Review of advancements in AI-assisted lung sound analysis for respiratory illness diagnosis in noisy environments

Reshma Sreejith, R. Kanesaraj Ramasamy, Wan-Noorshahida Mohd-Isa, Junaidi Abdullah
Faculty of Computing and Informatics, Multimedia University Cyberjaya, Cyberjaya, Malaysia

Article Info

Article history:

Received Nov 3, 2024
Revised Jan 27, 2026
Accepted May 23, 2026

Keywords:

Artificial intelligence
Auscultation analysis
Deep learning
Noisy sounds
Respiratory sounds
Systematic review

ABSTRACT

For several centuries, research has been carried out to address respiratory ailments, which are among the most detrimental to human health. The advent of the stethoscope in the 19th century has facilitated the identification of respiratory sounds. This innovation represents a significant advancement in the identification and diagnosis of numerous respiratory ailments. In Malaysia, public hospitals have traditionally employed stethoscopes in their emergency departments. However, the precision of readings obtained through this method is susceptible to interference from ambient noise, uneven terrain, and suboptimal acoustic performance, particularly during medical transportation. Consequently, this can result in erroneous diagnoses and inappropriate treatment. Potential remedies for addressing the challenges associated with assessing respiratory sounds during medical transportation include advancements in stethoscope technology, novel auditory techniques, and reduced levels of background noise within the transportation environment. The present investigation concerns the effects of developing a new machine learning (ML) algorithm for the assessment of lung sound in conditions of high ambient noise. The objective is to devise a ML algorithm that can categorize acute respiratory illnesses based on their level of urgency in the presence of ambient noise.

Corresponding Author:

R. Kanesaraj Ramasamy
Faculty of Computing and Informatics, Multimedia University Cyberjaya
Cyberjaya, Malaysia
Email: r.kanesaraj@mmu.edu.my

1. INTRODUCTION

In contemporary times, the utilization of advanced technologies such as machine learning (ML) and deep learning (DL) has facilitated the provision of viable remedies to challenges encountered in the medical field. Furthermore, the utilization of audio and medical image analysis enhances the accuracy of disease prognosis. Medical professionals value the use of technological support because it enables them to efficiently manage a larger patient population in light of the shortage of qualified human resources [1]. The incidence of respiratory ailments is progressively increasing and posing a threat to society, alongside severe illnesses such as cancer and diabetes. The utilization of chest X-rays and the analysis of respiratory sounds have demonstrated significant utility in the timely identification and management of respiratory pathologies. General practitioners or family doctors frequently include auscultation in their physical examinations. The process is expeditious, uncomplicated, and does not necessitate contemporary technology. The fundamental attributes of auscultation encompass non-invasiveness, simplicity, and mobility, irrespective of the stethoscope variant employed.

The result of this particular examination is inherently subjective and heavily dependent on the proficiency and perceptual abilities of medical professionals, rendering them vulnerable to considerable imprecision. Undoubtedly, there is a notable divergence of opinions among medical professionals. To attain an accurate and precise medical diagnosis, it can be imperative to differentiate between typical respiratory sounds and atypical ones. Respiratory sounds offer significant insights into the pathophysiology of lung diseases. and airway obstruction. Several scholarly articles have illustrated the extent and significance of this issue. Despite the advancements in the medical field, there is still a lack of a universally standardized classification system for the diverse respiratory system anomalies observed in humans. Medical practitioners often employ specialized terminology with specific semantic connotations, as outlined in various academic texts and institutions to characterize purported respiratory disorders. This phenomenon gives rise to challenges in introductory auscultation courses and, more notably, in subsequent professional practice when healthcare providers collaborate or exchange diagnoses [2].

Advances in artificial intelligence (AI) have greatly improved the potential for autonomous disease identification through the analysis of biosignals. Respiratory sounds that exhibit audible alterations due to inflammation in the respiratory system can potentially function as diagnostic markers for a range of disorders. The adaptability and scalability of AI-powered auscultation, which involves the utilization of respiratory sounds acquired from smartphones and electronic stethoscopes, have garnered considerable interest. Traditionally, respiratory physicians were responsible for conducting conventional auscultation. However, the process of training these professionals can prove to be both time-intensive and financially burdensome [3]. In addition, obtaining a medical diagnosis in a hospital or clinic setting results in elevated healthcare expenses and heightened susceptibility to viral exposure. The utilization of ML algorithms and acoustic signal processing techniques is imperative in attaining precise diagnoses in the applications. Two predominant ML methodologies that are frequently utilized are end-to-end DL and feature-based ML. To identify pathological sounds, ML models rely on prosodic temporal features such as pitch, duration, intensity, harmonic-to-noise ratio, jitter, and shimmer. Furthermore, the spectral features obtained from the logarithmic Mel-spectrogram have demonstrated encouraging outcomes in diverse domains. Subsequently, these characteristics are input into classifiers to facilitate the process of diagnosis.

In contrast, end-to-end DL methodologies entail the direct input of audio waves or their corresponding spectrograms into deep neural networks for the purpose of generating predictions [4]. This review will examine the algorithms, advancements, and disease applications of sound-based diagnostic tools for the lung and respiratory systems. Thus, the characteristics and contributions of this study are as follows:

i) This study can help researchers interested in sound-based illness analysis understand the development patterns and characteristics of employing such prediction techniques, allowing them to consciously select the most appropriate algorithms in their research process.
ii) The key trends in prospective medical diagnosis and trends in integrating digital processing are examined, demonstrating that audio-based illness algorithms combined with DL have a promising future.
iii) The review searches for the existing challenges of lung disease detection with DL, such as limited samples in the utilized dataset, low quality of data, unbalanced data, and poor interpretability, to propose the available suitable remedies.
iv) This article summarizes recent audio-based deep-learning algorithms for disease classification using several comparison tables.

