International Journal of Applied Mathematics
Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DESIGN AND IMPLEMENTATION OF DEEP LEARNING-BASED MODEL FOR LUNG DISEASE DETECTION AND CLASSIFICATION

Preeti Sharma[1], Devershi Pallavi Bhatt[2*]

[1,2] Department of Computer Science and Applications Manipal University Jaipur, Dehmi Kalan, Jaipur- 303007, India.
[1]preetisharma.bca@gmail.com [2*]devershipallavi.bhatt@jaipur.manipal.edu

Abstract

The limited availability of sufficiently labeled medical images poses a major challenge in developing reliable deep learning systems for lung disease diagnosis. The novel idea is to train popular pretrained convolutional neural network (CNN) models, such as VGG16, ResNet50, and DenseNet121, on subsets of different sizes of chest X-ray datasets and develop a multiclass classification model. The dataset includes images of five different lung diseases, i.e., lung tumor, bacterial pneumonia, viral pneumonia, normal, and COVID, sourced from Kaggle. Rotation, flipping, and brightness adjustment augmentation methods helped in creating a balanced dataset of 20,000 CXR images, each class having around 4000 images. This study demonstrates that the impact of image augmentation methods (balanced dataset), task-specific training of pre-trained models with a custom classification head, and optimization methods helped in achieving optimized outcomes. To ensure the outcomes are reliable, the experiments are done three times, and all models are trained on the same hyper-parameters. The advancement in accuracy from 73.6% (500 images) to 91.41% (20,000 images) exemplifies scalability for data-limited settings. This mathematically grounded study offers suggestions for model preference in resource-constrained healthcare, enabling deployment in low-resource clinics.

Keywords: chest x-ray, deep learning, lung diseases, medical imaging, convolutional neural network.

1. Introduction

The adverse effects of lung diseases have been a serious issue for many years. Diseases like pneumonia, tuberculosis (TB), cancer, etc., can take a severe form, causing an increase in fatal cases. Recently, COVID-19 was declared an epidemic that has caused worldwide casualties. A thorough detection of these diseases is possible with medical imaging modalities such as CT scans, X-rays, and ultrasound. The X-ray imaging modality is an old medical imaging (MI) technique, widely used because of its exceptionally low cost and easy availability [8]. In recent years, Artificial Neural Networks (ANN), specifically Deep learning (DL) models, have shown remarkable progress in the medical image classification field. DL models automatically learn complex patterns from images and provide excellent results in the detection of lung diseases on CXR [3].

Convolutional Neural Network (CNN) is a type of DL model that is considered the best for the classification of images [4]. Large-scale datasets with multiple classes serve as the training ground for some CNN models. These pre-trained models can be trained on new datasets to find new classes, offering the concept of transfer learning (TL). TL has been the subject of many research works in the medical image classification domain [5]. The significant issues with training CNN models on medical image classification are the shortage of well-labelled and unbalanced datasets specifically for multi-class classification problems. Studies have applied CNN to CXR datasets for multiclass classification using TL, but the major points stood the selection of an appropriate pre-trained model on a small dataset, while considering issues of an unbalanced dataset, model parameter counts for training time, and hardware requirements (RAM, GPU, parameters, training time, etc).

Received: August 10, 2025

Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

To overcome these issues, this work develops a CNN for multi-class classification of lung diseases utilizing the CXR dataset. The main contributions of this work are: (i) preparation of a balanced dataset through augmentation (rotation, flipping, brightness adjustment) [16]; (ii) application of TL with VGG16, ResNet50, and DenseNet121, each with a custom classification head; (iii) evaluation of scalability by training on subsets of increasing sizes (500–20,000 images); (iv) assessment under consistent training configurations with performance metrics such as accuracy, specificity, F1 score, ROC_AUC(v) comparative evaluation with prior studies using DenseNet121.

