International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 945-963
https://doi.org/10.70917/ijcisim-2026-3348
Article

A Multimodal Hybrid Deep Learning Ensemble Framework for Classification of COVID-19, Pneumonia, Lung Cancer, and Normal Chest Conditions

Jignesh Solanki[1], Vikram Agrawal[2*]

[1]Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India
[2*]Department of Computer Engineering, Bhailalbhai & Bhikhabhai Institute of Technology, Vallabh Vidya Nagar, Anand, Gujarat, India
[*]**Corresponding Author**: agrawal.vm@gmail.com

Abstract: - Respiratory disorders, including pneumonia, COVID-19, and lung cancer, continue to pose significant challenges to global healthcare systems due to their high prevalence, mortality rates, and overlapping radiological manifestations. Accurate differentiation among these conditions remains difficult because conventional diagnostic procedures based on chest X-ray (CXR) and computed tomography (CT) imaging depend extensively on expert interpretation, which may lead to inter-observer variability, delayed diagnosis, and reduced detection sensitivity, particularly during the early stages of disease progression. These challenges are further amplified in resource-limited clinical environments where access to specialized expertise is restricted. To overcome these limitations, this study presents a novel Multi-Objective Optimized Hybrid Deep Learning (MUOPTDL) framework for automated respiratory disease classification using medical imaging modalities. While pneumonia serves as the primary target disease, additional categories, including COVID-19, lung cancer, and normal cases, are incorporated to enhance diagnostic discrimination and minimize classification errors caused by overlapping imaging characteristics. The proposed MUOPTDL framework combines three complementary deep convolutional neural network architectures, namely DenseNet121, MobileNetV3, and Xception, to extract robust and diverse feature representations from chest CT and X-ray images. The framework employs transfer learning, data augmentation, dropout-based regularization, and adaptive optimization techniques to improve model generalization and reduce overfitting. Experimental results demonstrate that the proposed weighted hybrid ensemble significantly outperforms individual baseline models, achieving an overall accuracy, precision, recall, and F1-score of 98%, along with an area under the receiver operating characteristic curve (AUC) of 0.9866. The developed framework effectively differentiates among clinically similar radiographic patterns, including COVID-19-associated ground-glass opacities, pneumonia-related consolidations, and early-stage pulmonary malignancies. Furthermore, implementation through a web-based clinical interface highlights the practical applicability of the proposed system for real-time screening and computer-aided diagnostic decision support in healthcare environments. .

Keywords: Multi-class classification; Deep learning; Medical image analysis; respiratory disease diagnosis; Hybrid ensemble learning; multi-objective optimization

1. Introduction

Respiratory diseases remain among the leading causes of morbidity and mortality worldwide, posing substantial challenges to healthcare systems and public health infrastructure. Among these diseases, pneumonia, Coronavirus Disease 2019 (COVID-19), and lung cancer represent major clinical concerns due to their high prevalence, severe health consequences, and socioeconomic impact. Although these respiratory disorders originate from different pathological mechanisms, they frequently exhibit similar clinical symptoms and radiological manifestations, including cough, dyspnea, hypoxemia, pulmonary infiltrates, and chest imaging abnormalities. The considerable overlap in

clinical and imaging characteristics often complicates differential diagnosis and may lead to delayed treatment and adverse patient outcomes.

The global outbreak of COVID-19 highlighted the urgent need for reliable and automated diagnostic systems capable of supporting clinicians during large-scale healthcare emergencies. Chest imaging modalities, particularly chest X-ray (CXR) and computed tomography (CT), have become indispensable diagnostic tools because of their widespread availability, rapid acquisition, and cost-effectiveness. These imaging techniques play a crucial role in the diagnosis and monitoring of various pulmonary diseases. Recent advances in artificial intelligence (AI), especially deep learning (DL), have demonstrated significant potential in medical image analysis by automatically extracting hierarchical feature representations and achieving high diagnostic accuracy across diverse classification tasks [1].

Despite the remarkable progress achieved in AI-assisted medical imaging, several challenges continue to limit the clinical applicability and generalizability of existing diagnostic systems. These challenges include class imbalance, heterogeneity among imaging modalities, overlapping radiological characteristics, and the difficulties associated with real-world clinical deployment. Previous studies have shown that COVID-19, pneumonia, and other pulmonary infections frequently share common radiological features, such as ground-glass opacities and pulmonary consolidations, thereby increasing diagnostic complexity and inter-observer variability [2]. Furthermore, manual interpretation of medical images remains a time-consuming process and is often influenced by the expertise and experience of radiologists, highlighting the need for robust computer-aided diagnostic systems [3].

Deep convolutional neural networks (DCNNs) have emerged as powerful tools for medical image analysis due to their ability to learn discriminative feature representations directly from raw imaging data. Numerous studies have demonstrated that deep learning approaches outperform conventional machine learning methods based on handcrafted feature extraction techniques [4]. However, models based on a single neural network architecture often exhibit limited generalization capability when applied to heterogeneous datasets and varying clinical scenarios. Figure 1 illustrates a typical artificial intelligence-based diagnostic workflow for chest radiography, where medical images undergo preprocessing, feature extraction, classification, and clinical decision support processes [5]. Such automated systems can assist clinicians in disease identification, patient prioritization, and treatment planning.

The complementary use of chest CT and chest X-ray imaging provides a comprehensive approach for pulmonary disease assessment. Chest CT imaging offers detailed anatomical and structural information that facilitates the differentiation of COVID-19 and other pulmonary disorders, whereas chest radiography serves as a rapid, accessible, and cost-effective screening modality [6].

