IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141260

PULMONARY DISEASE PREDICTION USING MACHINE LEARNING

Mrs. Nethravathi K.G[1], Kavya S[2], Gagana Shree S[3], Keerthana B[4], Ganashree C. N[5]

Department of Computer Science & Engineering, K S School of Engineering and Management, Bangalore, India[1-5]

Abstract: Pulmonary disorders such as pneumonia remain a major clinical challenge, increasing the need for rapid and dependable diagnostic methods. This study proposes a machine learning–based system that examines both patient-reported symptoms and chest X-ray scans to estimate the probability of pneumonia. An initial risk score is generated using a weighted survey analysis, after which the chest radio-graphs are processed using MobileNetV2 for feature extraction and fed into a Convolutional Neural Network (CNN) for classification. By combining symptom evaluation with automated image interpretation, the system improves diagnostic accuracy and reduces reliance on manual assessment. This integrated approach supports faster screening and enhances the overall efficiency of pulmonary disease detection.

Keywords: Machine Learning, Deep Learning, CNN, MobileNetV2, Chest X-ray.

I. INTRODUCTION

Respiratory illnesses represent a major global health concern, contributing to hospital admissions, long-term complications, and high mortality rates. Diseases such as pneumonia, asthma, tuberculosis, COPD, and lung cancer require timely identification to prevent severe outcomes. Diagnosis is often delayed due to overlapping symptoms, limited specialist availability, and the complexity involved in interpreting radio-graphic images.

Recent advancements in Machine Learning (ML) and Deep Learning (DL) have enabled automated analysis of medical images with high precision. These techniques process chest X-rays, identify abnormalities, and detect early signs of lung disease that may be difficult for clinicians to observe manually. CNN-based architectures and lightweight models like MobileNetV2 have proven particularly effective due to their ability to learn hierarchical visual features while maintaining low computational cost, making them appropriate for real-time clinical use.

II. METHODOLOGY

The proposed system integrates both symptom-based evaluation and deep learning–based X-ray classification. The workflow includes four major stages: preprocessing, feature extraction, model development, and evaluation.

A. *Data Preprocessing*

All X-ray images are resized and normalized to ensure consistency. Augmentation techniques (rotation, flipping, contrast enhancement) are applied to improve robustness. The dataset is divided into training, validation, and testing sets in a 60:20:20 ratio.

B. Feature Extraction

MobileNetV2 is employed as the feature extractor because of its efficient architecture and pretrained weights, which help produce meaningful image representations.

C. *Model Development*

Two parallel approaches were utilized:

1. **Weighted Rule-Based System** – A scoring method assigns values to user-reported symptoms to classify them into initial risk categories.
2. **Convolutional Neural Network (CNN)** – Multiple convolution, pooling, and dense layers identify spatial features in X-ray images and classify them as Normal or Pneumonia.

D. Performance Evaluation

Standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix were used to assess model performance. These indicators help analyze classification strength and identify misclassification patterns.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141260

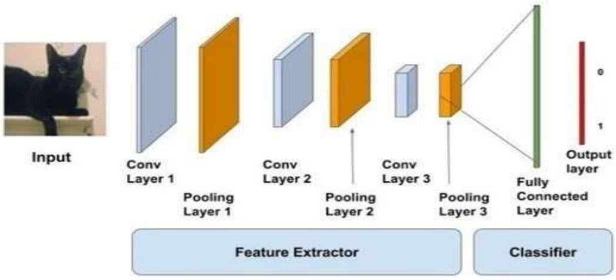

Fig. 1. CNN overview

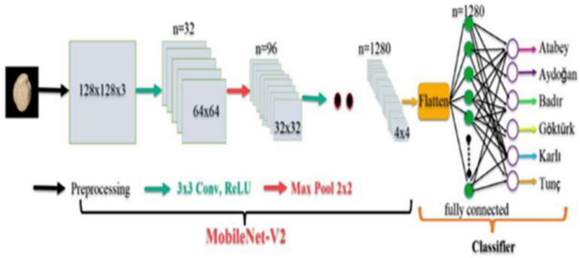

Fig. 2. MobileNet-V2 architecture

III. DATASET AND PARAMETER SETTINGS

Chest X-ray images of both pneumonia and healthy patients were collected primarily from Kaggle's open-source datasets. Additional real-world scans were obtained from local hospitals to improve the generalization ability of the model. Images with poor quality—such as blurred scans or those containing distracting artifacts—were removed to maintain dataset reliability.

Table 1. Chest X-ray images for Pneumonia

Category	No. Of Images
Pneumonia	5863
Normal	5852

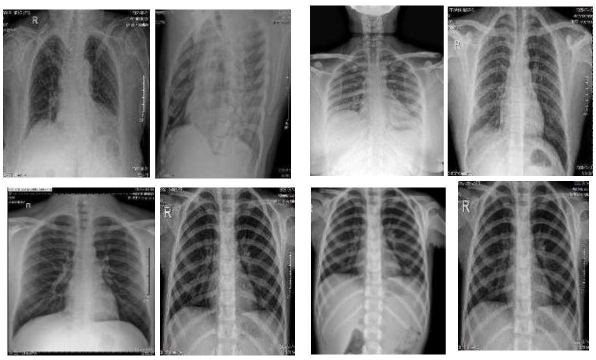

Fig.3 Normal and Pneumonia chest X-ray images

Cleaned and standardized images were finally used for MobileNetV2-based feature extraction and training.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141260

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Survey-Based Risk Assessment

The initial stage uses a symptom-based questionnaire, evaluating factors like age, respiratory history, fever, cough, and chest tightness. Based on a weighted scoring mechanism, users are categorized into Low, Medium, or High risk. This step helps prioritize users who truly require radiographic analysis.