2. RELATED WORKS

During the late 1980s, neural networks gained prominence in ML and AI due to advances in learning techniques and network architectures. However, interest waned with the rise of DL in 2006. DL, which relies on artificial neural networks, has since surged in popularity for its ability to learn from data, making it essential in ML, AI, data science, and analytics. Companies like Google, Microsoft, and Nokia have invested heavily in DL for tasks like classification and regression, leveraging its ability to mimic the human brain's data processing. Though DL requires long training times due to its vast number of parameters, it outperforms other ML algorithms during testing by executing faster.

According to Hazra and Majhi [5], convolutional neural network (CNN)-based model categorizes six respiratory diseases with 92.39% accuracy. Similarly, Tariq *et al.* [6] developed system for classifying lung diseases using multiple modalities achieved a 97% accuracy by integrating audio-based and image-based classification methods. Fraiwan *et al.* [7] highlighted the efficacy of support vector machines (SVM) for detecting breathing sounds like wheezing and rhonchi, with accuracies ranging from 97.7% to 98.8%. Their methodology utilized time-frequency analysis and advanced statistical methods, improving the accuracy of lung sound classification.

The use of hybrid CNN-long short-term memory (LSTM) models, as presented by Alqudah *et al.* [8] further demonstrated improved classification performance on datasets from King Abdullah University Hospital (KAUH) and International Conference on Biomedical and Health Informatics (ICBHI). They improved the model's credibility by tuning its hyperparameters, which led to better results than using CNN and LSTM models separately. Chanane and Bahoura [9] employed CNNs for a 4-class respiratory condition classification task, achieving an 80.4% accuracy using data normalization and hyperparameter optimization. Additionally, Pouyani *et al.* [10] developed a denoising technique using artificial neural networks and wavelet transforms, significantly improving signal quality in classifying lung sounds.

Haider and Behera [11] introduced the sparsity-assisted signal smoothing (SASS) method for denoising lung sounds, achieving exceptional signal-to-noise ratio (SNR) improvements, while Tamal *et al.* [12] developed a hybrid AI algorithm for diagnosing pulmonary sounds with a 100% accuracy rate using time-frequency analysis. Abdullah and Er [13] introduced a CNN-LSTM hybrid model with preprocessing techniques such as Butterworth filtering and wavelet decomposition, achieving over 90% accuracy in most instances. Recent research in ML and DL for respiratory sound classification continues to show enormous promise. Techniques such as hybrid ensembles combining different DL architectures are being explored to enhance classification accuracy further. The use of AI-assisted lung sound analysis offers significant potential for diagnosing respiratory conditions, particularly in noisy environments, although challenges related to noise mitigation and accurate classification persist. Recent research has focused on using ML and DL techniques for automatically classifying respiratory sounds, particularly pulmonary crackles. Vasava and Joshiara [14] attained a 97.5% accuracy rate with an SVM classifier for differentiating between normal and wheeze lung sounds, employing Mel-frequency cepstral coefficients (MFCC) and a Gaussian mixture model, which yielded a 94.2% accuracy rate. Rishabh and Kumar [15] used audio features like MFCC, Chroma short-time Fourier transform (STFT), and Mel-spectrogram with a CNN architecture, achieving 74.08% and 75.04% accuracy for 60-40 and 80-20 train-test splits, outperforming other methods. Bapa *et al.* [16] used a 2D CNN model for classifying various respiratory conditions, achieving 94.90% accuracy. Despite its impressive performance, the model struggled with conditions like asthma and lower respiratory tract infection (LRTI).

Taloba and Matoog [17] developed a ML model for chronic obstructive pulmonary disease (COPD) detection using MFCC feature extraction and an SVM-k-nearest neighbors (KNN) classifier. The study achieved high accuracy in distinguishing COPD patients from healthy individuals, showing the potential of ML-based systems for early respiratory disease diagnosis. However, limitations include dataset dependency, lack of real-time monitoring, limited clinical validation, and challenges in healthcare system integration. Sfayyih *et al.* [18] introduce a multimodal system that integrates audio and visual data, leveraging CNNs to achieve high diagnostic accuracy, though some misclassification issues remain. Abdullah *et al.* [19] present an AI-powered framework with exceptional performance in both disease classification and personalized medication recommendations, showcasing its robustness against noise and potential for real-world implementation. Each study offers valuable insights into advancing the field of lung disease diagnosis.

To summarize, the utilization of AI-assisted lung sound analysis exhibits great potential as a means of identifying respiratory ailments in environments with high levels of ambient noise. Additional investigation is required to authenticate its effectiveness; however, its capacity to enhance diagnostic precision and patient results renders it a stimulating domain for advancement in the healthcare sector. In general, the utilization of AI-assisted lung sound analysis represents a promising novel technology that has the capacity to enhance the diagnosis and management of respiratory illnesses. In spite of the remarkable advancements made thus far, it is undeniable that the smart stethoscope continues to encounter certain impediments [20]. Given the prevailing disruptions within the authentic medical settings where auscultation takes place, it becomes imperative to exercise caution in the process of capturing and deciphering respiratory sounds. The mitigation of noise constitutes a paramount and arduous endeavor in the realm of mechanical apparatuses and their corresponding analytical algorithms. While respiration sounds may occasionally appear on their own, they often coexist with one or more other auditory manifestations. These issues suggest that processing audio information under chaotic conditions is necessary to enhance sound quality. This would aid in the classification of a greater diversity of automatically audible sounds [21].

2.1. Performance of algorithms and generalization in noisy environments

Compared to several studies on respiratory sound analysis, handling noisy environments has consistently emerged as a critical challenge. Numerous researchers have investigated the effectiveness of various algorithms in such settings, with particular attention to how well these models generalize across different clinical environments and devices. A common finding in these studies is that DL models, especially CNNs, tend to outperform traditional ML models due to their ability to automatically learn hierarchical features from raw audio data. This feature extraction capability is particularly useful in noisy environments, where traditional methods often fail. For instance, Sfayyih *et al.* [18] demonstrated that CNNs show robust performance even in noisy settings, especially when models are trained with noise-augmented datasets, highlighting the effectiveness of CNNs in environments where environmental disturbances such as medical equipment sounds and conversations are prevalent. However, several studies have suggested that hybrid models, particularly those combining CNNs and LSTM networks, outperform traditional CNNs in highly dynamic and noisy conditions. These hybrid models, as shown by Alqudah *et al.* [8], benefit from the CNN's ability to extract spatial features and the LSTM's capacity to capture temporal dependencies in sequential data, thus enhancing their noise robustness. This combination allows the models to be more adaptable to real-world complexities found in clinical environments, where ambient noise from various sources such as hospital staff, equipment, and patient movements can significantly interfere with the accuracy of sound analysis.

