The Digital Stethoscope: A System for Multi-Modal Lung Health Analysis

Tushar Waykole
Computer Engineering Department
PCET's Nutan Maharashtra Institute of Engineering and Technology, Pune , India
tushar.waykole@nmiet.edu.in

Kunika Sirpor
Computer Engineering Department
PCET's Nutan Maharashtra Institute of Engineering and Technology, Pune , India
sirporkunika@gmail.com

Srushti Waghmare
Computer Engineering Department
PCET's Nutan Maharashtra Institute of Engineering and Technology, Pune , India
waghmaresushti832@gmail.com

Abscact— Lung-related illnesses, including pneumonia, tuberculosis, asthma, and chronic obstructive pulmonary disease (COPD), continue to affect millions of people worldwide [1]. Early and reliable detection is essential to prevent severe health outcomes, yet conventional diagnostic techniques like X-rays and auscultation rely heavily on human judgment. This dependency often leads to inconsistencies and delayed decisions [2]. The growing influence of Artificial Intelligence and machine learning in medical imaging and sound analysis has introduced quicker and more dependable diagnostic possibilities. This literature survey examines current development in AI-based detection of respiratory diseases [2][3]. One study explores a dual modality approach using Vision Transformers (ViT) and Convolutional Neural Networks (CNN) to combine X-ray images with lung sound data for improved diagnostic precision. Another Work Purpose, a CNN-Graph Convolutional Network (GCN) hybrid enhanced with an attention mechanism to classify non-speech respiratory sounds. Additionally, a comparative survey of ML algorithms highlights their application in diagnosing diseases such as COVID-19, pneumonia, and tuberculosis. Together, these discoveries suggest that multi-modal and hybrid deep learning models provide greater diagnostic accuracy and efficiency, paving the way for AI-assisted medical systems that can support medicals in both advanced and resource-limited environments

Keywords— *Artificial Intelligence, Machine Learning, Deep Learning, Vision Transformer, Graph Convolutional Network, Convolutional Neural Network, Respiratory Diagnostics, Multi-Modal Learning, Lung Sound Analysis*

I. INTRODUCTION

Respiratory diseases such as pneumonia, tuberculosis, chronic obstructive pulmonary disease (COPD), asthma, and lung cancer remain major global health concerns, affecting millions of people each year. According to international health reports, beyond one billion individuals suffer from acute or chronic lung conditions, with millions of deaths recorded annually [1]. The growing burden of these diseases places tremendous pressure on healthcare systems, particularly in regions with restricted medical facilities and diagnostic resources. Therefore, early and accurate diagnosis plays a important role in reducing mortality rates and improving treatment outcomes. Nonetheless, these techniques often depend on human proficiency, which may lead to subjective errors and delayed diagnosis [2]. To address these limitations, the integration of artificial intelligence and machine learning has gained significant attention in recent years. These technologies can examine huge amount of medical data efficiently, discover patterns that may not be easily visible to human observers.

Machine learning and deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in classifying medical images such as X-rays and CT scans. CNNs mechanically extract visual features from images, enabling faster and more accurate detection of diseases like pneumonia and tuberculosis. Similarly, advanced models such as Vision Transformers (ViTs) enhance diagnostic precision by utilizing self-attention mechanisms to identify detailed patterns in medical images.

Apart from visual analysis, respiratory sound data has also become an important diagnostic source. Lung sounds such as wheezes and crackles contain critical indicators of respiratory abnormalities. Deep learning approaches that convert these sounds into spectrograms have improved the precision of sound-based diagnosis. Current research has also explored hybrid models combining CNNs and Graph Convolutional Networks (GCNs) to capture both spatial and temporal patterns in lung sounds, achieving strong classification results [3].

Furthermore, multi-modal systems that integrate both X-ray images and lung sound data have proven to enhance diagnostic performance by offering a complete view of the patient's respiratory health. Such AI-driven frameworks are especially beneficial for rural and low-resource healthcare settings. The integration of image and audio analysis marks a notable advancement in automated medical diagnostics, providing reliable, fast, and scalable solutions for identifying respiratory diseases worldwide [1].

