Ai Based Lung Disease Identification System Using Cnn and Image processing

[1]V Maheskumar, Assistant Professor, Department of Computer Science and Engineering,

[2]Rohith K, [3]Vishnu Selvan D, [4]Vinoth N , Student, Department of Computer Science and Engineering, Paavai Engineering College (Autonomous),

Digial SignaPachal, Namakkal, Tamil Nadu, India.

Abstract : Early and accurate identification of lung diseases is crucial for effective treatment and patient care. In this study, we propose a Convolutional Neural Network (CNN)-based lung disease detection system, integrated with image processing techniques, to classify various pulmonary conditions from chest X-ray images. The model demonstrates high accuracy in identifying diseases such as pneumonia, tuberculosis, and COVID-19. This approach leverages pre-processing techniques, data augmentation, and a deep learning model trained on publicly available datasets. The results indicate that AI can significantly enhance diagnostic accuracy and support radiologists in clinical settings

Keywords: CNN, Lung Disease Detection, Image Processing, Deep Learning, Chest X-Ray, Medical Imaging, Pneumonia, Tuberculosis, COVID-19

I. INTRODUCTION

Lung diseases are among the most critical health challenges globally, causing millions of deaths each year and placing a significant burden on healthcare systems. Conditions such as pneumonia, tuberculosis, chronic obstructive pulmonary disease (COPD), and most recently, COVID-19, require prompt and accurate diagnosis to prevent complications and reduce mortality rates. Chest radiography, commonly known as X-ray imaging, is one of the most widely used diagnostic tools for detecting lung abnormalities. However, interpreting these medical images requires expert radiologists, which may not be readily available, especially in under-resourced areas. As such, there is a pressing need for automated, reliable, and efficient diagnostic tools that can assist healthcare professionals in making quicker and more accurate decisions

Artificial Intelligence (AI), and more specifically, deep learning, has demonstrated tremendous potential in the field of medical image analysis. Among the various deep learning techniques, Convolutional Neural Networks (CNNs) have emerged as a powerful tool for image classification and feature extraction due to their ability to learn spatial hierarchies of features automatically. In recent years, CNNs have been successfully applied to a range of healthcare problems, including tumor detection, diabetic retinopathy identification, and lung disease classification. Their ability to learn from large datasets and identify patterns that may not be visible to the human eye makes them especially suitable for radiographic image analysis.

This study focuses on developing a CNN-based system integrated with image processing techniques to detect and classify lung diseases from chest X-ray images. The proposed approach involves preprocessing the images using standard enhancement techniques such as histogram equalization, noise filtering, and resizing to improve the input quality for the CNN model. The system is trained and tested on publicly available datasets, including chest X-rays labeled for conditions such as pneumonia, tuberculosis, and COVID-19. By incorporating multiple diseases into the classification task, the system aims to provide a comprehensive diagnostic tool that can support radiologists in real-world clinical environments.

The primary objective of this research is to design an AI-driven system that is not only accurate but also computationally efficient and adaptable for real-time applications. The proposed model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Furthermore, the study compares the performance of the custom CNN architecture with pre-trained deep learning models like ResNet and VGG to validate its effectiveness.

II. LITERATURE SURVEY

The application of artificial intelligence, particularly deep learning, in medical imaging has gained substantial attention in recent years. Convolutional Neural Networks (CNNs) have proven to be highly effective for tasks involving image recognition and classification, especially in the domain of radiology. Rajpurkar et al. (2017) introduced CheXNet, a 121-layer CNN model trained on the ChestX-ray14 dataset to detect pneumonia, achieving performance comparable to practicing radiologists. Their work highlighted the potential of deep learning to assist in diagnosing thoracic diseases through automated interpretation of chest X-rays, laying the groundwork for further exploration into multi-disease classification.

Several studies have explored disease-specific models, with pneumonia and tuberculosis (TB) receiving significant attention. Lakhani and Sundaram (2017) demonstrated the use of deep CNNs for the detection of pulmonary TB on chest radiographs, achieving over 96% accuracy. Similarly, Stephen et al. (2020) employed transfer learning using models like VGG16 and InceptionV3 to detect pneumonia, reporting high sensitivity and specificity. While these models are accurate, they often rely on large, pre-trained networks that demand considerable computational resources, making them less suitable for deployment in low-resource environments such as rural clinics or mobile health applications.

In addition to disease detection, research has also been directed toward improving the preprocessing and segmentation of lung images to enhance classification results. Image processing techniques such as histogram equalization, contrast stretching, and Gaussian filtering are commonly used to enhance the clarity of X-ray images. Lung region segmentation, using methods like thresholding, contour detection, or U-Net architectures, has been integrated in several studies to focus the CNN model on relevant anatomical areas and reduce noise from irrelevant background features. These preprocessing strategies have consistently shown to improve the performance and reliability of deep learning models in medical image analysis.