2. Related Works

A number of related studies highlight both the potential and the challenges of transfer learning (TL) in lung disease classification. For example, authors [8] used DenseNet121 on CXR scans for lung cancer classification but pointed out the lack of a reliable dataset. ResNet18-based CNN model trained on 502 CXRs covering COVID, bacterial and viral pneumonia, TB, and normal cases achieved 88.9% accuracy [5]. Authors [9] examined six pre-trained models—including VGG16, Xception, ResNet50, and EfficientNetV2L—on 5,856 CXR images. Pneumonia detection was the main focus, and the best performance came from EfficientNetV2L with an accuracy of 94.4%.

Authors [10] utilized VGG19 model as the base model, followed by multiple CNN blocks that performed feature extraction. In this research work, the CXR images for all diseases were of different counts, with a minimum count of 3615 images of COVID and a maximum count of 20,000 of lung cancer, 1400 of TB, 5870 of pneumonia, and 10,192 of normal cases. The accuracy achieved by the proposed model was 96.48%. The authors suggested that TL can achieve higher accuracy in classifying chest diseases from chest X-rays. Authors [11] utilized 5,949 CXRs were to train VGG16, DenseNet121, and DenseNet201 for the classification of bacterial pneumonia, viral pneumonia, COVID, and normal cases. The COVID CXRs used were 76, while bacterial pneumonia CXRs were 2540, and normal and viral pneumonia CXRs were 1349 and 1355 showcases class imbalance problem. The accuracy achieved by VGG16, DenseNet121, and DenseNet201 is 79%, 78%, and 83%, respectively.

Looking across these works, one pattern becomes clear: while TL helps improve classification accuracy, most studies either rely on relatively small datasets or struggle with class imbalance. Moreover, there is little systematic analysis of how different dataset sizes influence the performance of commonly used pre-trained models. This is the specific gap that our work addresses. We evaluate three widely adopted architectures—VGG16, ResNet50, and DenseNet121—on balanced CXR datasets of gradually increasing size (500 to 20,000 images). By doing so, the study offers new insights into how data volume interacts with model capacity and helps determine which CNNs are better suited for multiclass lung disease classification.

3. Methodology

3.1 Dataset preparation and preprocessing methods. A comprehensive chest X-ray (CXR) dataset was constructed by incorporating four distinct types of CXR datasets from Kaggle. The dataset combines public CXR sources:

- "COVID-19 Radiography Database" [13]: 3,616 COVID-19 CXRs
- "Chest X-Ray Images (Pneumonia)" [14]: 2,530 CXRs of bacterial pneumonia and 1,345 CXRs of viral pneumonia.
- "Tuberculosis (TB) Chest X-ray Database" [15]: 1,341 normal CXRs.
- "NIH Chest X-ray dataset" [29]: 3500 lung nodule CXRs.

Augmentation (rotation ±15, horizontal flipping, brightness ±10%) balances to 20,000 images (~4,000 per class). Metadata (age, gender, symptoms, smoking) is simulated where unavailable.

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Preprocessing includes resizing to 224x224 (aligning with input requirements of standard CNN architectures such as VGG and ResNet). For ensuring consistent numerical stability during training, pixel intensities were scaled to the [0,1] range by dividing by 255 (normalization). To improve the visibility of subtle pathologies such as opacities and nodules, image contrast was enhanced by histogram equalization. To evaluate the effect of dataset size on model performance, we created seven training subsets of increasing sizes: 500, 1,000, 5,000, 10,000, 15,000, and 20,000 samples.

Each subset was randomly selected from the full dataset while maintaining class balance. The splitting of the datasets was fixed at 75% for training and 25% for testing. **Figure 1** demonstrates the workflow of the proposed study.

