural Processing Letters*, *55*(3), 2405–2428. https://doi.org/10.1007/s11063-022-11060-9

[38] Zhou, T., Chang, X., Liu, Y., Ye, X., Lu, H., & Hu, F. (2023). COVID-ResNet: COVID-19 recognition based on improved attention ResNet. *Electronics*, *12*(6), 1413. https://doi.org/10.3390/electronics12061413

[39] Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, *35*(23), 17095–17112. https://doi.org/10.1007/s00521-023-08568-z

[40] Ravi, V., Acharya, V., & Alazab, M. (2023). A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. *Cluster Computing*, *26*(2), 1181–1203. https://doi.org/10.1007/s10586-022-03664-6

[41] Abou Baker, N., Zengeler, N., & Handmann, U. (2022). A transfer learning evaluation of deep neural networks for image classification. *Machine Learning and Knowledge Extraction*, *4*(1), 22–41. https://doi.org/10.3390/make4010002

How to Cite: Shubham, & Chaurasia, B. K. (2026). Multi-network–Based COVID-19 Detector Using Chest X-ray Images. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS62024673