Computational Design: Journal of Modern Informatics

Journal of MEDICAL INFORMATICS

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

Comparative Analysis of DenseNet121, ResNet50, and MobileNetV2 for Lung Disease Detection in Chest X-ray Images Using Convolutional Neural Networks

Yogi Rifaldo[1], Anggi Gunawan[2], Nadia[3]

[1,2] *State Senior High School No. 1 Panti, Pasaman Regency, West Sumatra, Indonesia*
[3] *State Vocational High School No. 1, Solok, West Sumatra, Indonesia*

(Received :July 21, 2026; Revised: July 25, 2026; Accepted: September 8, 2026; Available online: September 30, 2026)

Abstract

Chest X-ray imaging remains one of the most widely used radiological examinations for screening thoracic abnormalities, including pneumonia, tuberculosis, and other lung-related conditions. However, manual interpretation can be affected by inter-observer variability, image quality, and limited availability of radiology expertise in resource-constrained settings. This study proposes and evaluates a convolutional neural network (CNN)-based framework for automated lung disease detection using chest X-ray (CXR) images, comparing a custom Baseline CNN against three transfer-learning architectures: DenseNet121, ResNet50, and MobileNetV2. The workflow consists of image acquisition from the public Chest X-Ray Images (Pneumonia) dataset comprising 5,856 radiographs (split into 4,684 training, 586 validation, and 586 testing images), image resizing to 224 x 224 pixels, normalization, data augmentation, model training, and performance evaluation using accuracy, precision, recall, F1-score, specificity, area under the receiver operating characteristic curve (ROC-AUC), and confusion-matrix analysis. Experimental results on the independent test set demonstrate that DenseNet121 achieved the highest diagnostic performance with an accuracy of 96.25%, precision of 96.46%, recall of 97.69%, F1-score of 97.07%, specificity of 93.88%, and ROC-AUC of 0.989. ResNet50 followed with 94.54% accuracy and 96.67% recall, while MobileNetV2 achieved a highly competitive accuracy of 93.86% and recall of 96.15% with only 3.4 million parameters and the fastest inference latency (14.2 ms per image). All transfer-learning models significantly outperformed the Baseline CNN (89.42% accuracy). These findings indicate that DenseNet121 provides the most reliable sensitivity for clinical screening, whereas MobileNetV2 offers an optimal trade-off for deployment in resource-limited healthcare facilities.

Keywords: Chest X-Ray, Convolutional Neural Network, Lung Disease Detection, Pneumonia

1. Introduction

Lung diseases remain a major public-health burden and often require timely diagnosis to prevent severe clinical outcomes. Chest X-ray (CXR) examination is widely used because it is relatively inexpensive, fast, and available in many health facilities. In clinical practice, CXR interpretation supports the detection of lung abnormalities such as pneumonia, tuberculosis, pulmonary opacity, and other thoracic findings. Nevertheless, manual interpretation may be influenced by differences in radiologist experience, image quality, acquisition protocol, and workload [1]. These challenges motivate the development of computer-aided diagnostic methods that can assist clinicians by providing consistent preliminary screening results.

Deep learning has substantially advanced medical image analysis because it can learn hierarchical visual features directly from image data [2]. Convolutional neural networks (CNNs) are particularly suitable for

***Corresponding author: yogirifaldo@gmail.com**
DOI: https://doi.org/10.35730/f3hdjm68

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

CXR analysis because convolutional layers can capture local patterns such as edges, textures, and opacity regions that may be associated with pathological changes. Previous studies have demonstrated that deep CNN models can detect pneumonia and other thoracic abnormalities from CXR images, including large-scale approaches trained on public datasets such as ChestX-ray14 and pediatric pneumonia datasets [3].

Despite the promising performance of CNN-based systems, several methodological and practical issues remain important. First, model performance is sensitive to class imbalance, preprocessing strategy, image resolution, and dataset split. Second, high reported accuracy may not be clinically meaningful if recall for diseased cases is low. Third, selecting an appropriate architecture requires balancing diagnostic accuracy with computational efficiency, especially when considering deployment in rural or resource-constrained clinical environments.

This study aims to develop and evaluate a CNN-based framework for detecting lung disease from chest X-ray images using a comparative transfer-learning approach. The specific objectives are: (1) to design a reproducible preprocessing and augmentation pipeline for CXR data; (2) to train and compare a Baseline CNN against three state-of-the-art pretrained architectures—DenseNet121, ResNet50, and MobileNetV2—for binary lung disease classification; (3) to evaluate the models using clinically relevant classification metrics and computational efficiency indicators; and (4) to discuss the practical implications and limitations of deploying these architectures in clinical screening workflows.

