Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 2 (2026)

Lung Disease Prediction using Deep Learning Techniques

Deepalakshmi B[1], Dr. K. Sumangala[2]

[1]PhD Scholar, Department of Computer Science, Kongunadu Arts and Science College, Coimbatore – 641029.

deepalakshmi941@gmail.com

[2]Associate Professor & Head, Department of Computer Applications, Kongunadu Arts and Science College, Coimbatore – 641029.

sumangalak_ca@kongunaducollege.ac.in

Abstract:

The timely and correct diagnosis of lung ailments has been an important challenge in the medical domain because of the similarity in symptoms and the inability to analyze chest X-ray images manually. However, the concept of automated image analysis in the medical domain by utilizing deep learning techniques has opened new doors for improving diagnosis accuracy and efficiency. In this paper, a new automated system has been introduced utilizing the concept of Convolutional Neural Networks for multi-class classification of lungs into bacterial pneumonia, Viral pneumonia, Tuberculosis, COVID-19, and Normal lungs. Advanced models of neural networks such as DenseNet201, ResNet50, VGG16, and Inception V3 are used in this paper for feature extraction from chest X-ray images for providing accurate classification. The models of AI and machine learning in this paper prove to have remarkable ability for distinguishing between different types of lung ailments without relying heavily on manual analysis by experts in the domain of medicine.

Keywords: Chest X-ray, Lung Disease Classification, Deep Learning, Convolutional Neural Networks, Medical Image Analysis.

I. Introduction

Lung conditions remain a major source of concern for global healthcare. Bacterial pneumonias, viral pneumonias, tuberculosis, and COVID-19 are some conditions that show overlapping symptoms. Chest X-rays are commonly used as a primary tool for diagnosing individuals with lung conditions. Chest X-rays are preferred due to their low costs and ease of accessibility. However, the manual analysis of these images is a time-consuming process, has inter-observer variations, and requires the presence of skilled professionals. These factors are more commonly a concern for healthcare organizations that have limitations. Timely diagnosis remains a challenge for such organizations.

Recent breakthroughs in deep learning technology have led to successful implementations in the field of medical image analysis and provided an easier and reliable method for the detection of diseases in an automated manner. Convolutional Neural Networks (CNN) have demonstrated state-of-the-art performance in feature extraction and classification tasks using images. The aim of this paper is to provide an automation method using deep learning technology for detection and classification of lung diseases such as bacterial pneumonia, viral pneumonia, tuberculosis, and coronavirus, and healthy lungs using chest X-rays. The system relies on the latest available CNN designs like DenseNet201, ResNet50, VGG16, and InceptionV3 for improving the accuracy level of the process. By implementing the automation process for the analysis, the purpose of this method is to ensure reduced workload for the radiologists, accuracy, and effective decision-making in the real world.

II. Literature Survey

1. **Shatnawi et al. (2025)** introduced an automated framework based on deep learning for classifying lung cancer through CT scan images. Their research utilized advanced Convolutional Neural Networks with pre-trained models to categorize lung cancer into adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cases. The findings indicated that fine-tuning CNN models significantly enhances classification accuracy and reliability, underscoring the promise of deep learning in facilitating early lung cancer diagnosis and aiding clinical decision-making.

2. **Kansal et al. (2024)** performed a comparative study of ResNet-50 and EfficientNet-B0 for identifying lung abnormalities using chest X-ray images sourced from various locations. The research highlighted the capability of deep learning models to minimize diagnostic variability and assist in the automated detection of lung diseases with reduced reliance on expert radiologists.

3. **Kanakaraddi et al. (2024)** investigated deep learning methodologies for diagnosing lung cancer through histopathological images. Their strategy employed the U-Net architecture for image segmentation alongside deep CNN models like ResNet-50, VGG-16, and EfficientNet-B5 for classification.

4. **Ifty et al. (2024)** explored explainable deep learning methods for classifying lung diseases through chest X-ray images. The research focused on various lung conditions, such as viral pneumonia, bacterial pneumonia, COVID-19, tuberculosis, and healthy lungs. By combining Explainable AI (XAI) techniques with Convolutional Neural Networks (CNNs), hybrid models, and transformer-based architectures, the study improved model transparency and interpretability, thus fostering trust and usability in clinical settings.

5. **Sheela et al. (2024)** introduced a machine learning-based framework for predicting lung diseases that utilizes Convolutional Neural Networks. Their research included sophisticated image preprocessing methods to enhance image quality and feature extraction. The findings revealed increased classification accuracy and highlighted the effectiveness of CNN-based techniques for early disease detection and supporting clinical decisions.

6. **Ashwini et al. (2023)** created an optimized deep convolutional neural network aimed at multi-class lung disease classification using chest X-ray images. The use of grid search for hyperparameter optimization and thorough preprocessing significantly boosted the model's performance, demonstrating the impact of deep learning on enhancing diagnostic accuracy and aiding healthcare professionals.

7. **Rajasekar et al. (2023)** examined deep learning models for predicting lung cancer through both CT scans and histopathological images. The results indicated that histopathological images provided greater diagnostic accuracy, highlighting their critical role in the early detection of lung cancer.