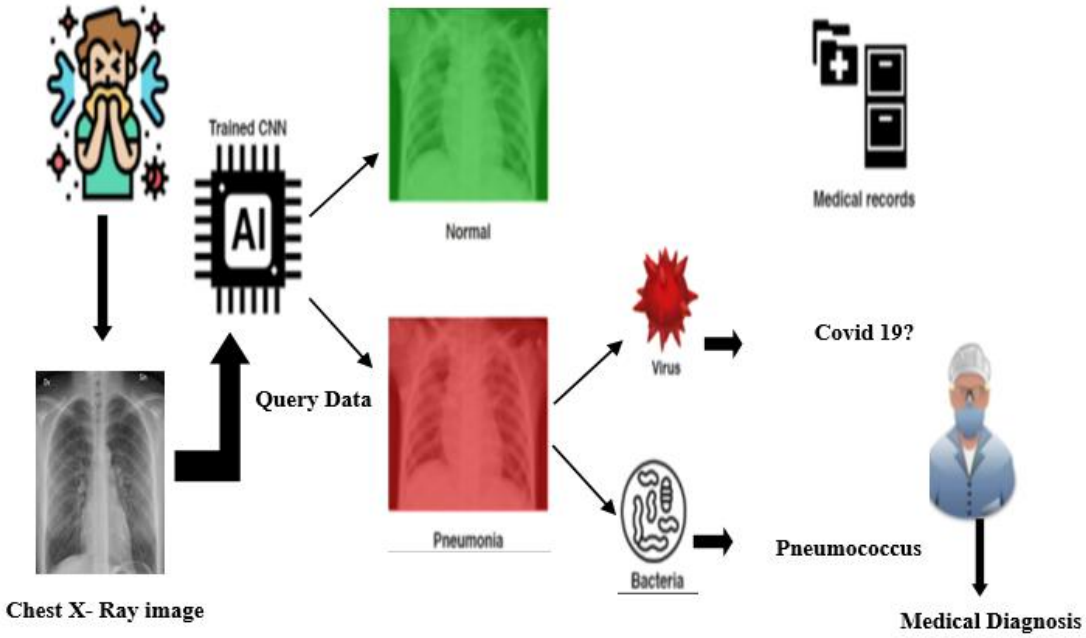

Fig. 1 AI-Based Chest X-Ray Diagnosis for Respiratory Diseases

However, the integration of multimodal imaging data presents several technical challenges due to differences in image resolution, contrast characteristics, and texture information. Previous studies have demonstrated that transfer learning and multi-scale feature extraction strategies can effectively capture both global and local image characteristics, thereby improving diagnostic performance [7].Another major challenge in medical image analysis is class imbalance, which can adversely affect model training and bias predictions toward majority classes [8]. To address these limitations, advanced convolutional neural network architectures have been developed to improve feature representation and model generalization. DenseNet121 employs dense connectivity mechanisms to enhance feature reuse, facilitate gradient propagation, and improve representation learning efficiency [9]. MobileNetV3 combines computational efficiency with high classification accuracy, making it suitable for real-time clinical applications and resource-constrained environments [10]. Furthermore, residual learning strategies have been shown to improve optimization stability and facilitate effective deep network training [11]. The evaluation of AI-assisted diagnostic systems requires comprehensive performance assessment beyond conventional accuracy measures. Metrics such as the receiver operating characteristic (ROC) curve, area under the curve (AUC), F1-score, and Matthews correlation coefficient (MCC) provide more reliable performance estimation, particularly for imbalanced medical datasets [12,13]. According to clinical guidelines published by the World Health Organization (WHO), rapid and accurate diagnosis remains essential for effective patient management and improved healthcare outcomes [14]. Recent studies have also demonstrated that AI-assisted diagnostic systems can significantly enhance the speed and accuracy of respiratory disease detection while maintaining high sensitivity and specificity [15].Motivated by these challenges, this study proposes a Multi-Objective Optimized Hybrid Deep Learning (MUOPTDL) framework for automated multi-class respiratory disease classification using chest CT and X-ray imaging. The proposed framework integrates three complementary deep learning architectures, namely Xception, MobileNetV3, and DenseNet121, to exploit their individual strengths in feature extraction and classification. Specifically, Xception is employed to capture complex spatial and textural patterns through depthwise separable convolutions, MobileNetV3 provides lightweight and computationally efficient feature representations suitable for real-time deployment, and DenseNet121 enhances feature propagation and representation learning through dense connectivity mechanisms. The outputs of these networks are combined using a weighted ensemble strategy to improve classification robustness and reduce prediction uncertainty.The primary objective of this research is to develop a reliable, computationally efficient, and clinically applicable multi-class diagnostic framework capable of accurately differentiating among COVID-19, pneumonia, lung cancer, and normal cases. By integrating multimodal chest imaging with hybrid deep learning and ensemble optimization techniques, the proposed framework aims to improve diagnostic accuracy, enhance model generalization, and provide effective decision support for respiratory disease diagnosis in real-world clinical environments.

2. Literature Review

2.1 Deep Learning in Medical Image Analysis

Deep learning has emerged as one of the most influential computational paradigms in medical image analysis due to its capability to automatically learn hierarchical feature representations from complex imaging data. Recent comprehensive reviews have demonstrated that convolutional neural networks (CNNs) have become the dominant methodology in healthcare imaging applications because of their superior feature extraction and classification capabilities compared with conventional machine learning approaches based on handcrafted features [1,4]. The increasing adoption of artificial intelligence in healthcare has further highlighted its potential for supporting disease diagnosis, clinical decision-making, and large-scale public health management, particularly during pandemic situations involving respiratory disorders [15]. The advancement of modern deep learning architectures has been largely driven by innovations in network connectivity and optimization strategies. Residual learning frameworks and densely connected neural networks have demonstrated significant improvements in gradient propagation, feature reuse, and training stability, thereby enabling the development of highly effective diagnostic models for medical imaging applications [11,16]. These architectural improvements have substantially enhanced the capability of deep neural networks to capture complex pathological patterns and improve classification performance.

Existing research findings suggest that accurate thoracic disease classification requires the integration of transfer learning strategies with hybrid deep learning architectures. Such approaches facilitate robust feature extraction, improve model generalization, and enhance diagnostic performance across diverse medical imaging modalities and clinical environments.

2.2 Deep Learning for Lung Disease Detection

Deep learning-based approaches for pulmonary disease diagnosis using chest X-ray and computed tomography (CT) imaging have attracted considerable research attention in recent years. Several studies have demonstrated the effectiveness of convolutional neural network (CNN) architectures for the automated detection and classification of respiratory diseases. Abdulahi et al. [17] proposed PulmoNet, a deep learning framework incorporating image segmentation techniques to improve the identification and classification of pulmonary abnormalities across multiple disease categories. Similarly, Kolhar et al. [18] investigated transfer learning strategies combined with lightweight CNN architectures to enhance computational efficiency and facilitate clinical deployment.

To improve pneumonia detection performance, Dardouri [19,34] explored convolutional neural network models integrated with data augmentation techniques, demonstrating enhanced model robustness and classification accuracy. Karimullah et al. [20] developed an intelligent diagnostic framework that combines CT image analysis with Internet of Things (IoT)-based data transmission for automated lung cancer detection. In addition, Rampriya et al. [27] introduced optimized deep learning approaches for pulmonary disease risk assessment, while Sanida et al. [28] proposed a multi-class classification framework for chest X-ray image analysis involving several respiratory disease categories. Mohan et al. [29] further demonstrated the feasibility of multi-class deep learning models for the simultaneous detection of COVID-19 and pneumonia from chest radiographs. More recently, ensemble learning strategies have gained increasing attention due to their ability to improve classification performance and model generalization. Kumar et al. [30] proposed a multi-model deep learning framework based on the integration of several pre-trained CNN architectures, resulting in enhanced diagnostic accuracy. Likewise, Angara et al. [35] combined convolutional neural networks with Vision Transformer architectures to develop a hybrid pneumonia detection system, demonstrating the advantages of integrating complementary feature extraction mechanisms. Despite these advancements, many existing studies remain limited to binary classification tasks or single-modality imaging datasets, which restrict their generalizability and practical applicability in complex clinical environments.

2.3 Data Augmentation, Class Imbalance, and Evaluation Metrics

Medical imaging datasets often suffer from limited sample availability and severe class imbalance, which can negatively affect the training and generalization capability of deep learning models. To address these challenges, numerous studies have investigated data augmentation and balanced learning strategies to improve classification performance. Yang et al. [22] presented a comprehensive review of image augmentation techniques and demonstrated their importance in enhancing the robustness and generalization ability of deep learning models. Similarly, Nagaye et al. [23] introduced an improved hide-and-seek augmentation approach that effectively increased model resilience by exposing networks to partially occluded image regions. Sriwiboon [24] further demonstrated that the combination of augmentation, model pruning, and quantization techniques can improve both computational efficiency and diagnostic accuracy in COVID-19 detection systems.

Kumar and Bhowmik [25] proposed an efficient convolutional neural network architecture that utilizes balanced augmentation strategies to optimize COVID-19 diagnosis performance. In addition, Nguyen and Hoang [8] investigated adaptive weighting techniques for imbalanced learning problems and emphasized the importance of balanced optimization during model training. Furthermore, Imani et al. [13] showed that conventional performance metrics such as the area under the receiver operating characteristic curve (ROC-AUC) may provide misleading results when applied to highly imbalanced datasets. Their findings highlighted the necessity of employing complementary evaluation metrics, including the F1-score and Matthews Correlation Coefficient (MCC), to obtain a more reliable assessment of model performance. These studies collectively indicate that effective data augmentation, balanced training methodologies, and comprehensive evaluation metrics are essential components for developing robust multi-class respiratory disease classification systems.