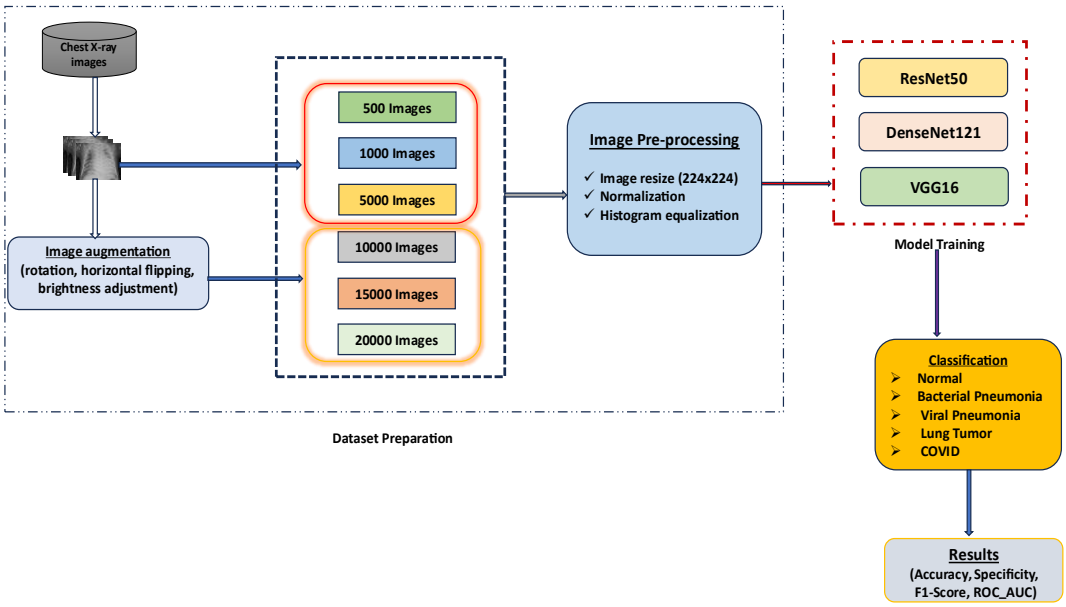

Figure 1. Workflow of proposed research work

3.2 Model Architecture. We propose three transfer learning-based CNN models—VGG16 (16 layers, 138M parameters), ResNet50 (50 layers, 25M parameters), and DenseNet121 (121 layers, 8M parameters)—for multi-class lung disease classification (bacterial pneumonia, viral pneumonia, lung tumors, COVID-19, normal) on a balanced 20,000-image CXR dataset. Selected for efficacy in medical imaging, VGG16 uses 3×3 filters, ResNet50 employs residual connections to mitigate vanishing gradients, and DenseNet121 leverages dense connectivity for feature reuse. Pre-trained on ImageNet, models are fine-tuned by freezing early layers and unfreezing top layers (13 for VGG16, 20 for ResNet50/DenseNet121). A custom classification head includes GlobalAveragePooling2D (GAP), Batch Normalization (BN), a 256-unit Dense layer with ReLU, a Dropout layer (rate 0.5), and a 5-unit softmax output. Optimization uses sparse categorical cross-entropy, ensuring computational efficiency for resource-constrained medical diagnostics. Figure 2 shows the VGG16 model with custom classification head, frozen and unfrozen layers.

$GAP(Z) = (1/(H * W)) * \Sigma_{i=1 \text{ to } H} \Sigma_{j=1 \text{ to } W} Z_{i,j}$, where $Z \in \mathbb{R}^{(H \times W \times C)}$

$BN(x) = \gamma * ((x - \mu) / \sqrt{(\sigma^2 + \varepsilon)}) + \beta$, where $\varepsilon = 10^{-3}$, momentum = 0.99

$Dense(x) = \sigma(Wx + b)$, where $\sigma(x) = \max(0, x)$, $x \in \mathbb{R}^{256}$