III. Methodology

3.1 Dataset Description

The dataset comprises chest X-ray (CXR) images depicting five lung conditions: bacterial pneumonia, viral pneumonia, tuberculosis, COVID-19, and healthy lungs. These images are obtained from publicly accessible repositories. Initially, the dataset contains approximately 6,381 images, which are enhanced to around 10,000 through techniques such as rotation, horizontal flipping, and random cropping. An 80:20 ratio is utilized for the division of training and testing sets. The images are saved in JPEG and PNG formats, ensuring adequate resolution (224×224 pixels) for deep learning applications, while personal patient information is excluded to uphold privacy.

Figure 1: Example of chest X-ray (CXR) Images

3.2 Data Preprocessing

Data preprocessing is crucial for ensuring consistency, enhancing model generalization, and minimizing training complexity. The following procedures are implemented:

Image Resizing: All images are resized to 224×224 pixels to align with the input specifications of CNN models.

Normalization: Pixel intensity values are normalized to diminish variation and stabilize the training process.

Data Augmentation: Techniques such as rotation, flipping, and scaling are employed to broaden the dataset and mitigate overfitting.

Data Splitting: Images are partitioned into training and testing sets (80%–20%) using stratified sampling to preserve class distribution.

One-Hot Encoding: Class labels are transformed into one-hot encoded vectors for multi-class classification.

3.3 Model Architecture

3.3.1 Custom Convolutional Neural Network (CNN)

A tailored CNN is developed specifically for the classification of lung diseases. The architecture features multiple convolutional layers (Conv2D) with diverse filter sizes to capture local patterns, max-pooling layers to decrease dimensionality, batch normalization for stable training, dropout for regularization, and fully connected dense layers with a softmax output for multi-class classification.

3.3.2 Transfer Learning Models

Pre-trained models are utilized to take advantage of existing feature representations:

DenseNet201: The dense connectivity improves gradient flow and allows for feature reuse.

ResNet50: The use of residual blocks with skip connections helps to avoid vanishing gradients and enhances the training process of deep networks.

VGG16: This sequential architecture employs small convolutional filters to effectively capture hierarchical features.

InceptionV3: Inception modules are designed to efficiently extract features at multiple scales.

EfficientNetB0: This model achieves high accuracy with reduced computational cost by balancing the scaling of depth, width, and resolution.

3.3.3 Hybrid Model

A hybrid architecture integrates a custom CNN with ResNet50, enabling the capture of both disease-specific local features and strong general feature representations. The feature vectors generated by both models are concatenated and then processed through fully connected layers with softmax activation for multi-class predictions.

3.4 Model Training and Optimization

Loss Function: Categorical cross-entropy is used for multi-class classification.

Optimizer: Adam optimizer is used with adaptive learning rates.

Regularization: Dropout and early stopping prevent overfitting.

Hyperparameter Tuning: Learning rate scheduling and selective unfreezing of layers for transfer learning improve convergence.

Epochs and Batch Size: Models are trained over multiple epochs with a batch size optimized for computational efficiency.

3.5 Prediction Process

The trained models predict lung disease categories on unseen test images. The workflow involves:

1. Loading and preprocessing images to 224×224×3 pixels.
2. Forward propagation through the trained model to generate probability scores.
3. Selecting the disease class with the highest probability.
4. Validation against true labels to evaluate accuracy and generalization.

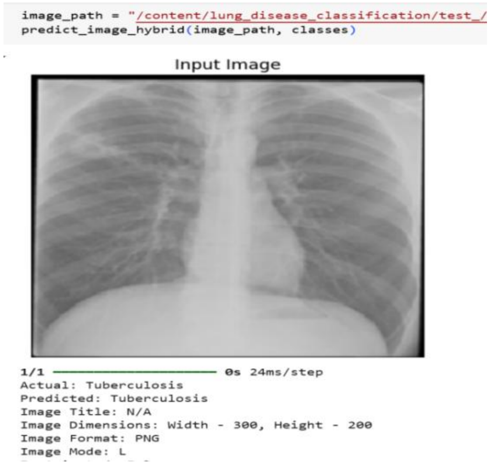

Figure 2: Chest X-ray Image with Prediction Result

3.6 Evaluation Metrics

Performance of the models is assessed using standard classification metrics:

Accuracy: Measures the proportion of correctly classified images among all images.

Precision: Evaluates the correctness of positive predictions, minimizing false positives.

Recall (Sensitivity): Measures the ability to identify actual diseased cases, minimizing false negatives.

F1-Score: Provides a harmonic mean of precision and recall, ensuring balanced evaluation.

Confusion Matrix: Presents a detailed comparison between predicted and actual labels, highlighting correctly classified instances and patterns of misclassification.

IV Results and Discussion

4.1 Model Performance

Performance of all deep learning models for lung diseases classification is given in Table I. Of all models considered, the Hybrid Model (Custom CNN + ResNet50) and Custom CNN models have been found to have the highest accuracy of 98%. The remaining models such as DenseNet201, ResNet50, VGG16, InceptionV3, and EfficientNetB0 have been found to have different levels of performances, with the highest F1-score of 86% given by EfficientNetB0.