Regarding the datasets used in respiratory sound analysis, they play a crucial role in determining the model's ability to generalize to noisy environments. The ICBHI dataset, for example, has been widely used due to its inclusion of real-world ambient noise from critical care units, where environmental disturbances such as alarms, equipment sounds, and staff conversations are common [22]. This makes it an ideal resource for training models that must perform reliably in noisy clinical settings. On the other hand, datasets like RALE, while comprehensive in terms of lung sound variations, lack higher levels of ambient noise, which may limit the generalization of models trained exclusively on such data. To address this limitation, many studies, including those by Bapa *et al.* [16] have utilized synthetic noise augmentation techniques. These methods artificially introduce noise into the training data, thereby improving the model's generalization in unpredictable, real-world environments where background noise is an unavoidable factor.

Ensuring that AI models generalize well across diverse clinical settings is an essential aspect of their development. A common strategy to evaluate this generalization is cross-site validation, where models trained on data from one hospital or healthcare facility are tested on data from other sites. This approach has been employed by researchers such as Alqudah *et al.* [8], who validated their CNN-LSTM hybrid model across multiple hospitals to ensure that the model's performance was not limited to data from a single institution. Cross-site validation is critical for assessing the model's ability to generalize to diverse patient populations and healthcare environments. However, cross-device validation remains a less explored but equally important area of study. Models trained on data collected from traditional stethoscopes may not perform well with data recorded using smartphones or electronic stethoscopes, primarily due to differences in microphone quality, noise filtering capabilities, and device-specific characteristics. This issue has yet to be fully addressed, and it is essential to ensure that models can perform reliably across various devices. This challenge is particularly pertinent as mobile health applications are becoming more widely used in healthcare, necessitating further research into cross-device validation to ensure the robustness of AI models across different recording technologies [23].

2.2. Existing models for respiratory sound classification: a comparative analysis

Table 1 provides a comparative overview of various sound analysis techniques, summarizing their accuracy ranges, feature types, strengths, limitations, and additional factors such as training time, evaluation metrics, and generalization capability. The table includes traditional ML models (e.g., SVM), DL models (e.g., CNN and CNN + LSTM). Hybrid approaches such as ANN + CNN and empirical mode decomposition (EMD)/multi-constrained nonnegative matrix factorization (MCNMF). Table 1 highlights the comparative strengths and limitations of different sound analysis models, showing clear performance differences between traditional ML approaches and DL based techniques.

Table 1. Comparison of various ML and DL models applied to respiratory

Model type	Reference	Accuracy range (%)	Feature types	Dataset	Strengths	Limitations	Applications
CNN	[5], [6], [9], [15], [16], [17], [18]	61-96.7	Spectrogram, MFCC, log quantization, and unintentional noise filters	ICBHI 2017 dataset	High accuracy and robust to feature variation	Lacks temporal dynamics and may overfit on small datasets	Medical diagnostics (e.g., lung disease)
CNN + LSTM / BDLSTM	[7], [8], [13], [19]	90-98.43	Spectrogram + sequential/ temporal features	ICBHI 2017 challenge dataset and KAUH dataset.	Captures both spatial and temporal patterns and better generalization	Resource-intensive, requires large, and labeled datasets	Early-stage disease detection
SVM	[24], [25]	74-92	Spectral roll-off, lacunarity, and denoising autoencoder	ICBHI 2017 dataset	Simple, interpretable and good for small or labeled datasets	Lower performance on complex or noisy inputs	Lung disease detection
ANN + CNN	[12]	95	Handcrafted features + integrated ML features	ICBHI 2017 dataset	Combines strengths of ML and DL and hybrid learning	May suffer from data/device variation and overfitting	Complex medical sound classification
EMD/MCNMF	[26], [27]	65-95	Instantaneous frequency and peak frequency	ICBHI 2017 dataset	Effective for signal decomposition and separation tasks	Limited generalizability and outdated compared to DL methods	Signal processing and separation tasks

While models such as SVM offer lower computational complexity and faster training, their performance and generalization capabilities are often limited when dealing with complex audio patterns. In contrast, DL and hybrid models. particularly CNN + LSTM consistently demonstrate higher accuracy and improved robustness by effectively learning both spatial and temporal features from audio data. Nevertheless, these models require larger training datasets, greater computational resources, and careful tuning to mitigate issues such as overfitting. Therefore, the selection of an appropriate model depends on the specific application requirements, dataset size, and available computational resources, emphasizing the trade-off between performance, efficiency, and implementation complexity [28], [29].

3. METHOD

3.1. Methodological approach: PRISMA framework

The study selection process followed a systematic review methodology, beginning with the identification of studies through database searches. A total of 251 records were identified from four databases: ACM (50), IEEE (65), ScienceDirect (6), and Google Scholar (130). After removing duplicates (22 records) and other irrelevant records (4 records), 225 records remained for screening. During the screening phase, 173 records were excluded based on title and abstract review, leaving 52 reports to be sought for retrieval. However, only 50 reports were successfully retrieved, while 2 reports could not be retrieved. These 50 reports were then assessed for eligibility, and none were excluded, indicating that all 50 reports met the eligibility criteria. Consequently, 50 studies were included in the final review. Figure 1 illustrates the PRISMA flow diagram summarizing the study identification, screening, eligibility, and inclusion process, in accordance with PRISMA guidelines to ensure transparency and rigor in the systematic review.