II. LITRATURE REVIEW

2.1 A Machine Learning Approach to Lung Disease Identification: Combining X-Ray and Data

Author: Prof. Dr. K. Chanthirasekaran, Department of Electronics and Communication Engineering, Saveetha Engineering College Published: 2025

The study concentrates on the growing role of Artificial Intelligence (AI) in automating lung disease diagnosis. The research is categorized into single-modality approaches, which use one type of data, and multi-modal approaches that combine different forms of data for improved accuracy.

For X-ray image analysis, deep learning techniques such as Convolutional Neural Networks (CNNs) have achieved remarkable success in detecting conditions like pneumonia and tuberculosis. Similarly, audio-based methods have been progressed to analyze lung sounds and detect unique patterns such as wheezes and crackles important indicators of diseases like asthma and bronchitis.

The study highlights that combining both image and sound data leads to more reliable outcomes. Previous work by Hatture, Koravanavar, and Saini (2022) supports this, showing that multimodal systems outperform single-modality ones.

Conclusion: The research concludes that integrating X-ray and audio data enhances diagnostic reliability, making multi-modal systems a promising approach for accurate respiratory disease detection The study focuses on the growing role of Artificial Intelligence in automating lung disease diagnosis. The research is categorized into single-modality approaches, which use one type of data, and multi-modal approaches that combine different forms of data for improved accuracy.

For X-ray image analysis, deep learning techniques such as Convolutional Neural Networks have achieved remarkable success in detecting conditions like pneumonia and tuberculosis. Similarly, audio-based methods have been progressed to analyze lung sounds and detect unique patterns such as wheezes and crackles important indicators of diseases like asthma and bronchitis. [1]

2.2 Enhancing Respiratory Disease Diagnosis with Graph Convolutional Networks Through Non-Speech Audio Synthesis

Authors: Raphael Anaadumba, Nazim A. Belabbaci, Mohammad Arif Ul Alam, Richard A. Miner School of Computer and Information Sciences, University of Massachusetts Lowell, USA

Published: 2025

This study investigates more developed deep learning architectures to better analyze respiratory sounds. Traditional CNN-based models were found to struggle in capturing relationships between features in audio data.

To address this drawback, the authors proposed a hybrid CNN, Graph Convolutional Network (GCN) model. The CNN extracts key audio features from spectrograms, while the GCN learns how these features relate to one another. An attention mechanism was added to help the model focuses on the most diagnostically important sound segments, leading to clearer differentiation between similar respiratory conditions.

Results: The hybrid CNN-GCN model achieved an accuracy of 93% in a six-class respiratory classification task.

Key Insight: The findings suggest that combining CNNs, GCNs, and attention layers can significantly improve diagnostic precision. This architecture can serve as a foundation for future multimodal systems that combine both audio and imaging data for better disease detection. [3]

2.3 A Survey on Lung Disease Diagnosis Using Machine Learning Techniques

Author: Mercy Rajaselvi V, Department of Computer Science and Engineering, Easwari Engineering College, Anna University, Chennai, India

Published: 2022

This paper presents various machine learning techniques used for identifying lung diseases. It emphasizes the dominance of CNNs and transfer learning in medical imaging. Pre-trained models such as ResNet50, DenseNet20l, VGG19, and InceptionV3 have been widely used to classify diseases including COVID-19, pneumonia, tuberculosis, and lung cancer.

Applications and Findings:

COVID-19: Models like COVID-SDNet have shown effective classification on chest X-rays.

Pneumonia: Deep Convolutional Neural Networks (DCNNs) have achieved high precision in detection.

Tuberculosis: CNN-based systems have proven useful for automated screening and early diagnosis.

Lung Cancer: 3D-CNNs have been used for correct nodule identification in CT scans.

Most reviewed studies reported accuracies between 80—85%, with some achieving up to 96.47% for tuberculosis and 92.8% for pneumonia. These results exhibit that transfer learning significantly enhances diagnostic performance.