$Output(x) = \text{softmax}(W_o x + b_o)$, where $x \in \mathbb{R}^5$

$L = (-1/N) * \Sigma_{i=1 \text{ to } N} \log P(y_i \mid X_i; \theta)$

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

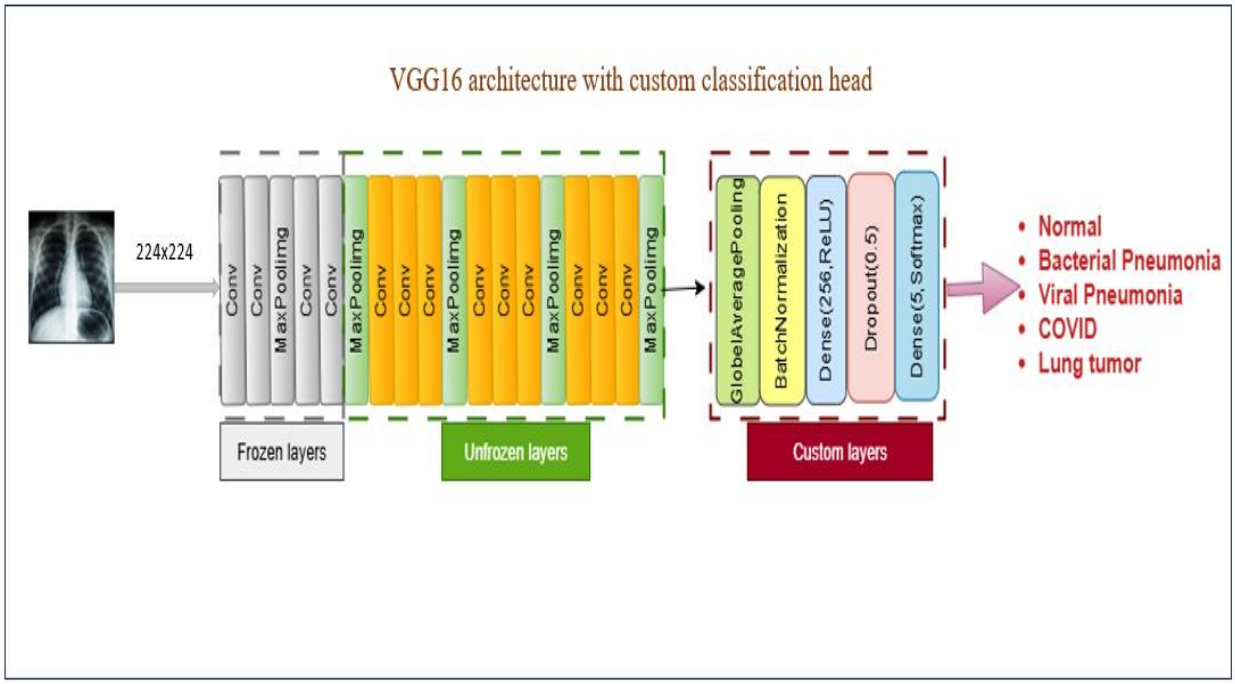

Figure 2. VGG16 Architecture showing frozen, unfrozen layer and custom classification head.

3.3 Model Training. The parameters for model training are discussed in this section; each parameter played a crucial role in model performance. Each parameter has various types, the right one for proposed study is chosen by reviewing various previous studies [7][8][9][10] and online resources. The parameters are:

- Loss function: sparse categorical cross-entropy loss function.

$L = (-1/N) * \Sigma_{i=1 \text{ to } N} \log P(y_i \mid X_i; \theta)$, where $y_i \in \{0,1,2,3,4\}$, $X_i \in \mathbb{R}^{(224\times224\times1)}$

This function calculates the loss by comparing the integer labels with the predicted probabilities for the five classes. This minimizes the negative log likelihood and enhances classification accuracy.

- Optimizer: Adam optimizer (learning rate 1e-4, $\beta1=0.9$, $\beta2=0.999$, $\varepsilon=1e-8$)

The Adam optimizer speeds up the training of the neural network by adjusting the model's weights based on a combination of past gradient information. The parameters $\beta1$, $\beta2$, and ε are hyperparameters that control the behavior of the Adam optimization algorithm.

- Training epochs: 30 (early stopping with patience of 7)
- Batch size : 32 (suitable for small to large datasets)
- Implementation platform: Python 3.11 Libraries: TensorFlow/Keras 2.x.

4. Results and Discussion

Performance metrics help evaluate and understand the performance of a deep learning model, thereby enabling quantitative measurement. 4.1 Performance metrics present a clear picture of the system's learning. With these measurements, we can decide whether the mode is usable or not. Accuracy, precision, recall, and f-measures are some examples of performance metrics. To calculate these parameters, the first four values are calculated first: TP, TN, FP, and FN. TP are True Positive. To represent true negative cases, TN is used. FP represents False Positive, meaning cases that are not positive but are recognized as positive. The last term is FN, which stands for False Negative, meaning cases that are actually positive but are incorrectly recognized as negative. All three models were trained and tested on the same dataset and hyperparameters to observe the performance of models in

Received: August 10, 2025

Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

identifying correct classes. Different accuracies achieved by these models showcase that each model learned the patterns in different ways, and the same hyperparameter tuning can't help each model to get higher results.