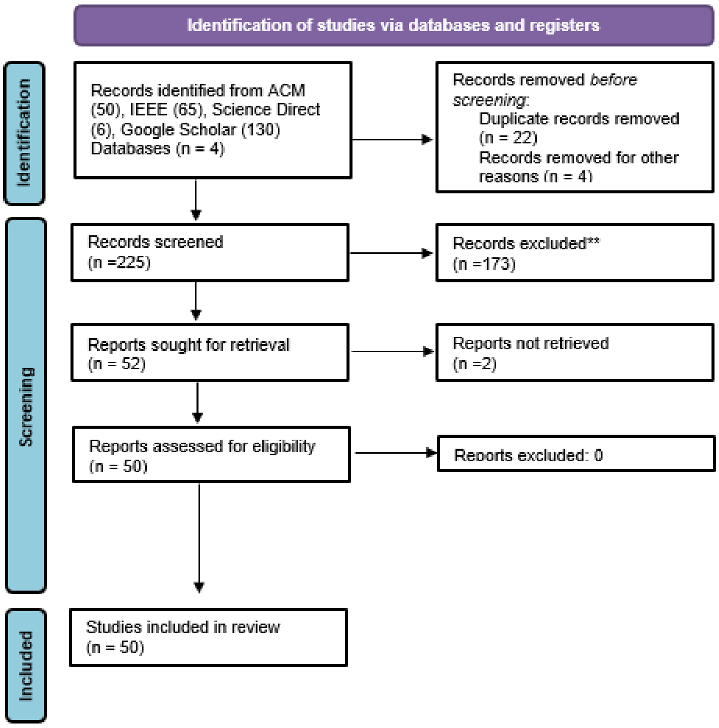

Figure 1. PRISMA flowchart on article selection

3.2. Proposed methodology for artificial intelligence-assisted lung sound analysis

The methodology for AI-assisted lung sound analysis consists of several stages, as depicted in Figure 2. The goal is to develop an effective noise cancellation method for improving lung sound auscultation. This involves comparing various DL models to determine which ones are most effective for handling noise in lung sound recordings. The stages of the methodology are as follows.

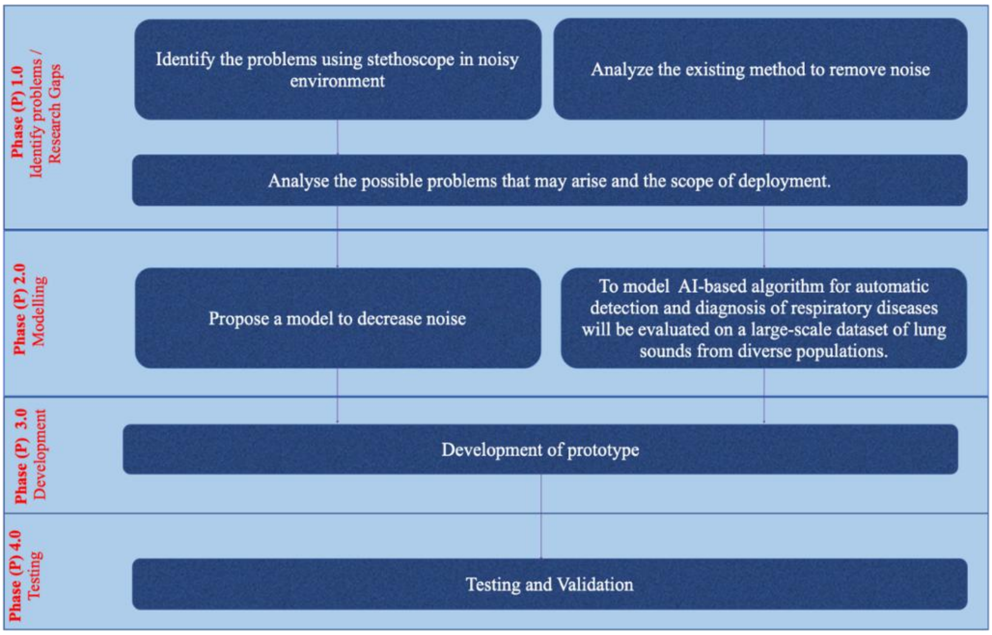

Figure 2. The workflow of the proposed approach

3.2.1. Stage A: identify problems

Stage A identifies the primary challenges involved in using a stethoscope in noisy environments. Despite its versatility as a diagnostic tool, the stethoscope is susceptible to interference from ambient noise, vibrations, physician noise, and patient touch noise. In high-noise environments like emergency rooms (with noise levels up to 70 dB) and aeromedical transport (over 75 dB), these noises make it difficult to detect lung sound anomalies accurately. While certain noise sources can be minimized through training or environmental management, others are persistent and unavoidable. Traditional noise cancellation methods, such as spectral filtering and active noise cancellation (ANC), have been explored but often fall short in completely removing unwanted noise in dynamic real-world conditions [30].

3.2.2. Stage B: deep learning models

To address the issues identified in stage A, DL models are developed to automatically classify and analyze lung sounds. The focus is on training models that can effectively cancel noise while identifying and classifying abnormal respiratory sounds. DL provides a powerful solution for respiratory sound analysis due to its ability to learn complex patterns from large datasets [31]. These models offer more precision and efficiency compared to traditional methods, eliminating the subjectivity of human auscultation. AI models such as CNNs, LSTM networks, and recurrent neural networks (RNNs) are employed, as they are capable of handling the complexity and noise inherent in lung sound data [24]. The multi-layered architecture of DL mimics the human brain's ability to learn from vast amounts of data, enabling more accurate identification and classification of lung sounds. The models used are designed to enhance precision and minimize noise interference, ultimately aiding in early detection of respiratory conditions [32].

3.2.3. Stage C: data science on respiratory sounds

Stage C focuses on the use of data science techniques to enhance the analysis of respiratory sounds. Differentiating between normal breathing sounds and adventitious sounds is essential for accurate lung evaluation. Various spectral analysis techniques, such as Fourier transform, fast Fourier transform (FFT), autoregressive models, and MFCC, are employed to extract meaningful features from lung sound signals. Additionally, techniques like multiscale entropy and EMD are used to analyze the non-stationary nature of lung sound data [33].

