



# Lung Sound–Driven Respiratory Disease Diagnosis using CNN–LSTM with GenAI and XAI

Shrineeta S Kurud, Dr. Ms. D. V. Gore

Department of Computer Engineering, P.E.S Modern College of Engineering, Pune, India

Department of Computer Engineering, P.E.S Modern College of Engineering, Pune, India

**ABSTRACT:** Respiratory diseases such as asthma, Chronic Obstructive Pulmonary Disease (COPD), and pneumonia remain significant global health concerns, requiring accurate and timely diagnosis for effective treatment. Traditional auscultation-based diagnosis is highly dependent on clinical expertise and may result in subjective interpretation. This paper presents a comprehensive survey of respiratory disease detection approaches using lung sound analysis, with a focus on hybrid Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) architectures, Explainable AI, and Generative AI integration.

The survey reviews existing frameworks that process respiratory sound recordings in WAV format and perform preprocessing, noise reduction, and feature extraction using Mel-Spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs). CNN layers are utilized for spatial feature extraction, while LSTM layers capture temporal dependencies in respiratory signals. The role of Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM and SHAP in improving interpretability and clinical trust is analyzed. Additionally, the emerging use of Generative AI modules for providing preliminary treatment recommendations and preventive healthcare suggestions is discussed.

Based on the reviewed literature, a conceptual framework integrating a hybrid CNN–LSTM model, XAI-based interpretability, and GenAI-assisted recommendations—deployed as a Flask-based web application—is synthesized. Analysis of existing studies demonstrates that hybrid CNN–LSTM models achieve superior classification performance compared to standalone models, making them a scalable and effective approach for early respiratory disease screening in resource-constrained environments.

**KEYWORDS:** Respiratory Disease Detection, Lung Sound Analysis, Hybrid CNN-LSTM, Deep Learning, MFCC, Explainable AI, Grad-CAM, SHAP, Generative AI, Healthcare AI.

## I. INTRODUCTION

Respiratory diseases such as asthma, Chronic Obstructive Pulmonary Disease (COPD), pneumonia, bronchitis, and other pulmonary disorders are major global health challenges, contributing significantly to morbidity and mortality worldwide. Early and accurate diagnosis is essential for effective treatment and prevention of severe complications. However, conventional diagnostic methods such as chest auscultation, pulmonary function tests, X-ray imaging, and CT scans often require experienced clinicians, expensive equipment, and advanced healthcare infrastructure, making timely diagnosis difficult in remote and resource-constrained regions.

Among the available diagnostic approaches, auscultation using a stethoscope remains one of the most widely used preliminary screening techniques because of its non-invasive and cost-effective nature. Lung sounds such as wheezes, crackles, and rhonchi provide valuable information about respiratory abnormalities. However, manual interpretation of these sounds is subjective and highly dependent on clinical expertise, which may lead to inconsistent diagnosis and delayed treatment. This has increased the demand for automated and intelligent respiratory disease detection systems.

Recent advancements in digital signal processing and artificial intelligence have enabled efficient computational analysis of lung sound recordings. Preprocessing techniques such as noise reduction, filtering, normalization, and segmentation improve signal quality, while feature extraction methods including Mel-Spectrograms, Short-Time Fourier Transform (STFT), and Mel-Frequency Cepstral Coefficients (MFCCs) capture important time–frequency characteristics of respiratory sounds. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have shown significant success in respiratory sound classification. CNNs

effectively extract spatial features from spectrogram representations, whereas LSTM networks model temporal dependencies in sequential audio signals. Hybrid CNN–LSTM architectures combine these capabilities to achieve improved feature learning and classification accuracy.

Despite the strong performance of deep learning models, their black-box nature remains a challenge in healthcare applications where interpretability and reliability are essential. Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM and SHAP improve transparency by identifying important spectrogram regions and feature contributions influencing predictions. In addition, recent advancements in Generative AI have enabled intelligent healthcare assistance systems capable of providing preliminary treatment recommendations and preventive guidance based on predicted respiratory conditions.

This paper presents a comprehensive review of respiratory disease detection systems based on lung sound analysis, focusing on signal preprocessing techniques, hybrid CNN–LSTM architectures, Explainable AI approaches, and AI-assisted healthcare systems. The paper highlights recent advancements, existing challenges, and future research directions in the development of intelligent and interpretable respiratory disease diagnosis systems.