2.4 Multi-Modal Learning and Hybrid Ensemble Frameworks

The integration of chest computed tomography (CT) and chest X-ray imaging has emerged as a promising approach for comprehensive respiratory disease diagnosis. By combining information obtained from multiple imaging modalities, multi-modal learning frameworks can capture complementary anatomical and pathological characteristics, thereby improving diagnostic accuracy. Amara et al. [2] reviewed recent advances in CT image segmentation techniques for COVID-19 diagnosis, whereas Qian et al. [7] investigated adaptive fusion methods for integrating heterogeneous imaging information from multiple sources.

Hybrid and ensemble learning frameworks have demonstrated considerable potential for improving the performance and robustness of medical image classification systems. Shaikh et al. [9] employed densely connected convolutional networks combined with stacking ensemble techniques and reported significant improvements in medical image analysis performance. In addition, the incorporation of attention mechanisms and multi-branch architectures has further enhanced feature extraction and classification capabilities. Li et al. [36] proposed a multi-branch residual attention network for thoracic disease diagnosis, demonstrating the effectiveness of extracting features through multiple information pathways. Similarly, Zhu et al. [37] introduced a dual-stream feature fusion architecture that highlighted the importance of structured integration mechanisms in multi-branch learning systems.

The interpretability of artificial intelligence models has also become an important consideration for clinical applications. Dalvi et al. [38] investigated various convolutional neural network architectures combined with explainable artificial intelligence techniques for pneumonia detection and emphasized the necessity of model transparency in healthcare decision-making. Furthermore, Lin et al. [39] developed a multi-class anomaly detection framework for medical image analysis, while Owda et al. [40] proposed a lightweight hybrid deep learning model for tuberculosis detection that achieved high diagnostic accuracy with reduced computational complexity. Although ensemble and hybrid learning approaches have demonstrated superior performance and improved stability, many existing methods rely on computationally intensive architectures that limit their practical implementation in resource-constrained clinical environments.

2.5 Research Gap and Motivation

Despite the substantial progress achieved in deep learning-based respiratory disease diagnosis, several important challenges remain unresolved. First, a large proportion of existing studies focus primarily on binary classification problems, limiting their applicability to real-world clinical scenarios that require differentiation among multiple respiratory diseases. Second, many ensemble-based frameworks employ computationally expensive architectures, making their deployment difficult in resource-constrained healthcare settings. Third, current multi-modal diagnostic systems often struggle to simultaneously achieve high classification accuracy, computational efficiency, model stability, and effective generalization across diverse imaging modalities.

To address these limitations, this study proposes a Multi-Objective Optimized Hybrid Deep Learning framework that integrates DenseNet121, MobileNetV3, and Xception architectures through a probability-level weighted ensemble strategy. The proposed framework combines the robust feature extraction capability of deep hierarchical networks with the computational efficiency of lightweight architectures to enable accurate and reliable multi-class classification of respiratory diseases using both chest CT and chest X-ray images. By integrating complementary feature representations and optimized ensemble learning techniques, the proposed system aims to provide a practical, stable, and clinically deployable solution for automated respiratory disease diagnosis

3. Research Methodology

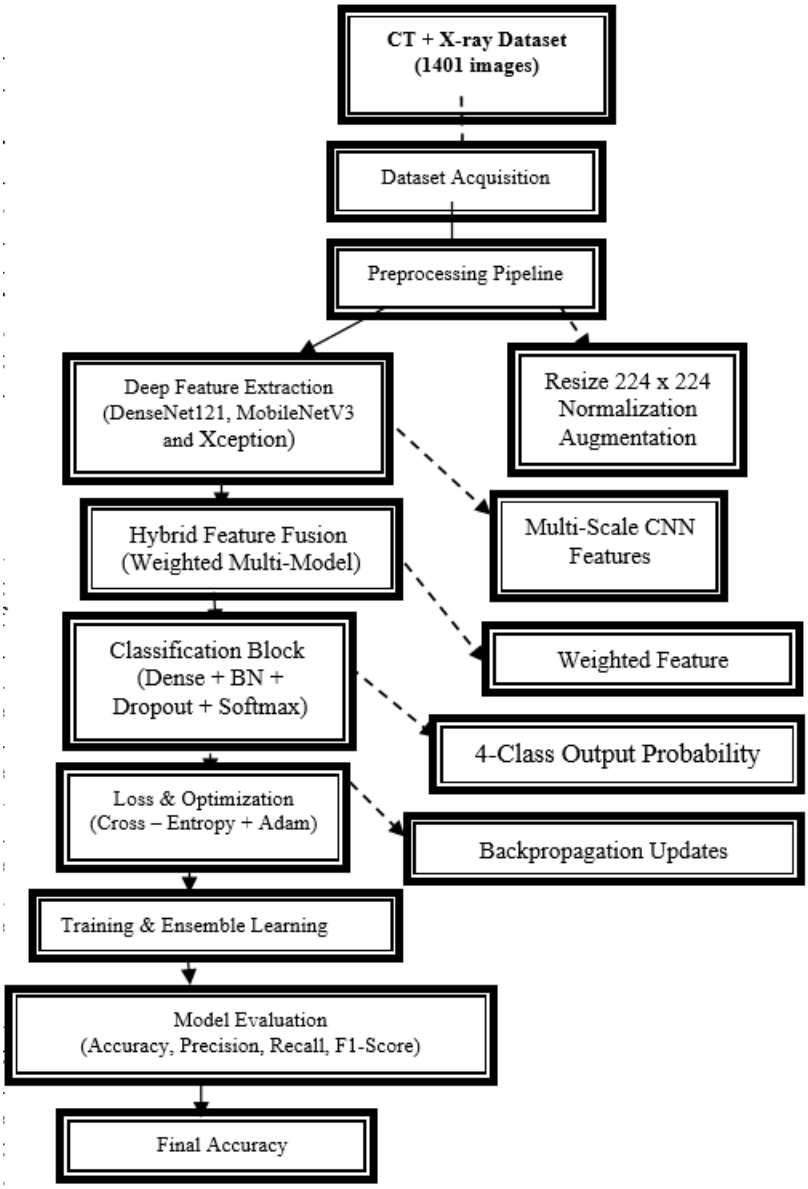

Fig. 2 Workflow of the Hybrid Deep Learning

The complete pipeline consists of dataset preparation, preprocessing, dual-model feature extraction, probabilistic fusion, optimization, and evaluation.

3.1 Dataset Acquisition

Let the dataset be defined as:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$$

where:

- $x_i \in \mathbb{R}^{H \times W \times C}$represents a chest image
- $y_i \in \{1,2,3,4\}$denotes the corresponding class label

Here, H, W, and Crepresent height, width, and channels, respectively. All images are resized to a fixed spatial resolution to ensure uniform feature extraction across models.

Table 1. Dataset distribution and augmentation configuration used for multi-class respiratory disease classification

Metric	Value
Train Images	5,592
Test Images	1,401
Classes	4
Input Size	224×224
Epochs	25
Effective Samples	139,800

Table 1 presents the dataset and training configuration for multi-class respiratory disease classification. The Softmax function transforms these logits into class probabilities for the system. The dataset contains 5,592 training images and 1,401 testing images, which are divided into four classes, and all images have been resized to 224×224. Data augmentation increased the effective samples to 139,800, which enhanced the model's robustness and generalisation capabilities.