To improve model generalization, data augmentation techniques are utilized. These include time stretching, frequency modulation, and adding background noise to enhance the training dataset. By simulating diverse real-world conditions, these augmentation techniques improve the model's robustness and prevent overfitting. The use of spectrogram inversion, frequency masking, and time-shifting further diversifies the dataset, contributing to better model performance in noisy environments [34].

3.2.4. Stage D: testing and validation

The final stage involves testing and validating the performance of the AI-assisted lung sound analysis algorithm. The dataset is split into training, validation, and testing sets to ensure robust evaluation. The model's performance is assessed using metrics such as accuracy, precision, recall, and F1-score. Additionally, validation experiments are conducted across different datasets and environmental conditions to ensure the model's generalizability [35], [36]. These experiments confirm the algorithm’s ability to perform consistently, even in noisy settings. Furthermore, to ensure the reliability and robustness of the model, the testing procedure incorporates additional validation through cross-site validation (using data from different healthcare facilities) and cross-device validation (to ensure compatibility across various recording devices, such as traditional stethoscopes and smartphones).

4. DISCUSSION

Over the past few decades, significant advancements in AI algorithms have greatly enhanced their potential for various applications, including the analysis of respiratory sounds. AI models, particularly those based on supervised learning, have shown significant promise in identifying and classifying lung sounds. Supervised learning approaches, such as SVM [37] and backpropagation in neural networks, rely on labelled input-output data pairs to train the model [38]. These algorithms effectively capture the relationship between features of respiratory sounds and their corresponding labels, enabling accurate predictions on new, unseen data. However, while these models perform well under controlled conditions, they are often susceptible to overfitting, particularly when trained on small or imbalanced datasets [39]. Overfitting limits the generalizability of models and reduces their performance when applied to diverse clinical settings. To mitigate this issue, cross-validation, regularization techniques, and data augmentation should be employed to improve the model’s robustness and prevent it from memorizing rather than generalizing from the data Conversely, unsupervised learning methods, such as self-organizing maps (SOMs), do not rely on labelled data, which makes them useful for detecting patterns and anomalies in unlabeled respiratory sound data [39]. However, unsupervised models typically require further labelled data or manual intervention to improve classification accuracy.

Although hybrid systems that combine supervised and unsupervised learning approaches show promise, these models introduce increased complexity, which can make them difficult to train and deploy in real-time clinical environments. As such, more careful evaluation is needed to determine whether hybrid systems truly offer substantial improvements in performance or if they are merely adding unnecessary complexity [40]. A significant challenge to the adoption of AI-assisted lung sound analysis in clinical practice is the lack of explainability of many DL models. AI models, especially DL architectures like CNNs, are often criticized as "black boxes" because they do not offer clear insights into how decisions are made. This lack of transparency undermines clinician trust, which is essential for the adoption of AI systems in healthcare [41]. For AI models to be successfully integrated into clinical workflows, it is critical to develop explainable artificial intelligence (XAI) techniques [42] that can provide clinicians with interpretable insights into how a model arrived at a particular diagnosis. Techniques such as saliency maps, layer-wise relevance propagation (LRP), and attention mechanisms can help demystify the decision-making process, thereby increasing confidence in AI tools [43].

Another significant barrier is the computational cost associated with training and deploying AI models, particularly DL models. Models like CNNs require substantial computational power and long training times, which can be prohibitive in resource-limited settings such as rural hospitals or emergency departments. To address this, AI systems should be optimized for efficiency, with lightweight models that can run on devices like smartphones or low-cost stethoscopes. Techniques such as model pruning and quantization can help reduce the computational cost, making AI models more accessible in resource-constrained environments [44]. Additionally, cloud-based solutions could be explored to offload heavy computational tasks while keeping the devices themselves lightweight. The issue of data privacy is also a significant concern when using AI models in healthcare. Respiratory sound data is inherently sensitive, and its use in AI systems must adhere to stringent ethical guidelines and regulatory frameworks like General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) [45]. To ensure patient privacy, researchers must explore privacy-preserving technologies such as federated learning, where data is kept on local devices and only model updates are shared, thus ensuring that patient data remains secure while still benefiting from collaborative learning. While AI-powered systems for lung sound analysis have demonstrated superior performance over traditional methods, particularly in noisy environments like emergency rooms and ambulances, they are not yet widely implemented in clinical practice [46]. The challenges of clinical deployment including device compatibility, integration with existing hospital infrastructure, and the training of healthcare providers must be addressed. AI models must be compatible with various stethoscopes and integrated into electronic health record (EHR) systems, facilitating seamless use in everyday clinical workflows. Furthermore, as AI models continue to improve, clinical validation through real-world trials will be essential to assess their safety, efficacy, and overall impact on patient care [47]–[49].

In conclusion, while AI-assisted lung sound analysis holds great promise for improving respiratory disease diagnosis, several challenges remain. Overfitting, the need for external validation datasets, explainability, computational costs, data privacy, and clinical deployment barriers all need to be addressed. Future research should focus on developing noise-robust datasets, improving cross-site and cross-device validation, integrating multi-modal data, and enhancing model explainability. These advancements, coupled with improved integration into clinical workflows, will be essential for realizing the full potential of AI in transforming respiratory sound analysis and improving healthcare outcomes.

5. LIMITATIONS AND FUTURE WORK

This research has limitations due to the use of various datasets, with some classes having a limited number of records, while others have a larger number of samples. The primary constraint is the limited number of samples per class, and further work is required to create more balanced datasets for all diseases. Additionally, the impact of medications on patient symptoms was not explored in this study, although it is recognized as a potential influencing factor.