Our major focus was to analyze these models on our customized dataset and achieve higher metric values, which were previously attained on 20,000 images. Table 1 showcases the results achieved by three models with custom heads, showcasing that out of all three models, VGG16 outperformed with an accuracy of 91.41% and an AUC of 98.85%, and DenseNet121 achieved moderate results with an accuracy of 85.49% and an AUC of 98.17%. ResNet50 performed low on all datasets compared to VGG16 and DenseNet121, with an accuracy of 77.50% and an AUC of 96.24%.

Table 1. Accuracy (%) of the proposed model on 20,000 CXRs using DenseNet121, VGG16 and ResNet50.

Performance metrics	DenseNet121	ResNet50	VGG16
Accuracy	85.49	77.50	**91.41**
Specificity	96.36	94.35	97.86
F1-score	85.46	77.54	91.38
ROC_AUC	98.17	96.24	98.85

4.1 Model Behavior Across Dataset Scales. We prepared even subsets of datasets in sizes of 500, 1000, 5000, 10000, 15000, and 20000 CXR images of five classes. Augmentation methods helped in achieving the required number of images. The table shows the classification accuracy of each model across increasing dataset sizes. At first the models were trained on 500 images to examine how these models perform on a small dataset. The experiments were performed multiple times to get the highest accuracy. VGG16 achieved 73.60% accuracy, which is good for a small dataset, while ResNet50 did not achieve accuracy >=50%. The table clearly shows that accuracy and other parameters increase with dataset sizes. Out of all image counts, the highest accuracy is achieved by VGG16 (91.41%) on 20,000 CXR, while DenseNet121 achieved a moderate result of 83.97%. The lowest accuracy on 20,000 images is achieved by ResNet50, i.e., 77.50% accuracy, which is 47.10% higher than the accuracy achieved on 500 images. This analytical study may help researchers in finding the proper pre-trained model for the available dataset size equal to or less than 20000 images. The epoch size was set to 30 for all studies, as it's suitable for all sizes of datasets. Table 2 highlights satisfactory results with less hyperparameter tuning. As hyperparameters were set, taking consideration of all dataset sizes, if set up on different parameters with various experiments on different optimizers, learning rates, and batch and epoch sizes, results could be more optimized and precise.

Table 2. Performance metrics of proposed models on varied sizes of CXR datasets.

Image counts	Model	Accuracy	Specificity	F1-score	ROC_AUC
500	DenseNet121	61.60	61.66	61.66	90.70
	ResNet50	30.40	82.38	17.15	83.00
	VGG16	73.60	93.41	73.59	95.31
1,000	DenseNet121	62.40	90.60	61.97	91.04
	ResNet50	59.20	89.78	59.23	88.04
	VGG16	70.40	92.60	70.28	94.69
	DenseNet121	79.12	94.80	79.17	96.48

Received: August 10, 2025

Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5,000	ResNet50	70.32	92.59	70.50	93.27
	VGG16	88.32	97.07	88.36	97.77
10,000	DenseNet121	82.12	95.53	82.14	97.36
	ResNet50	74.88	93.73	74.74	95.32
	VGG16	89.20	97.31	89.14	98.28
15,000	DenseNet121	83.97	95.98	83.97	97.67
	ResNet50	77.22	94.30	77.22	96.01
	VGG16	90.34	97.58	90.34	98.48
20,000	DenseNet121	85.49	96.36	85.46	98.17
	ResNet50	77.50	94.35	77.54	96.24
	VGG16	91.41	97.86	91.38	98.85

Proposed models were run three times on each dataset to analyse the learning behavior of models on datasets. Different performance metrics achieved by models on various runs is shown in Figure 3 in a graphical depiction. The results achieved by VGG16 are higher compared to other two models on RUN 1 to 3 in all dataset sizes. DenseNet121 performed well by achieving moderate results. ResNet50 failed to achieve higher results in all cases. At final, the best results achieved from RUN1, RUN2 and RUN3 were selected as final results achieved by proposed study.