## II. LITERATURE SURVEY

Respiratory disease detection using lung sound analysis has gained significant attention in recent years due to advancements in digital signal processing, deep learning, and artificial intelligence. Earlier research primarily relied on traditional machine learning techniques combined with handcrafted acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral contrast, zero-crossing rate, and Short-Time Fourier Transform (STFT) for identifying abnormal respiratory sounds. Although these methods achieved moderate classification performance, they required manual feature engineering and often struggled to capture complex temporal and spectral patterns present in lung sound recordings. With the emergence of deep learning, researchers shifted toward automated feature extraction and end-to-end classification systems capable of learning discriminative representations directly from respiratory audio signals.

Recent studies have focused on developing efficient and accurate deep learning architectures for respiratory disease classification. In [1], the RDLINet model introduced a lightweight inception-based deep learning architecture designed to reduce computational complexity while maintaining high diagnostic accuracy. The model utilized multi-scale convolutional operations for efficient feature extraction from respiratory spectrograms, making it suitable for real-time healthcare systems and edge-device deployment. In [2], a hybrid CNN–LSTM framework was proposed in which respiratory sounds were converted into STFT-based spectrograms before classification. CNN layers extracted spatial and frequency-domain features, while LSTM layers captured temporal dependencies and sequential respiratory patterns, resulting in improved classification of diseases such as asthma, COPD, and pneumonia. Similarly, ConvLSNet proposed in [4] integrated ConvLSTM, 3D CNN, and LSTM layers to simultaneously learn spatial and temporal characteristics from lung sound spectrograms. The model achieved approximately 97.4% classification accuracy and demonstrated improved efficiency compared to conventional CNN architectures. In [5], Lung SoundNet combined advanced statistical and spectral feature extraction methods with optimized LSTM classifiers to improve robustness and diagnostic accuracy. DeepRespNet presented in [6] introduced both 1D and 2D deep learning models trained on raw waveform signals and spectrogram representations, proving that combining multiple signal representations enhances generalization and classification performance across different respiratory conditions.

Several studies also emphasized comparative analysis, advanced feature learning strategies, and intelligent auscultation systems for automated diagnosis. The comparative study in [7] evaluated CNN, LSTM, and hybrid CNN–LSTM models for respiratory sound classification and reported that LSTM achieved very high accuracy due to its capability to effectively model temporal dependencies in sequential respiratory signals. Hybrid CNN–LSTM models also demonstrated competitive performance by combining spatial feature extraction and temporal sequence learning. In [8], researchers proposed a dual-channel CNN–LSTM architecture using Mel-spectral features and data augmentation techniques such as time shifting, pitch modification, and noise addition to improve feature diversity and model robustness. The proposed model achieved classification accuracy above 99%, highlighting the importance of multi-channel feature learning and augmentation strategies in biomedical audio analysis. In addition, [10] explored the application of Vision Transformer (ViT) architectures for adventitious lung sound classification using cephalogram representations. Unlike traditional CNN-based approaches, the transformer model utilized attention mechanisms to

capture complex global relationships within respiratory spectrograms and demonstrated competitive performance for respiratory disease detection.

Apart from classification performance, recent research has increasingly focused on intelligent healthcare systems, explainability, and telemedicine integration. The review presented in [3] discussed intelligent digital stethoscope systems integrated with deep learning algorithms for automated lung sound analysis. The study highlighted the role of AI-based auscultation in reducing subjectivity associated with manual diagnosis and enabling remote healthcare monitoring through telemedicine platforms. It also discussed preprocessing techniques, noise reduction methods, dataset limitations, and deployment challenges in real-world clinical environments. Similarly, the systematic review in [9] analysed recent developments in audio-based respiratory disease diagnosis using machine learning and deep learning approaches. The study reviewed commonly used feature extraction methods such as MFCC, STFT, wavelet transforms, and Mel-spectrograms along with classification architectures including CNNs, LSTMs, hybrid models, and transformer-based networks. Furthermore, the review emphasized major challenges such as dataset imbalance, environmental noise, lack of standardized datasets, model interpretability, and the need for Explainable Artificial Intelligence (XAI) in healthcare applications. Overall, the literature demonstrates a clear transition from conventional signal processing techniques to advanced hybrid deep learning, transformer-based, and explainable AI systems aimed at developing accurate, interpretable, and scalable respiratory disease diagnosis frameworks.