3.2 Preprocessing Pipeline

The system requires image input to go through resizing and normalisation processes before it can use the model. The system applies min–max normalization to scale pixel intensities through its normalisation method:

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

The process of transformation leads to two benefits, which include stabilising gradient updates and enabling faster convergence. The system uses stochastic augmentation for generalisation improvement, which applies:

$$\tilde{x}_i = \mathcal{T}(x_i')$$

Where $\mathcal{T}$ represent the random transformations that include rotation, horizontal flipping, and zooming. The method decreases overfitting while enhancing protection against various imaging conditions.

3.3 Dual-Backbone Feature Extraction

The two convolutional neural networks extract deep hierarchical features through their design. The learned representations x_i for an input image are measured through multiple factors:

where:

$$F_D = DenseNet121(x_i).$$

$$F_M = MobileNetV3(x_i).$$

$$F_X = Xception(x_i).$$

DenseNet121 facilitates efficient feature reuse through dense connectivity, MobileNetV3 provides computationally efficient representations suitable for real-time deployment, and Xception captures discriminative spatial features using depthwise separable convolutions. The integration of these architectures enables the extraction of diverse and robust feature representations, thereby improving the accuracy and reliability of multi-class respiratory disease classification.

3.4 Hybrid Feature Fusion Strategy

The system uses probability-level fusion instead of feature concatenation to decrease processing needs and create more reliable outcomes.

The system calculates the final hybrid prediction through the following equation:

$$P_{Hybrid} = \alpha P_D + (1 - \alpha) P_M$$

where:

- $0 \leq \alpha \leq 1$
- αcontrols the contribution of DenseNet121

In practice, equal weighting is used:

$$\alpha = 0.5$$

The system uses this averaging method to decrease prediction uncertainty while it creates equal decision boundaries that match the two networks' outputs.

The system outputs the predicted class label as follows:

$$\hat{y} = \arg\max_k P_{Hybrid}^{(k)}$$

Algorithm 1: Proposed Hybrid Model

1. **Input:** Medical image dataset $D = \{x_i, y_i\}$, where $x_i$as an image $y_i \in \{COVID, Pneumonia, LungCancer, Normal\}$
2. **Preprocessing:**
   a. Resize each image to 224×224.
   b. Normalise pixel values to [0, 1].
   c. Apply augmentation (rotation, flip, zoom).
3. **Feature Extraction:**
   a. Extract feature maps .$F_D = f_{DenseNet121(x_i)}$
   b. Extract feature maps .$F_M = f_{MobileNetV3(x_i)}$
   c. Extract feature maps $F_X = f_{Xception(x_i)}$
   d.
4. **Feature Fusion:**

   Combine both feature sets:

   $$F_{Weighted} = aF_D + bF_X + cF_M$$

   Where a+ b+ c =1

   where FX represents the feature embedding obtained from the Xception network, and a, b and c are non-negative weighting coefficients
5. **Classification Layer:**

   Pass F_{Hybrid}through a fully connected layer → Softmax → Output class probabilities.
6. **Loss Function:**

Use **Cross-Entropy Loss** for multi-class prediction.

7. **Optimization:**
   Use the **Adam optimizer** with a learning rate. 1×10^{-4}.
8. **Training:**
   a. Train both networks simultaneously for 25–30 epochs.
   b. Apply **early stopping** when validation accuracy plateaus.
9. **Evaluation:**
   Compute:
   i. Accuracy
   ii. Precision, Recall, F1-score
   iii. Confusion Matrix
10. **Output:**
    Predicted label $y_{pred} = \mathrm{arg\,max}(F_{Hybrid})$

3.5 Classification and Regularization

The combined probabilities create four different diagnostic categories. The training process uses categorical cross-entropy loss as its primary loss function:

$$\mathcal{L} = -\sum_{k=1}^{4} y_k \log(P_{Hybrid}^{(k)})$$

y_k describes the one-hot encoded ground truth, which the system uses as its reference point.

Batch normalization is used to maintain stability during internal covariate shift.

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

Dropout regularisation introduces stochastic neuron deactivation during training:

$$x' = m \odot x$$

where $m$is a Bernoulli mask. This improves generalisation and reduces overfitting.

3.6 Optimization Strategy

Model parameters are optimised using the Adam optimiser. Parameter updates follow:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where:

- $\eta$is the learning rate
- m_t and v_t are first and second moment estimates

The process of early stopping stops when the validation performance reaches its highest point to avoid creating overfitting issues.

3.7 Model Training and Baseline Comparison

Multiple pre-trained convolutional neural network architectures were trained on the same dataset to create a performance benchmark. These include ResNet-18, EfficientNet-B0, VGG16, DenseNet121, Xception, and MobileNetV3-Large.

All backbone models underwent identical training conditions to guarantee equal testing conditions for their performance evaluation. The independent evaluation of these baseline architectures enabled analysis of classification behaviour, convergence characteristics, and class-wise performance variations.

Based on comparative assessment, DenseNet121,and MobileNetV3-Large demonstrated complementary strengths in feature representation and generalisation stability. Therefore, a dual-model hybrid ensemble integrating these two architectures was constructed as the final proposed MUOPTDL framework.

It is important to note that Xception and other evaluated architectures were utilised solely for baseline comparison and were not incorporated into the final ensemble fusion strategy.

3.8 Model Evaluation

The performance of the model is assessed through the metrics of accuracy, precision, recall, F1-score, macro averages, and confusion matrix analysis. The formulas for precision and recall are given below:

$$Precision = \frac{TP}{TP + FP}, \qquad Recall = \frac{TP}{TP + FN}.$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}.$$

The last model provides outstanding and trustworthy results, generating 100% precision and recall for COVID-19 and lung cancer, and very good F1-scores of 0.98 for pneumonia and 0.95 for normal situations. The hybrid ensemble model, which reached an overall accuracy of 98%, has been shown to be very useful for the multi-class detection of respiratory diseases.

4. Result and Discussion

The section compares different deep learning models with the proposed hybrid ensemble framework, which performs multi-class respiratory disease classification (COVID-19, Lung Cancer, Normal, and Pneumonia). The performance evaluation uses a confusion matrix assessment together with training convergence analysis, ROC/AUC testing, and class-based precision-recall measurement. The research aims to assess how well lightweight architectures and deep architectures achieve generalisation and stability during operation and their ability to handle classification tasks.

4.1. Classification Performance

4.1.1 ResNet-18 Model:

The confusion matrix in Figure 3 unequivocally indicates that the model is performing wonderfully in the four disease categories. Among the COVID-19 cases, there are predictions of 116/116 correct ones, which means the model has a perfect recall and no false positives. The Lung Cancer image prediction means malignant patterns are completely isolated from other thoracic diseases. The Normal cases achieve 298 correct predictions, and only for 18 cases, the misclassification is into Pneumonia, which means there is a very slight visual overlap between normal tissue and mild infection patterns in chest X-rays, which, in a way, is in agreement with early radiological pneumonia being visually subtle.