4.2 Performance comparison with previous study. Manokaran et al. (2021) proposed a DenseNet201-based CNN model to find normal, COVID, and pneumonia classes on a CXR dataset of 8644 images. The training dataset included 644 COVID-19 images and 4000 for normal and pneumonia classes. The test dataset consisted of 1729 CXRs, with 800 images for both normal and pneumonia cases and 129 COVID cases. The DenseNet201 architecture was used as a backbone for TL with parameter tuning. The accuracy achieved by this study was 92.19% with a sensitivity of 94% for COVID cases.

To outperform these results, we created a balanced dataset of 9,999 (each class had 3,333 samples in total) of normal, COVID, and pneumonia from our proposed dataset and distributed it in 7499 images for training and 2500 CXR images for testing. We trained the mentioned models on our datasets and evaluate it on the test dataset. The parameters were set the same as defined in the previous study, with the Adam optimizer (0.0001 learning rate), a batch size of 32, and training for 30 epochs. The model was run for 10 epochs on the test dataset.

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

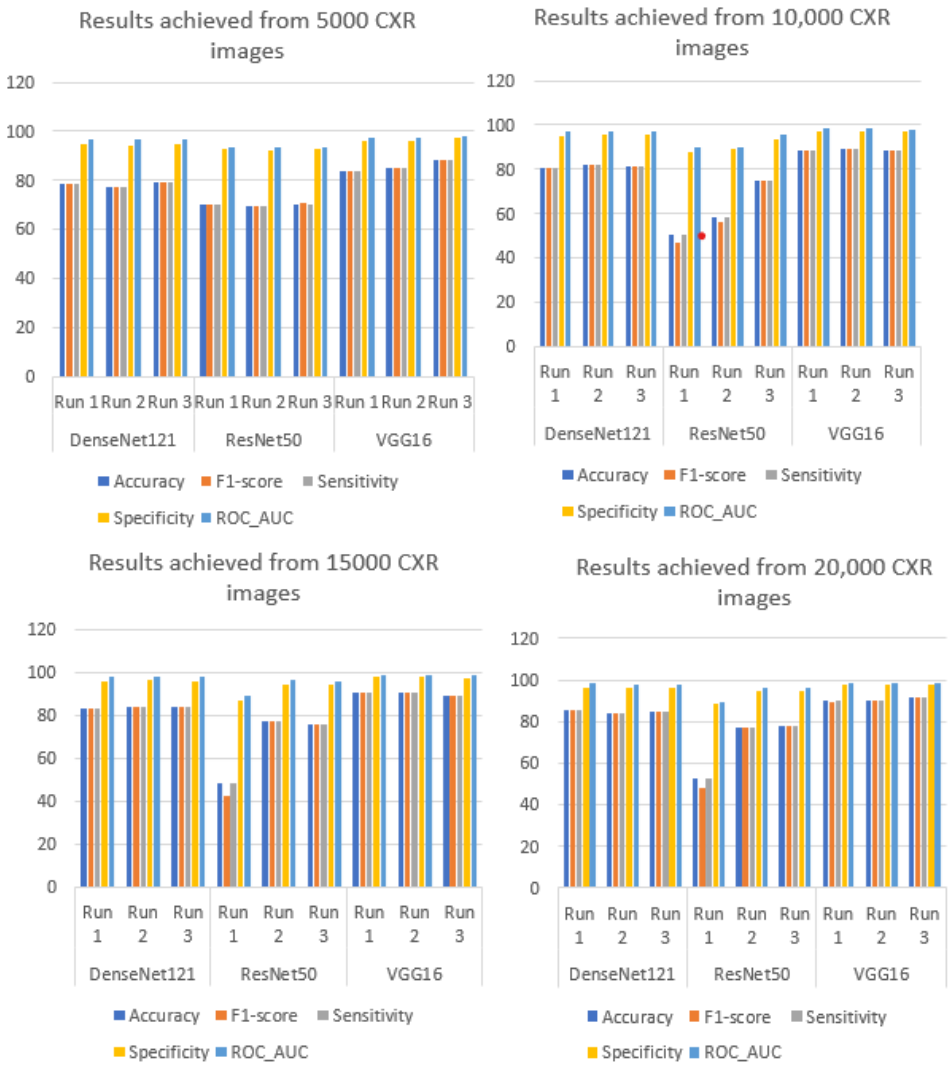

Figure 3. Accuracy trend of models on different datasets.