Challenges: The study notes that progress is limited by a shortage of large, high-quality datasets and the heavy computational power required to train deep learning models. It propose that future work should focus on expanding open-access datasets and adopting cloud computing for scalable diagnosis. [2]

Model Type	Description
Image-based CNN Models	Analyze chest X-ray or CT images to learn spatial lung patterns; provide strong performance for visible abnormalities but do not reflect breathing dynamics
Audio-based CNN Models	Process lung sound recordings using spectrogram representations; useful for identifying wheezes and crackles but affected by background noise
Graph-enhanced Audio Models (CNN + GCN)	Represent respiratory sounds as graphs to capture relationships between features; improve class discrimination at the cost of higher computation
Traditional Machine Learning Models	Use handcrafted acoustic features such as MFCCs; lightweight and interpretable but limited in representing complex patterns
Multi-Modal Deep Learning Models	Combine medical images and lung sounds to leverage complementary information; achieve better diagnostic reliability with added system complexity
Attention-based Models	Highlight important regions in images or audio signals; enhance focus on relevant features but increase architectural overhead
Transformer based Models	Employ self-attention on raw audio or images to model long-range dependencies; deliver strong results but require substantial computational resources

Table 2.4 Comparative Analysis of Models for Respiratory Disease Diagnosis

III. TRENDS AND PATTERNS

1. Shift from Single-Modality to Multi-Modality

Initially, most studies relied only on chest X-ray image analysis using Convolutional Neural Networks. These models executed well for conditions like pneumonia and tuberculosis but struggled to generalize across datasets. Over time, researchers began integrating audio data, such as cough and breathing sounds, recognizing that combining structural and functional information leads to more comprehensive results. This transition from image-only to multi-modal approaches has marked a major trend in recent years.

2. Advancement of Deep Learning Architectures

Another noticeable pattern is the utilization of advanced architectures such as Transformers, Graph Convolutional Networks, and hybrid CNN-GCN models. These models can operate complex relationships between visual and acoustic features, offering higher precision in classification tasks. Additionally, the use of transfer learning has allowed researchers to attain high accuracy even with smaller medical datasets.

3. Focus on Explainability and Trust

A growing trend in healthcare AI is the shift toward Explainable AI (XAI). Since medical predictions directly affect patient outcomes, doctors need to understand how AI systems reach conclusions. Researchers are now including visualization tools that highlight which parts of an X-ray or audio clip contributed most to the diagnosis, improving model transparency and clinical trust [1], [2], [3].

4. Data Quality and Ethical Awareness

Recent studies also emphasize ethical data sourcing, data diversity, and dataset sharing. Many projects now rely on open-access datasets like NIH ChestX-rayl4, Kaggle Pneumonia Dataset, and ICBHI Respiratory Sound Database to encourage reproducibility and fairness across research [2], [3].

IV. RESEARCH GAPS AND CHALLENGES

While existing research in automated lung disease diagnosis has made impressive progress, several gaps and challenges still remain unaddressed. The studies reviewed highlight that although achieving high accuracy rates, current systems are not yet reliable or practical enough for large-scale clinical deployment.

4.1 Limited and Imbalanced Datasets

Most research depends on publicly available datasets such as NIH ChestX-rayl4, Kaggle Chest X-ray (Pneumonia), and the ICBHI Respiratory Sound Database. However, these datasets are often imbalanced [2], with more samples of certain diseases and fewer of others. This irregular distribution can cause models to perform well on dominant classes but poorly on rare conditions. Additionally, some datasets lack patient diversity, reducing the generalizability of AI models across age groups, environments, and device variations.

4.2 Single-Modality Dependence

A major gap identified in the literature is that many studies still depend on either image-based or audio-based diagnosis instead of integrating both. While multimodal approaches have been shown to enhance performance, they are still in early research stages and require more consistent frameworks for effective fusion of image and sound data

4.3 Computational and Resource Constraints

Deep learning models such as CNNs, GCNs, and Transformer architectures demand high computational power and large memory resources. This limits their real-time applicability, especially in low-resource healthcare environments where such infrastructure is unavailable.