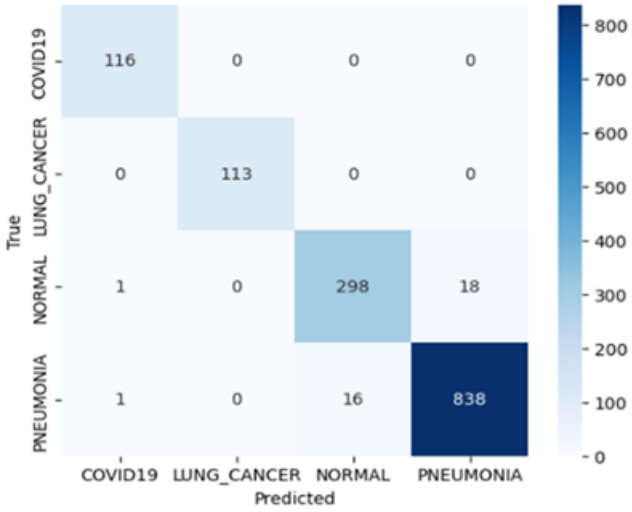

Fig. 3 Confusion Matrix of ResNet–18 Model

4.1.2 Efficient Net – B0 Model

The confusion matrix for EfficientNet-B0 shows in Figure 4 proof of the model's development. The recall for the COVID-19 and Lung cancer classes is again perfect, and this indicates the power of Efficient Net in detecting highly contrasting pathological signs. The normal class reached 311 correct predictions, and thus, 6 cases were incorrectly classified as pneumonia, which is a minor issue. On the other side, pneumonia shows a great reliability with 816 cases accurately classified, although 37 samples were mistakenly classified as Normal; this is a common problem because of the difficult identification of mild pneumonia in X-rays. The darker diagonal blocks and lighter off-diagonal values provide a clear picture of the dense diagonal dominance and also confirm the enhancement of feature extraction by EfficientNet-B0 and its powerful generalisation across all classes.

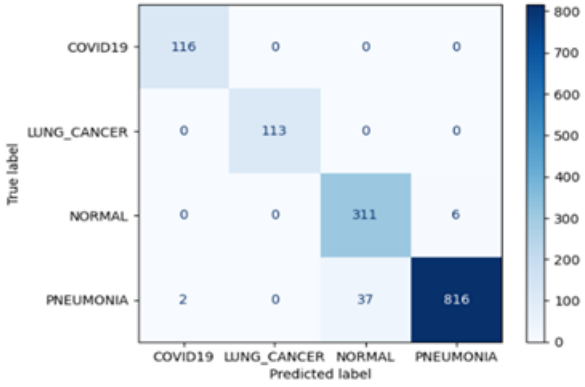

Fig. 4 Confusion Matrix of Efficient Net – B0 Model

4.1.3 MobileNetV3 Model

The accuracy curve for MobileNetV3-Large in the process of training and fine-tuning, displayed in Figure 5, shows a very clear upward trend across all epochs in both the training and validation accuracy. The training accuracy, which was initially at nearly 93%, is progressively going up to more than 99%, while the validation accuracy, which had started at almost 95.8%, is going up to around 97.8%. The difference between training and validation accuracy is very small throughout, which is a sign of good generalisation and very little overfitting. Such a trend points towards the fact that fine-tuning has a huge impact on the model's performance as well as its stability, thus making it a very good choice for a classification problem.

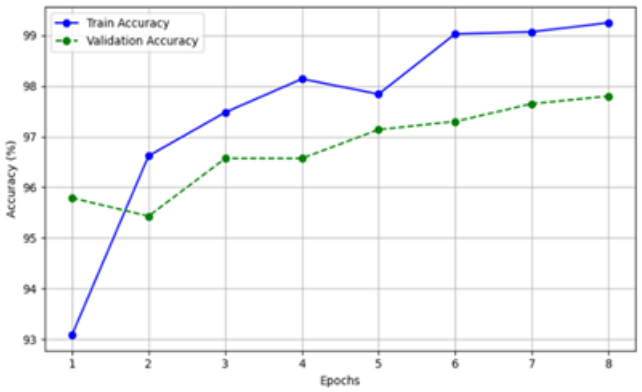

Fig. 5 MobileNetV3 – Large Combined Accuracy

4.1.4 VGG16 Model

The confusion matrix of VGG16 in Figure 6 of COVID-19, in which 98 are correctly identified, 8 are wrongly classified as Normal, and 10 are as Pneumonia, resulting in the lower recall of COVID-19 (~0.84). Lung cancer diagnoses are perfect. The Normal category has 310 correct and 7 misclassified as Pneumonia, which gives a high recall (~0.98) but leads to false positives in other categories, thus lowering precision. Pneumonia has 818 correct and 37 misclassified as Normal; its strong recall (~0.96) is thus affected by some overlap with Normal. In general, the confounding factors show most discrimination difficulties between Normal and Pneumonia, alongside a few missed COVID-19 cases.

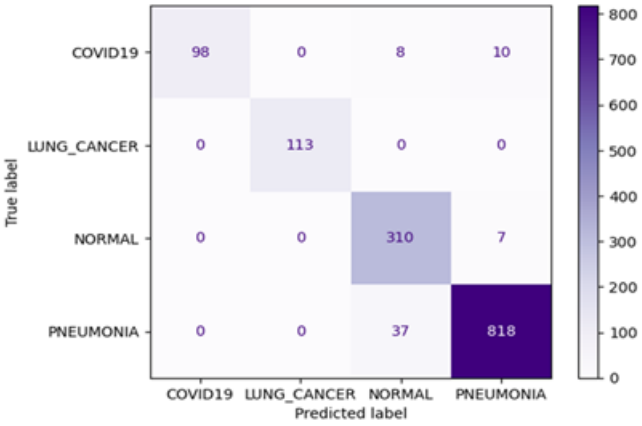

Fig. 6 Confusion Matrix of VGG16 Model

4.1.5 DenseNet121 Mode

The confusion matrix of DenseNet121 in Figure 7 shows the excellent performance in classification in all four classes. The retail of COVID-19 is done perfectly with 116 predictions being correct out of 116, which translates to 100% recall and no false positives. The Lung Cancer detection also achieved outstanding performance by making 113 correct predictions out of 113, thus validating the model's ability to discern malignant patterns with high precision. For the Normal category, the count of right identifications is 300, while 16 instances are incorrectly classified as Pneumonia and 1 as COVID-19, resulting in a recall of about 94.6% and a bit lower precision due to these errors. Pneumonia impressively reveals its presence through 837 correct predictions out of 855, though 18 samples are misclassified as Normal, which is generally a challenge owing to the subtle radiological signs of mild pneumonia.

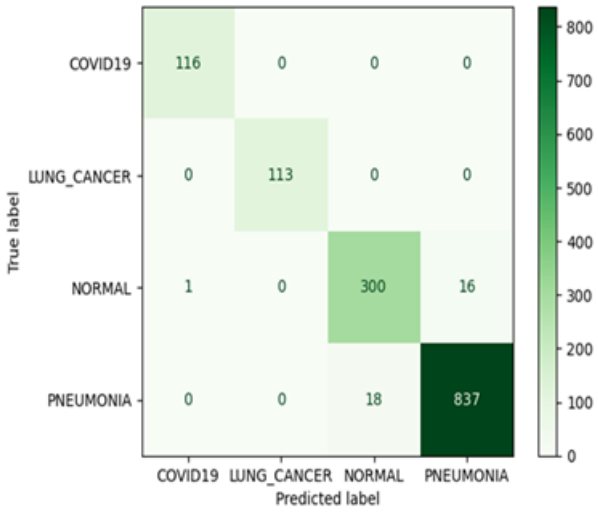

Fig. 7 Confusion Matrix of DenseNet 121 Model

4.1.6 Xception

Figure 8 shows that there are 116 out of 116 correct classifications; for Lung Cancer, there are 113 out of 113 correct. The Normal class has 305 correct, and 12 are misclassified as Pneumonia, thus coinciding with the high recall (~0.962). Pneumonia has 818 correct, with 35 misclassified as Normal and 2 misclassified as COVID-19; this factor is the main source of error for counting, which is in line with its recall (~0.957). In general, Xception confuses Normal with Pneumonia most of the time, but is faultless with COVID-19 and Lung Cancer.