Table 3 highlights the results achieved by previous study and proposed study, where models were kept same but dataset is different. Propose study with balanced dataset achieved **97.88% accuracy**, with a COVID-19 sensitivity of **99.51%** compared to previous study (94%). While their model showed higher specificity for COVID-19 (99%), our approach achieved superior overall sensitivity and F1-scores across all three classes, showing a more robust classifier for real-world multi-class diagnosis. The sensitivity and ROC-AUC results achieved for pneumonia and COVID is presented in Figure 4. This analysis highlights that a balanced and well-prepared dataset can achieve optimized results compared to an imbalanced dataset. This study was done to prove the effect of balanced dataset in performance of CNN model training. This three class classification model can be further improved with more fine tuning and training on large datasets.

Table 3. The comparative analysis of results achieved by both studies,

Study	Accuracy	Sensitivity	Specificity	F1 score
Manokaran et al.	92.19%	94.00%	99.00%	90.00%
Proposed study	97.88%	97.92%	98.95%	97.89%

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 7s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

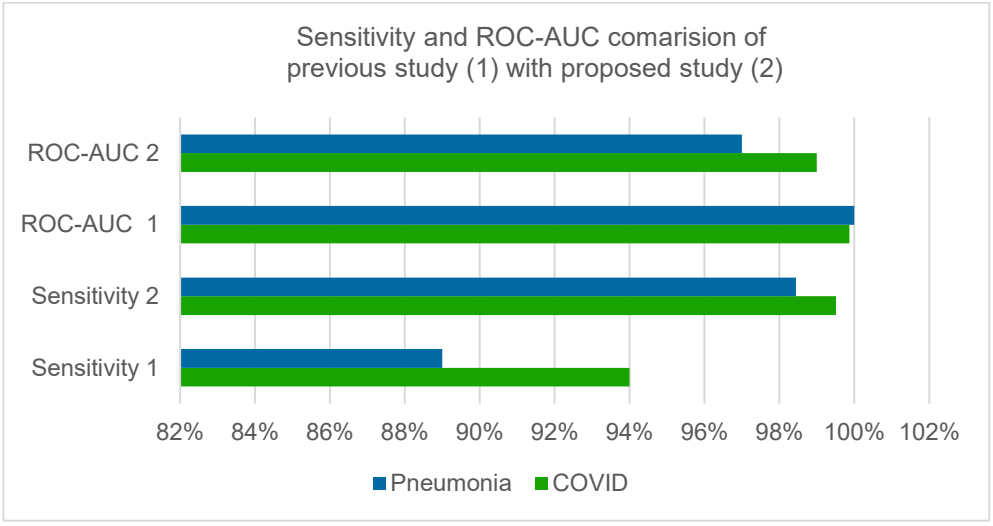

Figure 4. Sensitivity and ROC-AUC comparison for pneumonia and COVID.

Conclusion and future scope

This study assesses the performance of pre-trained Convolutional Neural Networks VGG16, ResNet50, and DenseNet121 in classifying five lung disease categories (bacterial pneumonia, viral pneumonia, COVID, lung tumor, and normal) across CXR dataset sizes of 500, 1,000, 5,000, 10,000, 15,000, and 20,000 images. The lack of medical images, specifically CXR images of diseases like lung tumors, has been a major concern in developing a well-trained classification model. VGG16 achieved the highest accuracy of 91.41% on 20,000 images, followed by DenseNet121 (85.49%) and ResNet50 (77.50%). On the smallest dataset (500 images), VGG16 attained 73.60% accuracy, surpassing ResNet50 (30.40%) and DenseNet121 (61.60%), showing its reliability for small datasets. This study shows that transfer learning with VGG16, ResNet50, and DenseNet121 enhances multi-class lung disease classification, with VGG16 achieving the best performance (91.41% accuracy) on larger datasets (≥5,000 images), guiding model selection for medical imaging applications.