4.4 Lack of Explainability

Another recurring challenge is the "black box" nature of many AI systems. Although some recent works use Explainable AI (XAI) to visualize attention areas, most models still do not provide sufficient interpretability for medical experts to confidently trust or verify the predictions [1], [3].

4.5 Integration into Clinical Practice

Even with promising accuracy, there is a gap between laboratory research and real-world implementation. Many studies stop at experimental validation without testing their models in real hospital environments. Concerns about model validation, patient privacy, and ethical approval also hinder smooth integration into clinical workflows.

V. CONCLUSION

This literature review explored the application of machine learning and deep learning techniques in the diagnosis of respiratory diseases. The collective findings emphasize that while traditional diagnostic approaches such as X-rays and auscultation remain important, their dependence on human expertise makes them prone to subjectivity and delays. Artificial Intelligence, particularly through machine learning and deep learning, has shown great potential in addressing these challenges by automating analysis, enhancing accuracy, and improving accessibility.

The reviewed studies consistently demonstrate that Convolutional Neural Networks (CNNs) are highly effective for medical image classification, especially in identifying lung-related abnormalities such as pneumonia and tuberculosis. Likewise, advanced architectures that analyze lung sounds including hybrid combinations of CNNs, Graph Convolutional Networks, and attention mechanisms have proven successful in distinguishing complex respiratory patterns.
The sound data outperform single-modality models by offering a more comprehensive and reliable evaluation of respiratory conditions.

Recent research trends also point toward the growing use of Vision Transformers (ViTs) for visual analysis and Explainable AI (XAI) techniques like Grad-CAM to increase interpretability and clinician trust. These advancements mark a shift toward more transparent and intelligent diagnostic frameworks.

Building upon these insights, the proposed future direction is to design a multi-modal diagnostic system that combines a CNN-based image model with a Wav2Vec2 Transformer for raw audio analysis. This hybrid framework aims to utilize the strengths of both data types, achieving higher diagnostic precision and real-world applicability. Overall, the conclusion drawn from this review is that multi-modal, explainable, and data-driven AI systems hold the key to developing next-generation respiratory diagnostic tools that are fast, accurate, and accessible across diverse healthcare environments.

VI. FUTURE SCOPE

Future research should focus on building more comprehensive diagnostic frameworks that merge medical imaging, audio data, and patient health records to deliver more personalized predictions. Expanding dataset diversity will improve the adaptability of AI models across populations and medical conditions. Another direction is the creation of lightweight, explainable, and energy-efficient AI architectures that can operate on mobile or portable diagnostic devices. This would allow real-time detection in remote or rural healthcare setups with limited access to specialists. Collaboration between medical institutions and AI researchers is also essential to develop standardized evaluation metrics and ensure ethical data handling. By combining interpretability, accuracy, and accessibility, future AI-driven systems can become vital tools for early respiratory disease detection, supporting global efforts toward affordable and inclusive healthcare.

VII. References

[1] V. P, D. R, and K. Chanthirasekaran, "A Machine Learning Approach to Lung Disease Identification: Combining X-Ray and Auscultation Data," in *2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 2025.

[2] M. Rajaselvi V, S. J, N. G M, and S. Koshy, "A Survey on Lung Disease Diagnosis using Machine Learning Techniques," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering*

[3] R. Anaadumba, N. A. Belabbaci, and M. A. U. Alam, "Enhancing Respiratory Disease Diagnosis with Graph Convolutional Networks Through Non-Speech Audio Synthesis," in *2025 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2025.

[4] . Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting

[5] lung diseases from X-ray images. Informatics in Medicine Unlocked, 20, 100391.

[6] Tabik et al. (2020). COVIDGR Dataset and COVID-SDNet Methodology for Predicting

[7] COVID-19 Based on Chest X-Ray Images. IEEE J. Biomed. Health Informatics.

[8] Sabry, A. H. et al. (2024). Lung disease recognition using audio-based analysis with machine learning. Heliyon, 10(4).

[9] Wanasinghe, T. et al. (2024). Lung Sound Classification for Respiratory Disease