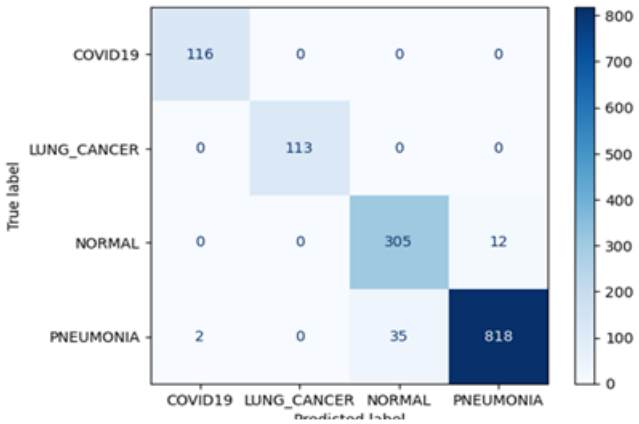

Fig. 8 Confusion Matrix of Xception

4.2 Training Dynamics & Convergence

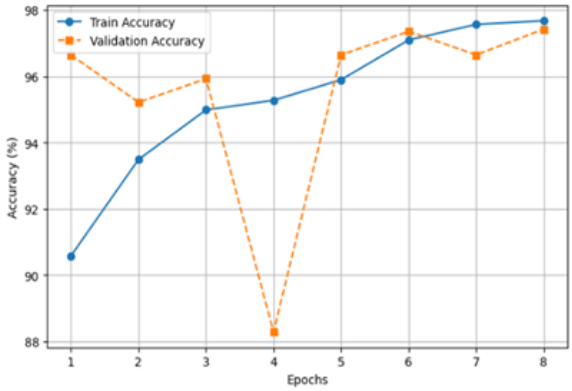

Fig. 9 Training vs Validation Accuracy Curve of ResNet-18

The training and validation accuracy results of the ResNet18 model appear in Figure 9. The training accuracy shows a steady and consistent increase across epochs, indicating effective learning of discriminative features from chest images. The validation accuracy is also on the same trend and is near the training curve throughout most of the epochs showing that it has good ability to generalize. The temporary reduction in the validation accuracy occurs at the middle epochs, which is possible due to the variability of the data or unequal distribution of the classes. The error margin is minimized to a great extent and becomes constant in the later epochs. The values of training and validation accuracy are similar that implies that the overfitting is at minimal rate and the ResNet18 model demonstrates steady learning trends during the entire training process.

The temporary decrease in validation accuracy observed around the fourth epoch is primarily attributed to stochastic mini-batch optimization and the effect of data augmentation applied during training. During the early stages of learning, the model is exposed to diverse augmented samples, resulting in short-term fluctuations in validation performance. Such behavior is common in deep learning models and does not indicate overfitting or optimization instability. As training progresses, the network learns more discriminative feature representations, leading to rapid recovery and stabilization of validation accuracy in subsequent epochs. The close alignment between training and validation accuracy curves in later epochs confirms the robustness and convergence stability of the proposed framework.

4.3 ROC/ AUC Curve

The ROC curve analysis shows exceptional ability to distinguish between four different classes of data. The AUC results for COVID-19 and Lung Cancer show near-perfect values, which demonstrate their ability to separate these conditions from all other disease categories. The Pneumonia and Normal classes show high AUC results, which display slight overlap according to their radiological similarities. The deep feature representations that the tested architectures learned show reliable performance through ROC analysis.

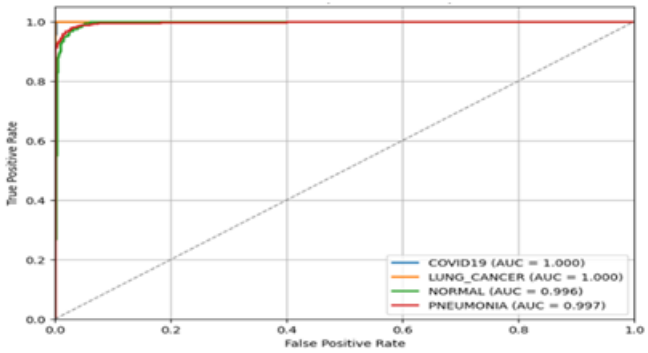

Fig. 10 ROC Curve of Efficient – B0

4.4 Precision Recall Matrix of Hybrid Ensemble And Weighted /Hybrid Model

The hybrid ensemble, which is shown in Table 2, further improves the overall accuracy to 97.43%. The categories of both COVID-19 and Lung Cancer have been given a perfect score (precision, recall, and F1 all = 1.0000), which means that there are no misses in detection, and no false alarms have occurred in these classes. The Normal class has displayed a noticeable precision increase (0.9350 vs 0.8971), while also sustaining very high recall (0.9527), which leads to a balanced F1 of 0.9437—this suggests that the number of incorrectly labelled non-normal cases as Normal in comparison with Xception has decreased. Pneumonia is up in terms of sensitivity with a new recall of 0.9754 (an increase from 0.9567), and also maintains high precision (0.9823), which results in F1 = 0.9789. The macro averages of the ensemble (P 0.9793, R 0.9820, F1 0.9807) and weighted averages (~ 0.9744) confirm that the improved performance has been consistent across all classes, and the major success has been in decreasing Normal↔Pneumonia confusion while keeping COVID-19/Lung Cancer perfect.

Table 2. Precision- Recall Matrix of Hybrid Ensemble

Class	Precision	Recall	F1-Score	Support
COVID19	1.0000	1.0000	1.0000	116
LUNG_CANCER	1.0000	1.0000	1.0000	113
NORMAL	0.9350	0.9527	0.9437	317
PNEUMONIA	0.9823	0.9754	0.9789	855
Accuracy			0.98	1401
Macro Avg	0.9793	0.9820	0.9807	1401
Weighted Avg	0.9745	0.974	0.974	1401

The matrix shown in Table 3 corroborates the metrics of the report: COVID-19 and Lung Cancer are both classified perfectly. The model can identify 303 Normal cases correctly, but it makes 14 false positive diagnoses of Pneumonia. This gives a recall rate of 0.956 and a precision rate of 0.94. The model for Pneumonia corrects 837 cases while 18 cases it diagnoses as Normal, which leads to a recall of 0.979 and a precision of 0.984. The system commits errors only between Normal and Pneumonia because it maintains complete separation between COVID-19 and Lung Cancer, which results in perfect scores for those categories and an overall accuracy of 98%.

Table 3. Precision- Recall Matrix of Weighted Hybrid Model

Class	Precision	Recall	F1-Score	Support
COVID19	1.00	1.00	1.00	116
LUNG_CANCER	1.00	1.00	1.00	113
NORMAL	0.94	0.96	0.95	317
PNEUMONIA	0.98	0.98	0.98	855
Accuracy			0.98	1401
Macro Avg	0.98	0.98	0.98	1401
Weighted Avg	0.98	0.98	0.98	1401

4.5 Discussion

The evaluation results prove that using ensemble learning systems for multi-class respiratory disease classification produces better results than using single deep learning models. DenseNet121 achieves the best-balanced accuracy among the three models because of its dense connectivity design, while MobileNetV3 enables efficient computations with consistent performance during training. The majority of classification errors occur between the

Normal and Pneumonia categories because these two categories share common radiological features, which create difficulty for the system to differentiate between them. The hybrid ensemble system achieves better prediction results because it combines different architectural elements through probability-based fusion, which helps to decrease prediction uncertainty. The system reaches 98% accuracy because it uses different feature types to improve multi-modal diagnostic results without needing synthetic data augmentation or advanced optimisation techniques.