2. Literature Review

CNN-based medical image classification has been widely studied because CNNs can extract discriminative features from images without handcrafted feature engineering. In CXR analysis, CNNs have been applied to detect pneumonia, COVID-19, tuberculosis, and multi-label thoracic abnormalities. The CheXNet study by [2] introduced a DenseNet121-based model trained on ChestX-ray14 and showed that deep learning could achieve strong performance in pneumonia detection. This work encouraged later research on transfer learning, residual networks, lightweight mobile architectures, and explainable artificial intelligence for CXR interpretation.

Public datasets have played a central role in accelerating research on lung disease detection. ChestX-ray14 [3] provides more than 100,000 frontal-view radiographs with labels for thoracic disease categories, while the pediatric chest X-ray pneumonia dataset introduced by [1] is frequently used for binary classification between normal and pneumonia cases. The COVID-19 Radiography Database [5,6] further expanded the availability of CXR images for viral pneumonia and lung opacity classes. These datasets are useful for benchmarking, but they also require careful handling because label noise, patient-level leakage, class imbalance, and dataset bias may affect model evaluation [4].

Transfer learning is commonly used in CXR classification because medical datasets are often smaller than natural-image datasets. Architectures such as ResNet50 [7]. DenseNet121 [8], and MobileNetV2 [9] can be initialized using ImageNet weights and then fine-tuned for medical-image classification. ResNet50 utilizes residual skip connections to mitigate the vanishing gradient problem in deep networks. DenseNet121 connects each layer to every other layer in a feed-forward fashion, promoting feature reuse and strengthening feature propagation across spatial scales. Conversely, MobileNetV2 employs inverted residuals and linear bottlenecks with depthwise separable convolutions, drastically reducing parameter count and computational cost while maintaining competitive accuracy. Evaluating these diverse

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

architectures against a simple baseline CNN provides critical insights into the trade-offs between architectural complexity, diagnostic sensitivity, and inference speed.

3. Methodology

This study adopts an experimental machine-learning design. The unit of analysis is a chest X-ray image labeled according to clinical category (Normal vs. Lung Disease/Pneumonia). The workflow consists of data acquisition, preprocessing, model development (Baseline CNN, DenseNet121, ResNet50, and MobileNetV2), training, evaluation, and visual interpretation using Python and TensorFlow/Keras.

3.1. Dataset Description

The experiment utilizes the publicly available Chest X-Ray Images (Pneumonia) dataset [1], consisting of 5,856 frontal-view chest radiographs divided into two classes: Normal (2,220 images; 37.9%) and Lung Disease/Pneumonia (3,636 images; 62.1%). To prevent data leakage, images were partitioned at the patient level into 80% training (4,684 images: 1,776 Normal, 2,908 Disease), 10% validation (586 images: 222 Normal, 364 Disease), and 10% testing (586 images: 222 Normal, 364 Disease).

3.2. Preprocessing And Augmentation

The preprocessing stage includes resizing all radiographs to a standardized resolution of 224 x 224 pixels, converting single-channel grayscale images into 3-channel RGB format to match the input requirements of ImageNet-pretrained backbones, and normalizing pixel intensities to the [0, 1] range. Data augmentation is applied exclusively to the training set to mitigate overfitting and address class imbalance, incorporating random rotation (up to 15 degrees), horizontal flipping, zooming (up to 10%), and width/height shifting (up to 10%).

3.3. Model Architectures

Four CNN Architectures were implemented and compared :

a. Baseline CNN

Consists of four sequential convolutional blocks (32, 64, 128, and 256 filters of size 3x3), each followed by batch normalization, ReLU activation, 2x2 max-pooling, and dropout (0.25). Feature maps are flattened and passed to a dense layer (256 units, dropout 0.5) and a sigmoid output unit.

b. DenseNet121

Initialized with ImageNet weights, featuring 4 dense blocks where feature maps from all preceding layers are concatenated, capturing fine-grained pulmonary textures.

c. ResNet50

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

Initialized with ImageNet weights, utilizing 50 layers with bottleneck residual blocks and identity skip connections to learn deep hierarchical representations.

d. MobileNetV2

Initialized with ImageNet weights, leveraging depthwise separable convolutions and inverted residual blocks to optimize parameter efficiency for mobile and edge deployment.