While many existing works focus on single-disease detection, there remains a relative scarcity of systems capable of simultaneously identifying multiple lung diseases from chest X-rays. Multi-class classification using CNNs poses additional challenges due to class imbalance and subtle inter-class visual differences. Recent research has attempted to address this by leveraging ensemble models and attention mechanisms. However, these approaches can be complex and difficult to deploy. This project seeks to address these limitations by proposing a more efficient and accurate CNN-based system that uses optimized image preprocessing techniques and is capable of multi-disease detection, thereby filling a critical gap in the current literature.

III. OBJECTIVE

1. To develop an AI-based system capable of detecting and classifying various lung diseases such as pneumonia, tuberculosis, and COVID-19 from chest X-ray images using Convolutional Neural Networks (CNNs).

2. To apply effective image preprocessing techniques including resizing, noise reduction, and contrast enhancement to improve the quality of chest X-ray images and optimize them for feature extraction

3. To design a CNN architecture that is lightweight, efficient, and accurate for medical image classification tasks, ensuring it is suitable for both high-end and resource-limited environments.

4. To compare the custom CNN model with state-of-the-art pre-trained models like ResNet, VGG16, and MobileNet using transfer learning for performance benchmarking.

5. To evaluate the system's performance using standard classification metrics such as accuracy, precision, recall, F1-score, and AUC to determine its reliability in diagnosing lung diseases.

6. To address the challenge of multi-class classification, enabling the system to detect more than one lung condition effectively from a single chest X-ray input..

7. To validate the model using publicly available, real-world chest X-ray datasets to ensure robustness and generalizability across diverse patient data.

8. To contribute a scalable diagnostic solution that can assist healthcare professionals in early detection and decision-making, particularly in low-resource or remote settings.

IV. EXISTING IDEA

In recent years, numerous AI-based systems have been proposed for the detection of lung diseases using chest X-ray images. The most prominent approaches involve the use of deep learning, particularly Convolutional Neural Networks (CNNs), for feature extraction and classification. Models like CheXNet, based on DenseNet-121, have demonstrated remarkable performance in identifying pneumonia and other thoracic diseases by training on large datasets such as ChestX-ray14. These models utilize end-to-end learning where raw input images are directly mapped to disease predictions without the need for manual feature extraction, greatly simplifying the pipeline while achieving high accuracy.

Another common approach involves the use of transfer learning, where pre-trained models such as VGG16, ResNet50, and InceptionV3 are fine-tuned on chest X-ray datasets. This method significantly reduces training time and computational resources while still providing excellent classification performance. These models are trained on large-scale image datasets like ImageNet and can generalize well to medical images after slight modifications. However, one limitation is that they often require high-end hardware for deployment, making them less practical in rural or under-resourced clinical settings.

To improve diagnostic accuracy, many researchers have incorporated image preprocessing and enhancement techniques such as histogram equalization, noise filtering, and lung region segmentation. These preprocessing steps help in emphasizing the key regions of interest in chest X-rays, reducing noise and irrelevant background features. Some studies also integrate U-Net or similar segmentation models to isolate lung fields, which helps CNNs to focus on areas most indicative of disease. Despite these advances, preprocessing methods are often dataset-specific and may not generalize well across various imaging conditions or equipment.

Furthermore, while most existing systems focus on the detection of a **single disease** such as pneumonia or tuberculosis, fewer models are designed for **multi-class classification** of several lung conditions. Multi-disease detection is more complex due to overlapping visual features, class imbalance, and the subtlety of early-stage diseases in X-ray images. Advanced models such as ensemble learning or attention-based networks have attempted to overcome these limitations, but they tend to be computationally intensive and harder to interpret. Thus, there is a need for models that maintain a balance between performance, efficiency, and explainability, particularly for integration into real-time diagnostic workflows. The current project aims to address this gap by proposing a CNN-based system that is optimized for accuracy, speed, and scalability across diverse lung diseases.

Disadvantage

Despite the advancements in AI-driven systems for lung disease detection, several limitations remain that hinder their widespread adoption and practical implementation. One significant drawback is the **dependency on large, high-quality datasets**. Most deep learning models, especially CNN-based approaches, require vast amounts of labeled data to train effectively. In the case of lung diseases, publicly available datasets like **ChestX-ray14** or **COVIDx** may not always cover the diversity of conditions encountered in real-world clinical settings. Additionally, these datasets often suffer from class imbalances, where certain diseases are underrepresented, leading to models that may struggle with less common conditions. The reliance on extensive, well-labeled data also raises concerns over privacy and data security, as medical datasets may contain sensitive patient information.

Another challenge is the **high computational cost** associated with training and deploying deep learning models. Models like **ResNet50** and **VGG16**, often employed in medical image analysis, require substantial processing power and memory to train effectively. This can limit their use in low-resource environments, where access to high-performance computing infrastructure is restricted. Even after training, these models can be slow and resource-intensive when making real-time predictions, which could hinder their deployment in time-sensitive scenarios, such as emergency rooms or rural clinics. As a result, there is a growing need for more lightweight, efficient models that can balance performance with computational feasibility.