### III. METHODOLOGY

This survey adopts a systematic literature review approach to identify, evaluate, and synthesize existing research on deep learning-based respiratory disease detection using lung sound analysis. The methodology is organized into five key phases: research question formulation, literature search, study selection, data extraction, and conceptual framework synthesis.

#### Research Questions

The study is guided by the following research questions: (1) What deep learning architectures are most used for lung sound classification? (2) Which audio feature extraction techniques are most effective for representing respiratory signals? (3) How do hybrid models such as CNN-LSTM compare with standalone architectures in classification performance? (4) What role do Explainable AI techniques play in improving model transparency for clinical applications? (5) How can Generative AI extend diagnostic systems into decision-support tools?

#### Search Strategy and Study Selection

A comprehensive literature search was conducted across IEEE Xplore, PubMed, ScienceDirect, Springer Link, and Google Scholar using keyword combinations such as "lung sound classification," "respiratory disease detection," "CNN-LSTM," "Mel-spectrogram," "MFCC," "Explainable AI," "Grad-CAM," "SHAP," and "Generative AI." The search was limited to peer-reviewed journal articles and conference papers published between 2020 and 2025.

Studies were selected through a three-stage screening process: title and abstract screening, full-text evaluation, and final quality assessment. Inclusion criteria required studies to focus on automated respiratory disease detection using audio signals, employ deep learning or machine learning models, and report quantitative evaluation metrics. Studies based solely on imaging modalities, non-peer-reviewed publications, duplicates, and papers with insufficient methodological detail were excluded. Through this process, 10 primary studies were selected for in-depth review along with 8 supplementary references covering foundational techniques and tools.

#### Data Extraction and Comparative Analysis

From each selected study, key information was systematically extracted, including the dataset used, preprocessing techniques, feature extraction methods, model architecture, evaluation metrics (accuracy, precision, recall, F1-score), and integration of explainability techniques. The extracted data was organized into a structured comparison to identify common trends, performance benchmarks, and research gaps. The analysis revealed the dominance of CNN and LSTM-based architectures, the widespread use of Mel-spectrograms and MFCCs as preferred features, and reported classification accuracies ranging from 94% to over 99% across reviewed studies.

#### Identification of Research Gaps

The comparative analysis also helped identify key research gaps across the surveyed studies. These include the limited availability of large-scale, clinically validated lung sound datasets, insufficient integration of Explainable AI

in existing diagnostic systems, limited exploration of advanced architectures such as vision transformers and attention mechanisms for lung sound classification, and the need for lightweight models suitable for deployment on edge devices in resource-constrained environments. These identified gaps inform the future scope of this survey and provide direction for subsequent research in the field.

#### IV. PROPOSED CONCEPTUAL FRAMEWORK

Based on the comprehensive analysis of existing literature and the identification of key techniques, architectures, and methodologies across the reviewed studies, this section presents a conceptual framework that synthesizes the most effective components into a unified respiratory disease detection system. This framework is not an independently developed or experimentally validated system, but rather a reference architecture derived from the collective insights of the surveyed literature, intended to demonstrate how the identified building blocks can be integrated into a cohesive and intelligent diagnostic platform.

##### Framework Overview

The proposed conceptual framework consists of five interconnected modules: (1) User Interface and Web Application Layer, (2) Audio Preprocessing and Feature Extraction Module, (3) Hybrid CNN–LSTM Classification Module, (4) Explainable AI (XAI) Module, and (5) Generative AI (GenAI) Recommendation Module. The workflow begins with user authentication and lung sound upload through a web interface, followed by signal preprocessing, feature extraction, deep learning–based classification, interpretability generation, and AI-assisted treatment recommendations.

##### Audio Preprocessing and Feature Extraction

The uploaded lung sound recordings undergo a series of preprocessing steps to improve signal quality and ensure consistency across inputs. These steps include resampling to a standardized sampling rate of 16 kHz, bandpass filtering in the range of 100–2000 Hz to remove irrelevant environmental noise, amplitude normalization to maintain uniform signal levels, and segmentation into fixed-length frames corresponding to respiratory cycles. Following preprocessing, time–frequency features are extracted using two widely adopted techniques identified across the reviewed literature: Mel-spectrograms, which capture perceptual frequency characteristics of respiratory sounds, and Mel-Frequency Cepstral Coefficients (MFCCs), which provide compact and discriminative spectral representations. These features are converted into two-dimensional representations suitable for input to the deep learning classification module.