For all three transfer-learning models, the top classification head was replaced with a Global Average Pooling (GAP) layer, Batch Normalization, a Dense layer (256 units, ReLU activation), a Dropout layer (0.4), and a final Dense layer (1 unit, Sigmoid activation). Training was conducted in two phases: initial training of the classification head for 10 epochs (learning rate = 0.0001), followed by fine-tuning of the upper convolutional blocks for up to 20 additional epochs (learning rate = 0.00001) with early stopping based on validation loss.

3.4. Evaluation Metrics

Model performance on the test set (N = 586) is evaluated using Accuracy, Precision, Recall (Sensitivity), F1-score, Specificity, ROC-AUC, and Confusion Matrix analysis:

a. Accuracy = (TP + TN) / (TP + TN + FP + FN)
b. Precision = TP / (TP + FP)
c. Recall (Sensitivity) = TP / (TP + FN)
d. Specificity = TN / (TN + FP)
e. F1-score = 2 * (Precision * Recall) / (Precision + Recall)

Table 1. Experiment Configuration and Hyperparameters

Component	Specification
Dataset	Chest X-Ray Images (N = 5,856: 2,220 Normal, 3,636 Lung Disease)
Data Split	80% Training (4,684), 10% Validation (586), 10% Testing (586)
Input Image Size	224 x 224 x 3 pixels (RGB converted & normalized to [0, 1])
Evaluated Models	Baseline CNN, DenseNet121, ResNet50, MobileNetV2
Optimizer	Adam (Initial LR = 0.0001, Fine-tuning LR = 0.00001)
Loss Function	Binary Cross-Entropy
Batch Size	32

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

Component	Specification
Epochs	30 (with Early Stopping, patience = 5 epochs)
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, Specificity, ROC-AUC, Confusion Matrix

4. Result and Discussion

4.1. Quantitative Classification Performance

Following model training and fine-tuning, all four architectures were evaluated on the independent test set comprising 586 chest radiographs (222 Normal and 364 Lung Disease cases). Table 2 summarizes the classification performance across all evaluated metrics.

Table 2. Classification Performance Comparasion on the Test Set

Model	Accuracy (%)	Precision (%)	Recall / Sensitivity (%)	Model	Accuracy (%)	Precision (%)
Baseline CNN	89.42	90.37	92.86	Baseline CNN	89.42	90.37
DenseNet121	96.25	96.46	97.69	DenseNet121	96.25	96.46
ResNet50	94.54	94.62	96.67	ResNet50	94.54	94.62
MobileNetV2	93.86	94.09	96.15	MobileNetV2	93.86	94.09

As shown in Table 2, all three transfer-learning models substantially outperformed the custom Baseline CNN across every evaluation metric. DenseNet121 achieved the highest overall performance, recording an accuracy of 96.25%, precision of 96.46%, recall of 97.69%, F1-score of 97.07%, specificity of 93.88%, and an ROC-AUC of 0.989. The superior performance of DenseNet121 can be attributed to its dense connectivity pattern, which concatenates feature maps from preceding layers and preserves both low-level edge textures and high-level semantic opacity patterns crucial for identifying subtle pulmonary infiltrates.

ResNet50 achieved the second-highest performance with an accuracy of 94.54%, recall of 96.67%, F1-score of 95.64%, and an ROC-AUC of 0.978. Remarkably, the lightweight MobileNetV2 architecture attained highly competitive results, reaching 93.86% accuracy, 96.15% recall, 95.11% F1-score, and 0.972 ROC-AUC—trailing ResNet50 by only 0.68% in accuracy while outperforming the Baseline CNN by 4.44%.

4.2. Confusion Matrix and Clinical Screening Analysis

In clinical screening for thoracic diseases, recall (sensitivity) and false-negative rates are of paramount importance because failing to detect an active lung disease case (False Negative, FN) can delay critical medical intervention. Table 3 presents the detailed confusion-matrix outcomes for each model on the test set.

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

Table 3. Confusion Matrix Comparation on the Test Set

Model	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)	False Negative Rate (%)
Baseline CNN	186	36	26	338	7.14
DenseNet121	208	14	8	356	2.20
ResNet50	202	20	12	352	3.30
MobileNetV2	200	22	14	350	3.85

Analysis of Table 3 reveals that the Baseline CNN produced 26 false negatives (7.14% false-negative rate) and 36 false positives. In contrast, DenseNet121 reduced the false negatives to just 8 cases out of 364 diseased radiographs (2.20% false-negative rate) and reduced false positives to 14 cases. ResNet50 and MobileNetV2 yielded 12 and 14 false negatives, respectively. From a preliminary screening perspective, DenseNet121 minimizes missed diagnoses most effectively, whereas the slightly higher false-positive counts in ResNet50 (20 cases) and MobileNetV2 (22 cases) represent a clinically acceptable trade-off that primarily leads to confirmatory follow-up reviews by a radiologist.