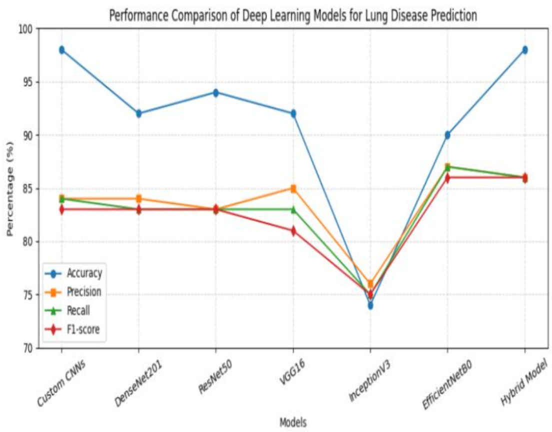

Figure 3: Performance Metrics of Deep Learning Models for Lung Disease Prediction and Classification.

The Hybrid Model outperformed other models due to its successful incorporation of the complementary characteristics of both the Custom CNN and ResNet50 models, resulting in improved accuracy and an appropriate precision recall curve. The custom CNN model had an outstanding pattern extraction skill for the disease without requiring pre-trained models.

Analysis of misclassification indicates that the highest rates of misclassification are in the area between bacterial pneumonia and viral pneumonia, which indicates where further refinement will be needed. From the above studies, it can be seen that deep learning models, in particular hybrid models, are capable of efficiently identifying various types of pneumonias from chest X-ray images. When efficiency reaches its peak, it can automatically ease intense work on the part of radiologists, fast-track decision-making processes in hospitals, and massive screenings in hospitals.

V. Challenges and Future Directions

Technical Challenges: It requires considerable computational power for designing and training the deep learning model for analyzing medical images. It is also challenging to combine various CNN models for handling different chest X-ray images for efficient feature extraction.

Availability of Data: Unavailability of large, well-balanced, and high-quality health-care information might impact the accuracy of the model. Heterogeneity, image resolution, and inconsistencies in the labels can be challenges for the model to form generalizations.

Scalability and Deployment: The deployment of deep learning models into real-world clinical environments needs to adopt scalable approaches. It is essential to ensure that the models are efficient across hospitals, irrespective of the underlying infrastructural setup.

Interpretability and Trust: "Black boxes" are how deep models are perceived, which might limit adoption in the clinical setting. There is a need to integrate explainable AI tools to enhance interpretability and gain trust among medical practitioners.

Ethics and Privacy: Dealing with sensitive patient information raises issues related to privacy and compliance.

Future Directions: Increasing the size of datasets, using multi-modal imaging such as CT and histopathology images, and developing hybrid or ensemble models could help in improving the diagnostic accuracy. More clinically relevant approaches should be stage-wise disease detection, attention mechanisms, and explainability in AI.

VI. Conclusion

Deep learning has emerged as a highly effective method for lung disease classification based on chest X-rays. This has been made evident by the various implementations of different architectures for the convolutional neural network. The Hybrid Model that combines both CNN models and ResNet50, both of which show an astonishing accuracy level of 98%. This accuracy has been reinforced by the analysis of the confusion matrix as well as the effective prediction function, both of which show effective classification. There is also more to be learned from other models like DenseNet201, VGG16, and InceptionV3, which add different viewpoints to the assessment process. There are also different approaches that can be made for future improvements related to increased effectiveness, including these models being implemented in actual clinical setups.

References

[1] M. Q. Shatnawi, Q. Abuein, and R. Al-Quraan, "Deep learning-based approach to diagnose lung cancer using CT-scan images," 2025.

[2] K. Kansal, T. B. Chandra, and A. Singh, "ResNet-50 vs. EfficientNet-B0: Multi-centric classification of various lung abnormalities using deep learning," 2024.

[3] S. G. Kanakaraddi, V. S. Handur, S. Giraddi, A.Jalannavar, and A. Chikaraddi, "Segmentation and classification of lung cancer using deep learning," 2024.

[4] T. T. Ifty, S. A. Shafin, S. M. Shahriar, and T.Towhid,"Explainable lung disease classification from chest X-ray images utilizing deep learning and XAI," 2024.

[5] M. S. Sheela, G. Amirthayogam, J. J. Hephzipah, S. Gopalakrishnan, and S. R. Chand, "Machine learning-based lung disease prediction using convolutional neural network algorithm," 2024.

[6] S. Ashwini, J. R. Arunkumar, R. T. Prabu, N. H. Singh, and N. P. Singh, "Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network," 2023.

[7] A. M. Q. Farhan and S. Yang, "Automatic lung disease classification from chest X-ray images using hybrid deep learning algorithm," 2023.

[8] V. Rajasekar, M. P. Vaishnnave, V. Premkumar, V. Sarveshwaran, and V. Rangaraaj, "Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques," 2023.

[9] A. Hamdi, A. Aboeleneen, and K. Shaban, "MARL: Multimodal attentional representation learning for disease prediction," 2021.