The accuracies achieved by the models show that these models require aggressive fine-tuning to achieve more optimized results by fine-tuning hyperparameters (e.g., learning rate, batch size) and performance and mitigating potential overfitting in the future. Also, proposed research could expand the classification to include other thoracic diseases, such as tuberculosis and chronic obstructive pulmonary disease, and increase dataset sizes to evaluate models.

References

[1] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Inform. Med. Unlocked*, vol. 20, p100391, 2020.

[2] T. Rahman *et al.*, "TB-CXRNet: Tuberculosis and drug-resistant tuberculosis detection technique using chest X-ray images," *Cogn. Comput.*, pp. 1–20, Jan. 2024.

[3] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p103792, Jun. 2020.

[4] R. Siddiqi, "Automated pneumonia diagnosis using a customized sequential convolutional neural network," in *Proc. 3rd Int. Conf. Deep Learn. Technol.*, Jul. 2019, pp. 64–70.

[5] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.

[6] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, "Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI," *Comput. Biol. Med.*, vol.150, p106156, Nov. 2022.

Received: August 10, 2025

[7] M. Woźniak, D. Połap, G. Capizzi, G. L. Sciuto, L. Kośmider, and K. Frankiewicz, "Small lung nodules detection based on local variance analysis and probabilistic neural network," *Comput. Methods Programs Biomed.*, vol. 161, pp. 173–180, Jul. 2018.
[8] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic lung cancer prediction from chest X-ray images using the deep learning approach," in *Proc. 11th Biomed. Eng. Int. Conf. (BMEiCON)*, Nov. 2018, pp. 1–5.
[9] M. Ali *et al.*, "Pneumonia detection using chest radiographs with novel EfficientNetV2L model," *IEEE Access*, vol. 12, pp. 34678–34690, Feb. 2024.
[10] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Eng. J.*, vol. 64, pp. 923–935, Feb. 2023.
[11] Chouat et al., "Lung Disease Classification in X-ray Images using a Modified Deep Convolutional Neural Networks," *Journal of Medical Imaging*, 2023.
[12] I. Sirazitdinov *et al.*, "Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database," *Comput. Elect. Eng.*, vol. 78, pp. 388–399, Sep. 2019.
[13] T. Rahman, "COVID-19 Radiography Database," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database
[14] P. T. Mooney, "Chest X-ray Pneumonia," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
[15] T. Rahman, "Tuberculosis (TB) Chest X-ray Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset
[16] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscip. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122.
[23] D. P. Bhatt, V. Bhatnagar, and P. Sharma, "Meta-analysis of predictions of COVID-19 disease based on CT-scan and X-ray images," *J. Interdiscip. Math.*, vol. 24, no. 2, pp. 381–409, Mar. 2021.
[24] P. Sharma and D. P. Bhatt, "Importance of deep learning models to perform segmentation on medical imaging modalities," in *Data Eng. Smart Syst.: Proc. SSIC 2021*, Singapore: Springer, 2022, pp. 593–603.
[25] P. Sharma and D. P. Bhatt, "COVID-19 identification on chest X-rays with deep learning technique," in *Proc. Data Anal. Manag.: ICDAM 2021*, vol. 2, Singapore: Springer, 2022, pp. 113–123.
[26] P. Sharma and D. P. Bhatt, "Importance of deep learning models in the medical imaging field," in *Approaches Appl. Deep Learn. Virtual Med. Care*, Hershey, PA, USA: IGI Global, 2022, pp. 1–23.
[27] Sakirin, T., & Said, R. B. (2025). Application of deep learning and transfer learning techniques for medical image classification. *Edraak*, *2025*, 38-46.
[28] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, "Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI," *Comput. Biol. Med.*, vol. 150, p106156, Nov. 2022.
[29] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2097-2106).

Received: August 10, 2025