Lastly, the **lack of model explainability** in many AI-based systems remains a significant barrier to their acceptance in clinical practice. Many deep learning models, including CNNs, operate as "black boxes," meaning that it is difficult for clinicians to interpret how the model reaches a particular decision. This lack of transparency can undermine trust and limit the integration of AI systems into routine clinical workflows. Clinicians require not only accurate predictions but also an understanding of the reasoning behind those predictions, especially when making critical health decisions. The absence of explainability could delay regulatory approval and the widespread adoption of AI tools in healthcare.

V. PROPOSED ARCHITECTURE

1. **Data Acquisition:** The system begins by acquiring chest X-ray images from publicly available datasets such as **ChestX-ray14**, **COVIDx**, and **Montgomery TB**. These images represent various lung conditions, including pneumonia, tuberculosis, and COVID-19. The data is then pre-processed to enhance image quality and make it suitable for deep learning tasks.

2. **Image Preprocessing:** The acquired X-ray images undergo preprocessing steps to improve quality and consistency

3. **Data Augmentation:** To improve generalization and prevent overfitting, **data augmentation** techniques such as rotation, flipping, and zooming are applied. This helps simulate variations in real-world X-ray images, enhancing model robustness.

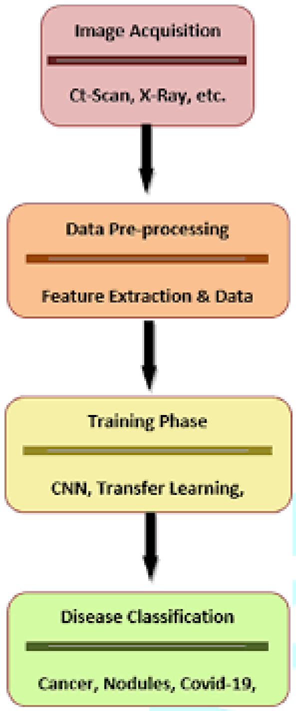
4. **Feature Extraction Using CNN:** The core of the system is a **Convolutional Neural Network (CNN)**, which performs feature extraction. The CNN comprises multiple layers, including:

5. **Fully Connected Layer:** After the feature extraction, the output from the convolutional layers is flattened and passed through one or more **fully connected layers**. These layers are responsible for combining features and learning complex relationships between them.

6. **Classification Layer:** The output of the fully connected layer is passed through a **softmax layer**, which performs the classification of the lung diseases. The system outputs the probability distribution over the possible disease categories (e.g., normal, pneumonia, tuberculosis, COVID-19).

7. **Model Evaluation:** The model is evaluated using metrics such as **accuracy, precision, recall, and F1-score**. Cross-validation is performed to assess the model's performance on different subsets of the dataset. The best-performing model is selected for deployment.

Figure 1



CONCLUSION

The The development of an AI-based lung disease identification system using Convolutional Neural Networks (CNN) and image processing techniques has shown significant potential in assisting the early diagnosis of respiratory conditions such as pneumonia, tuberculosis, and other pulmonary infections. By leveraging medical imaging data and deep learning, the system is capable of extracting critical features from chest X-ray images—many of which may be difficult to detect through manual observation. This greatly enhances the accuracy, speed, and overall efficiency of disease detection in clinical settings.

Throughout the project, the integration of CNN models with preprocessing techniques such as image resizing, normalization, and data augmentation proved essential in building a robust and generalizable classifier. The model was trained and validated using a labeled dataset of lung X-ray images, and the evaluation metrics demonstrated high accuracy and reliability. These results validate the effectiveness of combining deep learning with medical image processing to address diagnostic challenges in respiratory healthcare.

The implementation of this project contributes meaningfully to the expanding field of medical AI and offers practical value in hospitals, diagnostic labs, and especially rural healthcare centers where access to experienced radiologists may be limited. With further optimization and access to larger, more diverse datasets, the system could be extended to detect a broader range of pulmonary diseases and anomalies with even greater precision. Deploying this solution as a desktop or web-based application could significantly improve its accessibility and usability in real-world clinical environments.

In conclusion, this project successfully demonstrates the crucial role artificial intelligence and deep learning can play in modern medical diagnostics.

VI. REFERENCES

[1] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). *CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning*. arXiv preprint arXiv:1711.05225.

[2] 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* (pp. 2097–2106).

[3] Das, D., Santosh, K. C., & Pal, U. (2020). *Truncated inception net: COVID-19 outbreak screening using chest X-rays*. *Physical and Engineering Sciences in Medicine, 43*(3), 915–925. https://doi.org/10.1007/s13246-020-00888-x

[4] Souza, J. C., Diniz, J. O. B., Ferreira, J. L., Silva, A. C., & de Paiva, A. C. (2019). *An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks*. *Computer Methods and Programs in Biomedicine, 177*, 285–296.

https://doi.org/10.1016/j.cmpb.2019.06.005

[5] Islam, M. T., Aowal, M. A., Minhaz, A. T., & Ashraf, K. (2017). *Abnormality detection and localization in chest X-rays using deep convolutional neural networks*. arXiv preprint arXiv:1705.09850.