##### Hybrid CNN–LSTM Classification Module

The core classification component of the conceptual framework employs a hybrid CNN–LSTM architecture, reflecting the strong consensus across the reviewed studies that combining spatial and temporal learning yields superior performance for lung sound classification. The CNN module applies multiple convolutional layers with pooling operations to the spectrogram input, extracting spatial features such as frequency patterns, spectral structures, and localized acoustic characteristics. The output feature maps from the CNN module are then reshaped and passed sequentially to the LSTM module, which processes the temporal sequence of features to capture long-range dependencies and dynamic patterns across respiratory cycles. The LSTM module utilizes memory cells and gating mechanisms to selectively retain and propagate relevant temporal information. The final classification is performed through fully connected dense layers with a SoftMax activation function, producing probability scores across multiple disease classes including Normal, Asthma, COPD, and Pneumonia.

##### Explainable AI (XAI) Module

To address the critical need for transparency and clinical trust identified across the literature, the conceptual framework integrates two complementary Explainable AI techniques. Gradient-weighted Class Activation Mapping (Grad-CAM) generates visual heatmaps overlaid on the input spectrograms, highlighting the specific time–frequency regions that most strongly influence the model's prediction. This allows clinicians to visually verify whether the model focuses on acoustically meaningful regions such as wheezes, crackles, or other abnormal sound patterns. Shapley Additive explanations (SHAP) provide quantitative feature importance scores, indicating the contribution of individual input features to the final classification output. Together, these techniques enable both visual and numerical interpretability of the model's decision-making process.

### Generative AI (GenAI) Recommendation Module

The framework extends beyond classification by incorporating a Generative AI module that transforms the system into a decision-support tool. Based on the predicted disease class and the associated confidence score, the GenAI module generates preliminary treatment suggestions, preventive healthcare measures, and lifestyle recommendations relevant to the detected respiratory condition. The generated content is clearly presented as informational guidance and not as clinical prescriptions, intended to support awareness and initial decision-making in settings where immediate access to specialist consultation may be limited

## V. CONCLUSION

This survey paper presented a comprehensive review of deep learning–based respiratory disease detection systems using lung sound analysis. Through a systematic examination of recent literature, the study identified and analyzed key advancements in audio preprocessing techniques, feature extraction methods, deep learning architectures, and model interpretability approaches applied to automated respiratory diagnosis.

The review revealed that hybrid architecture, particularly those combining Convolutional Neural Networks (CNNs) for spatial feature extraction with Long Short-Term Memory (LSTM) networks for temporal sequence modeling, consistently achieve superior classification performance compared to standalone models across the surveyed studies. Time–frequency feature representations such as Mel-spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs) were found to be the most widely adopted and effective techniques for capturing discriminative acoustic patterns in lung sounds, with reported classification accuracies ranging from 94% to over 99% across the reviewed literature.

The survey also highlighted the growing significance of Explainable Artificial Intelligence (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley Additive explanations (SHAP), in enhancing model transparency and fostering clinical trust—addressing a critical barrier to the adoption of deep learning in medical practice. Additionally, the integration of Generative AI (GenAI) for providing preliminary treatment suggestions and preventive recommendations was identified as a promising direction that extends diagnostic systems into comprehensive clinical decision-support tools.

Based on the insights gathered from the literature, a conceptual framework integrating a hybrid CNN–LSTM model, XAI-based interpretability, and GenAI-assisted recommendations—deployed as a Flask-based web application—was discussed to illustrate how these components can be unified into an accessible, scalable, and intelligent respiratory disease detection system. Overall, this survey provides researchers and practitioners with a structured understanding of the current state of the art and identifies key directions for advancing intelligent respiratory healthcare systems.

## VI. FUTURE SCOPE

While the reviewed studies demonstrate significant progress in deep learning–based respiratory disease detection, several areas remain open for further research and improvement:

**Dataset Expansion and Standardization:** Most existing studies rely on limited or proprietary lung sound datasets. Future research should focus on developing large-scale, clinically validated, and publicly available respiratory sound datasets with diverse patient demographics, recording conditions, and disease categories to enable fair benchmarking and improved model generalization.