The comparative analysis demonstrates that the proposed MUOPTDL framework achieves competitive performance with recent state-of-the-art methods published during 2024-2025, as shown in Table 4.

Table 4. Comparative Performance Analysis of Proposed MUOPTDL with Existing Lung Disease Detection Models

Reference	Accuracy (%)	Precision	Recall	F1-score	AUC
Kaya et al. (2024) [32]	97.23	—	—	—	—
Chihaoui et al. (2024) [33]	~97	—	—	—	—
Dardouri et al. (2025) [34]	96	96	94	96	—
Angara et al. (2024) [35]	~94.87	—	—	—	—
Proposed MUOPTDL	98.00	98	98	98	99

The proposed framework shows better disease classification performance because it uses multiple disease classification methods, although recent literature uses single-architecture methods and training specific to particular modalities. The results show that hybrid ensemble methods create an effective balance between diagnostic accuracy and operational efficiency, which makes the system appropriate for fast-paced clinical screening settings.

4.6 Training Stability and Generalization Analysis

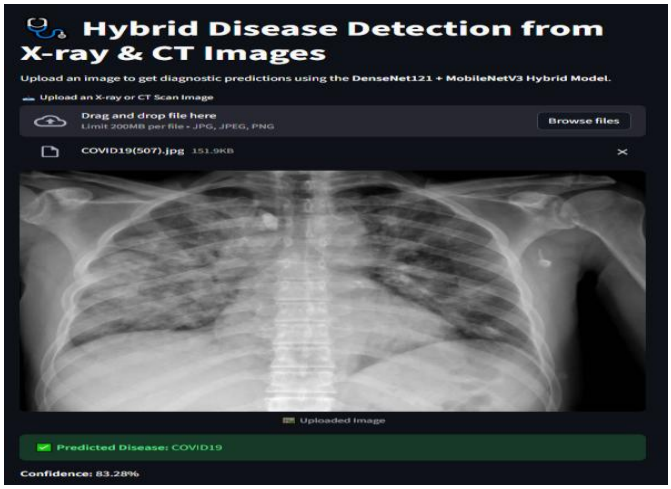

Fig 11: Web-Based Interface Showing Uploaded Chest Image and Predicted Disease Output

Figure 11 and 12 shows the web-based interface developed for real-time disease prediction. The system enables users to upload chest images and obtain predicted labels with confidence scores.

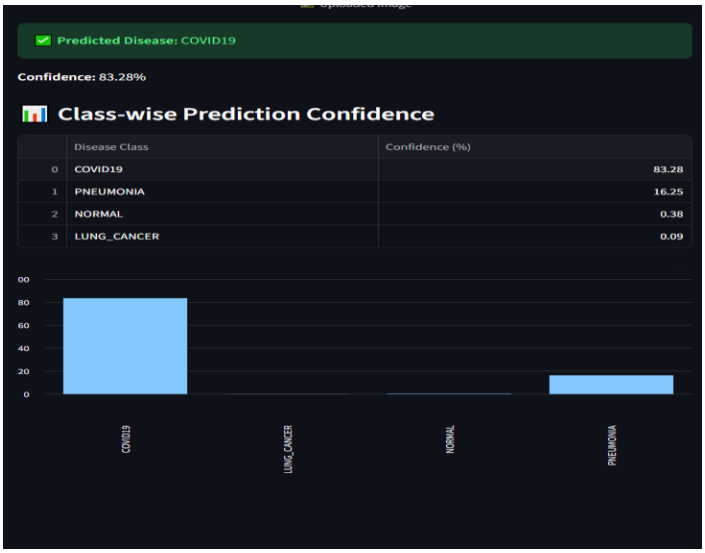

Fig 12: Class-wise Prediction Confidence Visualization of the Proposed Hybrid Model

Conclusion

The integration of multi-scale deep convolutional neural networks (CNNs) with weighted fusion techniques significantly enhances the multi-class classification performance of COVID-19, pneumonia, lung cancer, and normal cases using CT and chest X-ray imaging modalities. The proposed model demonstrates high diagnostic accuracy while effectively addressing class imbalance and adapting to variations in imaging conditions. These capabilities make it a promising approach for real-time clinical applications. Future research can further improve this framework by incorporating three-dimensional (3D) volumetric CT analysis, transformer-based vision architectures, federated learning for privacy-preserving multi-institutional training, and explainable artificial intelligence (XAI) techniques to support clinical interpretation by radiologists. Additionally, integrating electronic health records (EHRs), biological markers, and longitudinal patient imaging data may enhance predictive performance and facilitate the development of fully automated, AI-assisted respiratory disease diagnosis and management systems.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted as part of the authors' academic research activities without external financial support.

Acknowledgments

The authors express sincere gratitude to their respective institutions for providing computational facilities and research support. The authors also thank medical professionals and data contributors whose publicly available datasets facilitated this study. All authors contributed significantly to the conceptualisation, experimentation, analysis, and manuscript preparation of this research work.

References

1. H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. El-Amir, "A review of deep learning algorithms and their applications in healthcare," Algorithms, vol. 15, no. 2, p. 71, 2022.
2. K. Amara, O. Kerdjidj, M. A. Guerroudji, N. Zenati, and N. Ramzan, "Advancements and challenges in CT image segmentation for COVID-19 diagnosis through augmented and virtual reality: A systematic review and future perspectives," Journal of Radiation Research and Applied Sciences, vol. 18, no. 2, p. 101374, 2025.
3. K. Khare, Z. Aizaz, and N. Khare, Fostering Healthcare through Artificial Intelligence: A High-Performance Computing Approach. Boca Raton, FL, USA: CRC Press, 2025.
4. M. Trigka and E. Dritsas, "A comprehensive survey of deep learning approaches in image processing," Sensors, vol. 25, no. 2, p. 531, 2025.