This systematic literature review (SLR) addresses several gaps in the current research, including issues related to unbalanced data, the lack of publicly available voice data, and the absence of longitudinal studies. However, there are still several areas that require further investigation. The future research directions to address these gaps are outlined as follows:

i) Develop standardized, noise-robust lung sound datasets: there is a need to create and standardize lung sound datasets that are robust to environmental noise. These datasets should incorporate real-world noise from diverse clinical and emergency settings to improve the generalization of ML models.
ii) Implement cross-device validation: future work should focus on validating models across different devices, such as electronic stethoscopes and smartphone microphones, to ensure that they generalize well regardless of the recording equipment used.
iii) Integrate multi-modal fusion (audio, imaging, and metadata): combining audio features with imaging data (e.g., chest X-rays or CT scans) and metadata (e.g., patient demographics and clinical history) could improve the accuracy and robustness of lung sound analysis models, offering a more holistic approach to diagnosis.
iv) Discuss data ethics and feasibility for real-time inference in resource-limited settings: as AI-assisted lung sound analysis moves toward clinical deployment, it is essential to discuss the ethical implications, including patient consent and data privacy. Additionally, future work should explore the feasibility of real-time inference in resource-limited environments where computational resources may be constrained.

By addressing these directions, future research can significantly improve the reliability and applicability of AI-powered lung sound analysis, particularly in real-world, noisy environments like emergency rooms and ambulances.

6. CONCLUSION

This SLR highlights the significant advancements in AI-assisted lung sound analysis, particularly in noisy environments, for early detection of acute respiratory illnesses. Traditional lung sound auscultation has limitations such as subjectivity and interference from ambient noise, which can affect diagnostic accuracy. AI-powered methods, such as DL models like CNN, CNN + LSTM, and hybrid models, have demonstrated improvements in accuracy and reliability. These models can process large datasets efficiently, offering a promising solution for more precise and timely diagnoses in clinical settings. However, challenges such as data quality, class imbalance, and the need for larger, more diverse datasets remain. Future research should focus on refining noise cancellation techniques, improving feature extraction methods, and addressing issues related to dataset variability. The integration of AI systems into clinical workflows requires further validation, particularly in real-world, and high-noise environments like emergency rooms and ambulances. DL methods have shown promise in detecting lung diseases, but further research is required to substantiate these findings and gain wider acceptance in the medical community.

FUNDING INFORMATION

This research was funded from Telekom Research and Development Sdn Bhd: RDTC/241124.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Reshma Sreejith	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
R. Kanesaraj Ramasamy	✓			✓				✓		✓	✓	✓	✓	✓
Wan-Noorshahida Mohd-Isa	✓			✓						✓		✓	✓	
Junaidi Abdullah	✓			✓						✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] R. S. Alice, L. Wendling, and K. Santosh, "2D respiratory sound analysis to detect lung abnormalities," in *Recent Trends in Image Processing and Pattern Recognition*, Cham, Switzerland: Springer, 2023, pp. 46–58, doi: 10.1007/978-3-031-23599-3_5.

[2] A. Gopi and T. Sajini, "Machine hearing a cognitive service for aiding clinical diagnosis," in *Artificial Intelligence and Speech Technology*, Cham, Switzerland: Springer, 2022, pp. 288–304, doi: 10.1007/978-3-030-95711-7_26.

[3] S. Sreejyothi, A. Renjini, V. Raj, M. N. S. Swapna, and S. I. Sankararaman, "Unwrapping the phase portrait features of adventitious crackle for auscultation and classification: a machine learning approach," *Journal of Biological Physics*, vol. 47, no. 2, pp. 103–115, 2021, doi: 10.1007/s10867-021-09567-8.

[4] R. Liu, S. Wang, F. Tian, and L. Yi, "SIR-3DCNN: a framework of multivariate time series classification for lung cancer detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, 2025, doi: 10.1109/TIM.2025.3563000.

[5] R. Hazra and S. Majhi, "Detecting respiratory diseases from recorded lung sounds by 2D CNN," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Oct. 2020, pp. 1–6, doi: 10.1109/ICCCS49678.2020.9277101.

[6] Z. Tariq, S. K. Shah, and Y. Lee, "Multimodal lung disease classification using deep convolutional neural network," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2020, pp. 2530–2537, doi: 10.1109/BIBM49941.2020.9313208.

[7] M. Fraiwan, L. Fraiwan, M. Alkhodari, and O. Hassanin, "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 10, pp. 4759–4771, 2022, doi: 10.1007/s12652-021-03184-y.

[8] A. M. Alqudah, S. Qazan, and Y. M. Obeidat, "Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds," *Soft Computing*, vol. 26, no. 24, pp. 13405–13429, 2022, doi: 10.1007/s00500-022-07499-6.

[9] H. Chanane and M. Bahoura, "Convolutional neural network-based model for lung sounds classification," in *Midwest Symposium on Circuits and Systems*, 2021, pp. 555–558, doi: 10.1109/MWSCAS47672.2021.9531887.

[10] M. F. Pouyani, M. Vali, and M. A. Ghasemi, "Lung sound signal denoising using discrete wavelet transform and artificial neural network," *Biomedical Signal Processing and Control*, vol. 72, 2022, doi: 10.1016/j.bspc.2021.103329.

[11] N. S. Haider and A. K. Behera, "Respiratory sound denoising using sparsity-assisted signal smoothing algorithm," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 481–493, 2022, doi: 10.1016/j.bbe.2022.03.005.

[12] M. Tamal, Y. AlMania, A. J. AlHubaishi, N. Alotaibi, and S. Tulleimat, "Classification of lung sounds using deep neural network (DNN) for accurate and robust diagnosis of lung diseases," in *IET Conference Proceedings*, 2022, pp. 264–271, doi: 10.1049/icp.2022.2473.

[13] K. H. Abdullah and M. B. Er, "Lung sound signal classification by using cosine similarity-based multilevel discrete wavelet transform decomposition with CNN-LSTM hybrid model," in *2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST)*, 2022, pp. 1-4, doi: 10.1109/AIST55798.2022.10065345.

[14] R. P. Vasava and H. A. Joshiara, "Different respiratory lung sounds prediction using deep learning," in *2 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2023, pp. 1626–1630, doi: 10.1109/ICESC57686.2023.10193040.