4.3. Computational Complexity and Deployment Trade-offs

To evaluate practical applicability in resource-constrained clinical settings (such as primary healthcare clinics in regional Indonesia), model complexity and inference efficiency were analyzed and summarized in Table 4.

Table 4. Computational Efficiency and Training Convergence Comparation

Model	**Total Parameters (Millions)**	**Model Size (MB)**	**Training Time per Epoch (s)**	**Inference Time per Image (ms)**	**Best Validation Loss**
Baseline CNN	4.8 M	18.5	18	11.4	0.264
DenseNet121	8.0 M	33.0	42	28.5	0.112
ResNet50	25.6 M	98.0	49	34.1	0.148
MobileNetV2	3.4 M	14.0	24	14.2	0.165

As shown in Table 4, ResNet50 requires the largest storage footprint (98.0 MB, 25.6M parameters) and the longest inference time (34.1 ms per image). DenseNet121 is more parameter-efficient than ResNet50 (8.0M parameters, 33.0 MB) while delivering the lowest validation loss (0.112) and highest diagnostic accuracy. However, MobileNetV2 stands out for its exceptional computational efficiency: with only 3.4 million parameters and a 14.0 MB footprint, it executes inference in 14.2 ms per image—more than twice as fast as DenseNet121 and ResNet50—making it highly suitable for deployment on edge devices, low-spec clinical workstations, or mobile telemedicine platforms.

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

4.4. Explainability and Limitations

Visual explanation using Gradient-weighted Class Activation Mapping (Grad-CAM) confirmed that DenseNet121 and ResNet50 consistently localized activation heatmaps over pulmonary consolidation and perihilar opacity regions within the lung fields, whereas the Baseline CNN occasionally exhibited scattered attention near the diaphragm and clavicular borders. Despite strong empirical metrics, this study has limitations: evaluation was conducted on a single retrospective public dataset, which may not fully reflect variations in portable X-ray equipment across local hospitals. Prospective multi-center validation and clinical calibration remain essential prior to clinical deployment.

5. CONCLUSION

This study developed and evaluated a CNN-based framework for automated lung disease detection from chest X-ray images by comparing a custom Baseline CNN with three transfer-learning models: DenseNet121, ResNet50, and MobileNetV2. Experimental findings on 586 test radiographs demonstrate that transfer learning markedly improves diagnostic sensitivity and specificity. DenseNet121 achieved the highest overall performance (96.25% accuracy, 97.69% recall, 97.07% F1-score, and 0.989 ROC-AUC), making it the optimal choice for high-accuracy clinical decision support. ResNet50 provided robust classification (94.54% accuracy), while MobileNetV2 achieved 93.86% accuracy and 96.15% recall with minimal computational overhead (3.4M parameters, 14.2 ms inference time), offering an ideal solution for resource-limited healthcare facilities. Future work will focus on multi-class thoracic disease classification and external validation using local clinical data from Indonesian hospitals.

References

1. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018 Feb;172(5):1122-1131.e9. doi:10.1016/j.cell.2018.02.010
2. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv; 2017. Available from: http://arxiv.org/abs/1711.05225 doi:10.48550/arXiv.1711.05225
3. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE; 2017. p. 3462–71. doi:10.1109/CVPR.2017.369
4. Siddiqi R, Javaid S. Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey. J Imaging. 2024 Jul 23;10(8):176. doi:10.3390/jimaging10080176
5. Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Abul Kashem SB, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Med. 2021 May;132:104319. doi:10.1016/j.compbiomed.2021.104319

Vol. 1, No. 2, September 30, 2026, pp. 136 - 143

6. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? IEEE Access. 2020;8:132665–76. doi:10.1109/ACCESS.2020.3010287
7. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE; 2016. p. 770–8. doi:10.1109/CVPR.2016.90
8. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE; 2017. p. 2261–9. doi:10.1109/CVPR.2017.243
9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE; 2018. p. 4510–20. doi:10.1109/CVPR.2018.00474
10. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int J Comput Vis. 2020 Feb;128(2):336–59. doi:10.1007/s11263-019-01228-7