B. Image Preprocessing

To ensure uniformity and improve feature clarity, each X-ray is processed through:

- Resizing to 224×224
- Normalization
- Grayscale conversion
- Canny edge detection to highlight lung boundaries and opacities

C. Machine Learning Prediction

The CNN model differentiates Normal and Pneumonia scans by observing features like opacities, cloudiness, and tissue consolidation.

Model performance:

- **Accuracy:** ~90%
- **High Sensitivity:** Effective detection of pneumonia
- **High Precision:** Fewer false positives

The confusion matrix confirmed strong classification across both classes.

D. Combined Decision Strength

Integrating survey scoring with deep learning prediction produces more consistent results:

- Low-risk users mostly matched Normal X-ray outputs
- High-risk users frequently aligned with pneumonia-positive predictions
- False negatives reduced significantly

E. User Evaluation and System Behavior

The system delivered predictions within 1–2 seconds. Users reported that the interface was clean, intuitive, and easy to navigate. Automatically generated PDF summaries also enhanced clarity and usability, offering transparency by displaying both raw and processed X-ray images.

F. Limitations

- Limited dataset size restricts clinical-grade deployment
- Some hospital images still contain noise or artifacts
- Currently supports only pneumonia classification

Future improvements include expanding datasets, adding multi-disease detection, and integrating more patient clinical parameters.

G. Accuracy Analysis

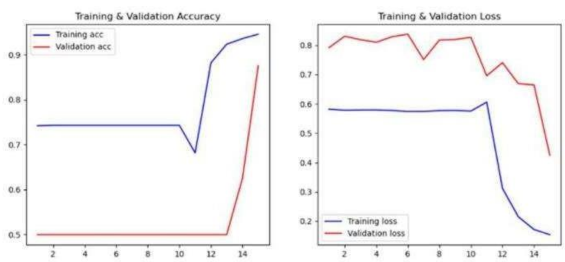

Fig.5 Training and Validation Accuracy Curve for the CNN

Training accuracy gradually increased and stabilized around 97%, while validation accuracy remained between 91–94%, indicating limited overfitting and good generalization.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141260

V. CONCLUSION AND FUTURE WORK

This work presents an AI-based screening method that combines rule-based symptom analysis with deep learning of chest X-ray images. The use of MobileNetV2 for feature extraction and CNN for classification enables fast and highly accurate detection of pneumonia. The system reduces manual dependency and is well-suited for remote healthcare, telemedicine, and resource-limited environments.

Future advancements may include:

- Secure patient-record storage and retrieval
- Multi-disease classification (TB, COPD, lung cancer, etc.)
- Larger and more diverse datasets
- Real-time deployment for hospitals and mobile applications

REFERENCES

[1]. A. P. Singh, Ankita, and D. C. Savinod, “An Analysis of Deep Learning Algorithms for Detection of Pulmonary Illness,” in Proc. 2nd Int. Conf. Comput. Intell., Commun. Technol. Netw. (CICTN), 2025, pp. 20–25, doi: 10.1109/CICTN64563.2025.10932572.

[2]. V. Tidke, S. Zade, H. Bhimarapu, R. Agrawal, N. C. Morris, and C. Dhule, “Multi Disease Prediction Related to Pulmonary Area by Leveraging Deep Learning,” in Proc. 2025 Int. Conf. on Computational, Communication and Information Technology (ICCCIT), 2025, pp. 58–61, doi: 10.1109/ICCCIT.2025.10927913.

[3]. T. Gupta and S. P. Panda, "A Comparison of Machine Learning Algorithms in Predicting Pulmonary Diseases," 2022 International Conference on Machine Learning , Big Data , Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, May 2022, pp. 791–795, doi: 10.1109/COM-IT CON54601.2022.9850510.

[4]. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, et al. (2017) "CheXNet: RadiologistLevel Pneumonia Detection on Chest X-Rays with Deep Learning." Retrieved from the arXiv preprint server (arXiv:171105225)

[5]. Fu Y, et al. (2023). "A Deep Learning Approach to Pneumonia Detection in X-ray Images." IEEE Transactions Medical Imaging. Volume, Pages.

[6]. Huynh, B. Q., et al (2021). "Transfer Learning from Deep Convolutional Neural Networks for Digital Mammographic Tumor Classification." Journal of Medical Imaging. Volume,Pages

[7]. Shin, H. C., et al. (2022). "Deep Convolutional Neural Networks for Computer-Aided Detection: Architectures, Dataset Characteristics, and Transfer Learning." IEEE Transactions on Medical Imaging. Volume. Pages.

[8]. Johnson, L., et al. (2021). "Data Collection Techniques for Supervised Learning in Medical Imaging." Journal of Biomedical Informatics, Volume, Pages

[9]. Wang, L, et al (2022). "Transfer Learning Strategies for Medical Image Analysis: A Comprehensive Review." IEEE Journal of Biomedical and Health Informatics, Volume, Pages

[10]. Smith, J., et al. (2022). "Advancements in Pulmonary Disease Detection Systems Using Deep Learning." Journal of Medical Imaging. Volume, Pages.