[15] Rishabh and D. Kumar, "Multi spectral feature extraction to improve lung sound classification using CNN," in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, 2023, pp. 186–191, doi: 10.1109/SPIN57001.2023.10116295.

[16] A. Bapa, O. Bandgar, A. Ekapure, and J. Sisodia, "Respiratory disorder classification based on lung auscultation using MFCC, Mel spectrogram and Chroma STFT," in *2023 International Conference on Artificial Intelligence and Applications, ICAIA 2023 and Alliance Technology Conference, ATCON-1 2023*, 2023, doi: 10.1109/ICAIA57370.2023.10169299.

[17] A. I. Taloba and R. T. Matoog, “Detecting respiratory diseases using machine learning-based pattern recognition on spirometry data,” *Alexandria Engineering Journal*, vol. 113, Feb. 2025, pp. 44-59, doi: 10.1016/j.aej.2024.11.009.

[18] A. H. Sfayyih, N. Sulaiman, and A. H. Sabry, “Non-invasive diagnosis of lung diseases via multimodal feature extraction from breathing audio and chest dynamics,” *Computers in Biology and Medicine*, vol. 191, 2025, doi: 10.1016/j.compbiomed.2025.110182.

[19] Abdullah, Z. Fatima, J. Abdullah, J. L. O. Rodríguez, and G. Sidorov, “A multimodal AI framework for automated multiclass lung disease diagnosis from respiratory sounds with simulated biomarker fusion and personalized medication recommendation,” *International Journal of Molecular Sciences*, vol. 26, no. 15, 2025, doi: 10.3390/ijms26157135.

[20] M. Miyamoto *et al.*, “Lung sound analysis for predicting recurrent wheezing in preschool children,” *Journal of Allergy and Clinical Immunology: Global*, vol. 3, no. 1, 2024, doi: 10.1016/j.jacig.2023.100199.

[21] Y. S. Wu, C. H. Liao, and S. M. Yuan, “Automatic auscultation classification of abnormal lung sounds in critical patients through deep learning models,” in *2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII),* 2020, pp. 9–11, doi: 10.1109/ICKII50300.2020.9318880.

[22] F. Ahmad, “Early lung cancer diagnosis using a hybrid CNN–BiLSTM deep learning model: a step toward precision medicine,” *Research Square*, 2025, doi: 10.21203/rs.3.rs-7954420/v1.

[23] T. Bikku *et al.*, “Deep learning-driven early diagnosis of respiratory diseases using CNN-RNN fusion on lung sound data,” *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-28832-7.

[24] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, “A neural network-based method for respiratory sound analysis and lung disease detection,” *Applied Sciences*, vol. 12, no. 8, 2022, doi: 10.3390/app12083877.

[25] Y. Kim, Y. K. Hyon, S. Lee, S. D. Woo, T. Ha, and C. Chung, “The coming era of a new auscultation system for analyzing respiratory sounds,” *BMC Pulmonary Medicine*, vol. 22, no. 1, 2022, doi: 10.1186/s12890-022-01896-1.

[26] W. Wang, W. Shubo, Q. Dimei, F. Yu, and Z. Y. Kang, “Heart-lung sound separation by nonnegative matrix factorization and deep learning,” *SSRN Electronic Journal*, vol. 79, 2022, doi: 10.2139/ssrn.4017034.

[27] H. Melbye, J. Ravn, M. Pabiszczak, L. A. Bongo, and J. C. A. Solis, “Validity of a deep learning algorithm for detecting wheezes and crackles from lung sound recordings in adults,” *medRXiv*, vol. 75, pp. 1–12, 2022, doi: 10.1101/2022.11.18.22282442.

[28] B. A. Tessema, H. Nemomssa, and G. L. Simegn, “Acquisition and classification of lung sounds for improving the efficacy of auscultation diagnosis of pulmonary diseases,” *Medical Devices: Evidence and Research*, vol. 15, pp. 89–102, 2022, doi: 10.2147/MDER.S362407.

[29] A. Puneet, P. Shankar, S. R. R. R. Koluguri, and A. Srivastava, “Edge-enabled portable classifier for lung sounds using convolutional neural networks,” in *2025 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2026, pp. 21–25, doi: 10.1109/biocas67066.2025.00016.

[30] S. Jayalakshmy, B. L. Priya, and N. Kavya, “CNN based categorization of respiratory sounds using spectral descriptors,” in *Proceedings of the 2020 IEEE International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2020, doi: 10.1109/C2I451079.2020.9368933.

[31] S. Dongre, O. Dudooknale, K. Jacob, and G. Dharmale, “Classification of lung diseases from respiratory acoustic sounds,” in *2025 9th International Conference on Computing, Communication, Control and Automation (ICCCBEA),* Pune, India, 2025, pp. 1–7, doi: 10.1109/iccubea65967.2025.11283969.

[32] M. M. Ramyasri, M. Yoga, C. Harish, M. M. Rizwan, and K. Monisha, “Utilizing neural networks to analyze respiratory sounds for the detection of lung diseases,” in *ICACC - International Conference on Advances in Computing and Communications*, 2024, doi: 10.1109/ICACC63692.2024.10845461.

[33] M. T. G. Ordás, J. A. B. Andrades, I. G. Rodríguez, C. Benavides, and H. A. Moretón, “Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data,” *Sensors*, vol. 20, no. 4, 2020, doi: 10.3390/s20041214.

[34] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara, and I. D. L. T. Díez, “Lung sound classification with multi-feature integration utilizing lightweight CNN model,” *IEEE Access*, vol. 12, pp. 21262–21276, 2024, doi: 10.1109/ACCESS.2024.3361943.

[35] G. Harman, “Lung sounds ventilation cycle segmentation and classify healthy, asthma and COPD,” *Fortune Journal of Health Sciences*, vol. 07, no. 01, 2023, doi: 10.26502/fjhs.160.

[36] J. S. Park, K. Kim, J. H. Kim, Y. J. Choi, K. Kim, and D. I. Suh, “A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model,” *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-27399-5.