**Real-Time Monitoring and IoT Integration:** Integration of deep learning models with digital stethoscopes, wearable sensors, and Internet of Things (IoT)–based monitoring devices can enable continuous and real-time respiratory health monitoring, particularly for patients with chronic conditions.

**Edge AI and Lightweight Model Deployment:** Deploying optimized and compressed deep learning models on edge computing platforms and embedded devices such as Raspberry Pi or ESP32 can reduce latency, enable offline functionality, and extend accessibility to remote and resource-constrained environments.

**Advanced Architectures:** Exploration of transformer-based models, attention mechanisms, and self-supervised learning approaches may further improve feature representation and classification accuracy for complex respiratory sound patterns that are difficult to distinguish using conventional architectures.

**Multi-Modal Diagnosis:** Combining lung sound analysis with other diagnostic modalities such as chest X-ray imaging, spirometry data, and patient clinical history can enable more comprehensive and robust multi-modal diagnostic systems.

**Enhanced Explainability and Clinical Validation:** Further development and clinical validation of Explainable AI methods tailored for healthcare applications will be essential for building clinician trust and facilitating regulatory approval of AI-assisted diagnostic systems.

**Personalized and Adaptive Systems:** Future systems could incorporate patient-specific adaptation and longitudinal monitoring capabilities, enabling personalized diagnosis and treatment tracking over time.

## REFERENCES

- [1] A. Sharma, R. Gupta, and P. Verma, "RDLINet: A Lightweight Deep Learning Model for Respiratory Disease Classification Using Lung Sounds," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023.
- [2] S. Patel and M. K. Singh, "Hybrid CNN–LSTM Based Respiratory Disease Classification Using Lung Sound Analysis," *International Journal of Power Electronics and Drive Systems*, vol. 15, no. 2, pp. 1123–1132, 2024.
- [3] J. Kim, H. Lee, and S. Park, "Deep Learning Approaches for Intelligent Stethoscope Systems: A Review," *Military Medical Research*, vol. 10, no. 1, pp. 1–15, 2023.
- [4] Y. Zhang, X. Chen, and L. Wang, "ConvLSNet: A Deep Learning Model Combining ConvLSTM and CNN for Lung Sound Classification," *Artificial Intelligence in Medicine*, vol. 150, pp. 102–110, 2024.
- [5] M. Rahman, T. Hossain, and A. Islam, "Lung SoundNet: An Optimized Deep Learning Framework for Respiratory Disease Detection," *Biomedical Signal Processing and Control*, vol. 90, pp. 105–115, 2025.
- [6] K. Verma and D. Sharma, "DeepRespNet: Deep Learning Framework for Respiratory Sound Classification Using Time and Frequency Domain Features," *Biomedical Signal Processing and Control*, vol. 85, pp. 104–112, 2024.
- [7] L. Garcia, P. Martinez, and R. Lopez, "Comparative Analysis of CNN, LSTM, and Hybrid Models for Respiratory Disease Detection," *Frontiers in Medicine*, vol. 10, pp. 1–12, 2023.
- [8] H. Wang, Y. Liu, and Z. Sun, "Dual-Channel CNN–LSTM Model for Lung Sound Classification Using Mel-Spectral Features," *Biomedical Signal Processing and Control*, vol. 88, pp. 104–113, 2024.
- [9] R. Kumar, S. Jain, and A. Mishra, "A Systematic Review of Audio-Based Respiratory Disease Detection Using Machine Learning," *Sensors*, vol. 24, no. 4, pp. 1173, 2024.
- [10] F. Ali, M. Khan, and N. Ahmad, "Vision Transformer-Based Lung Sound Classification Using Cochleogram Features," *Sensors*, vol. 24, no. 2, pp. 682, 2024.
- [11] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.
- [12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. ICLR*, 2015.
- [13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [14] R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc. IEEE ICCV*, pp. 618–626, 2017.
- [15] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [16] B. McFee *et al.*, "Librosa: Audio and Music Signal Analysis in Python," in *Proc. Python in Science Conference*, 2015.
- [17] T. Virtanen, M. D. Plumbley, and D. Ellis, *Computational Analysis of Sound Scenes and Events*, Springer, 2018.
- [18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details