5. C. A. Jayawardena, N. Wedasingha, N. Kolambage, and S. Perera, “Pneumonia detection and lung disease assessment from chest X-rays: Developing a diagnostic support system,” 2025.
6. N. Enshaei, A. Mohammadi, F. Naderkhani, N. Daneman, R. A. Mughli, R. Anconina, and F. H. Berger, “Differentiation of COVID-19 from other types of viral pneumonia and severity scoring on baseline chest radiographs: Comparison of deep learning with multi-reader evaluation,” PLOS ONE, vol. 20, no. 7, p. e0328061, 2025.
7. Y. Qian, R. An, G. Liu, H. Tang, G. Xiao, and D. P. Bavirisetti, “TransFusion: Transfer learning-driven adaptive fusion network for infrared and visible image,” Infrared Physics & Technology, vol. 150, p. 105906, 2025.
8. D. Nguyen and V.-D. Hoang, “Towards enhancing learning on imbalanced data: A novel adaptive weighting strategy,” Neurocomputing, 2026, Art. no. 132886.
9. A. Shaikh, S. Amin, M. A. Zeb, A. Sulaiman, M. S. Al Reshan, and H. Alshahrani, “Enhanced brain tumour detection and segmentation using densely connected convolutional networks with stacking ensemble learning,” Computers in Biology and Medicine, vol. 186, p. 109703, 2025.
10. Z. Cao, J. Li, L. Fang, Z. Li, H. Yang, and G. Dong, “Research on efficient classification algorithm for coal and gangue based on improved MobileNetV3-small,” International Journal of Coal Preparation and Utilisation, vol. 45, no. 2, pp. 437–462, 2025.
11. G. Xu, X. Wang, X. Wu, X. Leng, and Y. Xu, “Development of residual learning in deep neural networks for computer vision: A survey,” Engineering Applications of Artificial Intelligence, vol. 142, p. 109890, 2025.
12. N. Mikami, K. Aizawa, A. Kurihara, and Y. Ueki, “Unsupervised learning-based acoustic detection of gas leakage in liquid: Evaluation of noise resistance based on parametric ROC analysis,” AI Thermal Fluids, p. 100029, 2026.
13. M. Imani, M. Joudaki, A. Bagheri, and H. R. Arabnia, “Why ROC-AUC is misleading for highly imbalanced data: In-depth evaluation of MCC, F2-score, H-measure, and AUC-based metrics across diverse classifiers,” Technologies, vol. 14, no. 1, p. 54, 2026.
14. World Health Organisation, Clinical Management of COVID-19: Living Guideline, WHO/2019-nCoV/clinical/2023.2, Geneva, Switzerland, 2023.
15. M. M. Hasan, M. U. Islam, M. J. Sadeq, W.-K. Fung, and J. Uddin, “Review on the evaluation and development of artificial intelligence for COVID-19 containment,” Sensors, vol. 23, no. 1, p. 527, 2023.
16. C. Zhang, S. Bengio, and Y. Singer, “Are all layers created equal?” Journal of Machine Learning Research, vol. 23, no. 67, pp. 1–28, 2022.
17. A. T. Abdulahi, R. O. Ogundokun, A. R. Adenike, M. A. Shah, and Y. K. Ahmed, “PulmoNet: A novel deep learning-based pulmonary disease detection model,” BMC Medical Imaging, vol. 24, no. 1, p. 51, 2024.
18. M. Kolhar, A. M. Al Rajeh, and R. N. A. Kazi, “Augmenting radiological diagnostics with AI for tuberculosis and COVID-19 disease detection: Deep learning detection of chest radiographs,” Diagnostics, vol. 14, no. 13, p. 1334, 2024.
19. S. Dardouri, “A deep convolutional neural network model for lung disease detection using chest X-ray imaging,” Pulmonary Medicine, vol. 2025, no. 1, p. 6614016, 2025.
20. S. Karimullah, M. Khan, F. Shaik, B. Alabduallah, and A. Almjally, “An integrated method for detecting lung cancer via CT scanning via optimisation, deep learning, and IoT data transmission,” Frontiers in Oncology, vol. 14, p. 1435041, 2024.
21. A. Whata, K. Dibeco, K. Madzima, and I. Obagbuwa, “Uncertainty quantification in multi-class image classification using chest X-ray images of COVID-19 and pneumonia,” Frontiers in Artificial Intelligence, vol. 7, p. 1410841, 2024.
22. S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, “Image data augmentation for deep learning: A survey,” arXiv preprint, arXiv:2204.08610, 2022.
23. W. Nagaye, Y. Afriyie, and S. B. Egala, “An improved hide-and-seek augmentation technique for image classification using convolutional neural networks,” Journal of Electrical Systems and Information Technology, vol. 13, no. 1, p. 2, 2026.
24. N. Sriwiboon, “Efficient and lightweight CNN model for COVID-19 diagnosis from CT and X-ray images using customised pruning and quantisation techniques,” Neural Computing and Applications, vol. 37, no. 18, pp. 13059–13078, 2025.
25. S. Kumar and B. Bhowmik, “EffiCOVID-net: A highly efficient convolutional neural network for COVID-19 diagnosis using chest X-ray imaging,” Methods, vol. 240, pp. 81–100, 2025.
26. S. H. Chowdhury and Y. Kutlu, “Deep learning-based multi-class classification of lung diseases on chest X-ray: A comparative study,” Journal of Intelligent Systems with Applications, pp. 1–10, 2025.
27. R. Rampriya, R. S. Reddy, K. L. Prasad, H. K. Kumar, and R. V. Raja, “Diagnosis and risk analysis of pulmonary disorder in CXR images using optimised deep learning techniques,” in Proc. 6th Int. Conf. Inventive Research in Computing Applications (ICIRCA), IEEE, pp. 1221–1226, 2025.
28. M. V. Sanida, T. Sanida, A. Sideris, and M. Dasygenis, “An advanced deep learning framework for multi-class diagnosis from chest X-ray images,” J, vol. 7, no. 1, pp. 48–71, 2024.
29. G. Mohan, M. M. Subashini, S. Balan, and S. Singh, “A multiclass deep learning algorithm for healthy lung, COVID-19 and pneumonia disease detection from chest X-ray images,” Discover Artificial Intelligence, vol. 4, no. 1, p. 20, 2024.
30. R. Kumar, C.-T. Pan, Y.-M. Lin, Y.-L. S., T.-S. Chung, and U. G. S. Janesha, “Enhanced multi-model deep learning for rapid and precise diagnosis of pulmonary diseases using chest X-ray imaging,” Diagnostics, vol. 15, no. 3, p. 248, 2025.

31. P. Kalaivani, A. Dhivya, G. Dharani, S. Bharathi, and C. Rajan, “Deep learning with multi-class classification for detection of COVID-19 and pneumonia,” in Proc. Int. Conf. Inventive Computation Technologies (ICICT), IEEE, pp. 287–292, 2023.
32. M. Kaya and Y. Cetin-Kaya, “A novel ensemble learning framework based on a genetic algorithm for the classification of pneumonia,” Engineering Applications of Artificial Intelligence, vol. 133, p. 108494, 2024.
33. M. Chihaoui, N. Dhibi, and A. Ferchichi, “Optimisation of convolutional neural network and visual geometry group-16 using genetic algorithms for pneumonia detection,” Frontiers in Medicine, vol. 11, p. 1498403, 2024.
34. S. Dardouri, “Pneumonia detection in X-ray chest images based on convolutional neural networks and data augmentation methods,” in Proc. ICT4AWE, pp. 165–172, 2025.
35. S. Angara, N. R. Mannuru, A. Mannuru, and S. Thirunagaru, “A novel method to enhance pneumonia detection via a model-level ensembling of CNN and vision transformer,” arXiv preprint, arXiv:2401.02358, 2024.
36. D. Li, H. Huo, S. Jiao, X. Sun, and S. Chen, “Automated thorax disease diagnosis using multi-branch residual attention network,” Scientific Reports, vol. 14, no. 1, p. 11865, 2024.
37. B. Zhu, L. Chen, X. Li, S. Zhao, S. Yu, and Q. Sun, “ADFF-Net: An attention-based dual-stream feature fusion network for respiratory sound classification,” Technologies, vol. 14, no. 1, p. 12, 2025.
38. T. Dalvi, M. Patley, V. Bhatt, K. Kadam, S. Gite, and B. Pradhan, “Pneumonia detection using chest X-rays: A comparative study of CNN architectures and explainable AI techniques,” in Advancements in IoT Sensors and Security: Harnessing AI/ML for Secured IoT Sensor Data, Cham, Switzerland: Springer Nature, pp. 219–245, 2026.
39. S.-C. Lin, J.-N. Hwang, and S.-H. Lai, “MCMIAD: Multi-class model for medical image anomaly detection,” in Medical Imaging with Deep Learning, 2026.
40. M. Owda, A. Abumihsan, A. Y. Owda, and M. Abumohsen, “A lightweight hybrid deep learning model for tuberculosis detection from chest X-rays,” Diagnostics, vol. 15, no. 24, p. 3216, 2025.