[37] J. Li *et al.*, “Explainable CNN with fuzzy tree regularization for respiratory sound analysis,” *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1516–1528, 2022, doi: 10.1109/TFUZZ.2022.3144448.

[38] F. Meng, Y. Shi, N. Wang, M. Cai, and Z. Luo, “Detection of respiratory sounds based on wavelet coefficients and machine learning,” *IEEE Access*, vol. 8, pp. 155710–155720, 2020, doi: 10.1109/ACCESS.2020.3016748.

[39] P. Shrivastava and N. Tripathi, “Comparison of different classification techniques for the detection of speech affected due to respiratory disorders,” *Journal of Physics: Conference Series*, vol. 2273, no. 1, 2022, doi: 10.1088/1742-6596/2273/1/012013.

[40] F. Tong, L. Liu, X. Xie, Q. Hong, and L. Li, “Respiratory sound classification: from fluid-solid coupling analysis to feature-band attention,” *IEEE Access*, vol. 10, pp. 22018–22031, 2022, doi: 10.1109/ACCESS.2022.3151789.

[41] P. Zhang, A. Swaminathan, and A. A. Uddin, “Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks,” *Frontiers in Medicine*, vol. 10, 2023, doi: 10.3389/fmed.2023.1269784.

[42] K. C. Burçak and H. Uğuz, “A new hybrid breast cancer diagnosis model using deep learning model and relieff,” *Traitement du Signal*, vol. 39, no. 2, pp. 521–529, 2022, doi: 10.18280/ts.390214.

[43] S. A. Shehab, K. K. Mohammed, A. Darwish, and A. E. Hassanien, “Deep learning and feature fusion-based lung sound recognition model to diagnoses the respiratory diseases,” *Soft Computing*, vol. 28, no. 19, pp. 11667–11683, 2024, doi: 10.1007/s00500-024-09866-x.

[44] S. Ghrabli, M. Elgendi, and C. Menon, “Challenges and opportunities of deep learning for cough-based COVID-19 diagnosis: a scoping review,” *Diagnostics*, vol. 12, no. 9, 2022, doi: 10.3390/diagnostics12092142.

[45] S. Gupta, M. Agrawal, and D. Deepak, “Classification of auscultation sounds into objective spirometry findings using MVMD and 3D CNN,” in *2022 National Conference on Communications (NCC)*, 2022, pp. 42–47, doi: 10.1109/NCC55593.2022.9806737.

[46] Z. Zhang, J. Han, K. Qian, C. Janott, Y. Guo, and B. Schuller, “Snore-GANs: improving automatic snore sound classification with synthesized data,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 300–310, 2020, doi: 10.1109/JBHI.2019.2907286.

[47] N. Asatani, T. Kamiya, S. Mabu, and S. Kido, “Classification of respiratory sounds using improved convolutional recurrent neural network,” *Computers and Electrical Engineering*, vol. 94, 2021, doi: 10.1016/j.compeleceng.2021.107367.

[48] U. Hassan, A. Singhal, and G. Gupta, "Neural network based AI model for lung health assessment," *Scientific Reports*, vol. 15, no. 1, Jul. 2025, doi: 10.1038/s41598-025-09524-8.
[49] Y. Jeong, J. Kim, D. Kim, J. Kim, and K. Lee, "Methods for improving deep learning-based cardiac auscultation accuracy: data augmentation and data generalization," *Applied Sciences*, vol. 11, no. 10, 2021, doi: 10.3390/app11104544.

BIOGRAPHIES OF AUTHORS

Reshma Sreejith is a research scholar in the Faculty of Computing and Informatics at Multimedia University. She is a dedicated Ph.D. student specializing in AI. She is deeply committed to advancing the frontiers of AI through research that integrates innovative algorithms, ML techniques, and ethical considerations. Her research interests include AI, ML, and DL. She can be contacted at email: 1231400108@student.mmu.edu.my.

Dr. R. Kanesaraj Ramasamy is a lecturer in the Faculty of Computing and Informatics, Multimedia University Cyberjaya, Malaysia. He obtained his Ph.D. from Multimedia University with a thesis titled: Adaptive and dynamic web service composition for cloud-based mobile application. He was also awarded professional technologist (Ts) by the Malaysian Board of Technology (MBOT) and is a Microsoft Office 2016 specialist, Microsoft certified professional, and specialist in both web development and database technology. His research is based on service-oriented computing and the internet of things (IoT). He has also been published in several conferences and journals. He has nine years of experience in the software industry in both the development and implementation phases. He is also certified by the International Software Testing Qualification Board (ISTQB), which allows him to practice as a professional software tester. He was also involved in a research project funded by JICA and SASTREPS to develop an early warning system for floods and landslides in Malaysia. He can be contacted at email: r.kanesaraj@mmu.edu.my.

Dr. Wan-Noorshahida Mohd-Isa is affiliated to Faculty of Computing and Informatics (FCI), Multimedia University. She is currently providing services as lecturer. She has authored and co-authored multiple peer-reviewed scientific papers and presented works at many national and international conferences. She contributions have acclaimed recognition from honorable subject experts around the world. She is actively associated with different societies and academies. She academic career is decorated with several reputed awards and funding. She researches interests include computing and informatics. She can be contacted at email: wan.noorshahida.isa@mmu.edu.my.

Dr. Junaidi Abdullah is an associate professor in the Faculty of Computing and Informatics, Multimedia University (MMU), Malaysia. He is currently the dean of the faculty and the chairperson of the Assistive Technology Special Interest Group. He obtained a bachelor degree in engineering (B.Eng., first class honors) from the University of Bristol, United Kingdom, and a Ph.D. in Computer Science in the area of computer vision and augmented reality from the University of Southampton, United Kingdom, in 2005. His research includes augmented reality, image and video processing, and computer vision. He has published more than 50 internationally multi-disciplinary refereed conference papers, journal articles, and books. He can be contacted at email: junaidi.abdullah@mmu.edu.my.