

Implementation of a Hybrid CNN-BiLSTM-Attention Model for Detecting Respiratory Sound Abnormalities

Fikky Aprico^{1)*}, Kristiawan Nugroho²⁾

^{1,2)}Program Studi Teknologi Informasi, Universitas Stikubank, Semarang

¹⁾fikkyaprico0013@mhs.unisbank.ac.id, kristiawan@edu.unisbank.ac.id²⁾

Submitted : Jul 23, 2026 | **Accepted :** Aug 12, 2026 | **Published :** Oct 5, 2026

Abstract: Early detection of respiratory disorders generally relies on digital stethoscope devices which are expensive and have limited availability. Although the use of smartphones offers a highly practical alternative for independent screening purposes in the community, respiratory sounds recorded directly through built-in microphones are highly susceptible to environmental noise and acoustic characteristic differences (domain shift). Consequently, artificial intelligence (AI) models trained exclusively on public medical datasets often fail to adapt in real-world clinical scenarios, yielding an initial accuracy rate of only 57.78%. To address this critical limitation, this study proposes the application of a Transfer Learning technique with a Full Fine-Tuning approach on a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) architecture. Furthermore, an Attention Mechanism is specifically integrated so the model can focus its computational weight on very short-duration spectral anomalies that are often masked by background noise. Following the pre-training phase, the model was comprehensively calibrated using 45 primary respiratory sound samples from smartphones. The testing results demonstrated a substantial performance leap, with the overall accuracy rapidly increasing to 93.33%. The proposed system successfully identified all abnormal pulmonary cases without a single miss, achieving 100% Sensitivity. This finding Indicates that the integration of CNN-BiLSTM and Attention Mechanism is promising in handling acoustic distortion. Thus, the model is highly viable to be implemented as an early clinical decision support system with a showing high sensitivity on the evaluated localized cohort predictions.

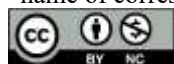
Keywords: Deep Learning, Transfer Learning, Lung Sound, Smartphone, Attention Mechanism.

INTRODUCTION

Respiratory diseases represent one of the leading causes of global mortality, making the early detection of pulmonary function abnormalities a highly crucial clinical necessity (Malviya et al., 2024; Mufarah et al., 2023). Recent systematic reviews demonstrate that artificial intelligence combined with acoustic audio analysis offers a viable non-invasive approach for early respiratory disease detection (Kapetanidis et al., 2024). Current automated auscultation practices still heavily rely on expensive digital stethoscopes (Brunese et al., 2022). This condition makes it difficult for the general public to conduct widespread self-screening. As an alternative, people can utilize smartphone microphones to record respiratory sounds. However, smartphone devices produce environmental noise distortion and acoustic characteristic differences (domain shift) (Huang et al., 2023; Saky et al., 2026). Furthermore, recent advancements in acoustic classification strongly emphasize the critical need for robust domain adaptation strategies when dealing with multi-source hardware discrepancies (Islam & Nirjon, 2021). Artificial intelligence models trained on public medical datasets often experience a drastic drop in accuracy when processing smartphone recording data in the field.

Previous researchers have developed various algorithms to classify respiratory audio signals. (Petmezas et al., 2022) combined Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures to extract spectral features and model the time sequence of lung sounds. Nevertheless, anomalous sounds such as crackles or wheezes often appear in just milliseconds. (Salor-b, 2022) added an Attention Mechanism to LSTM

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

models so the system could focus its mathematical calculations on the most critical acoustic features. Regarding the hardware discrepancy issue, (Wang et al., 2024) applied a Transfer Learning approach to align the data distribution between the source dataset and the target dataset.

Referring to these studies, this research builds an early detection framework using a hybrid architecture of CNN and Bidirectional LSTM (Bi-LSTM) equipped with an Attention Mechanism. This model undergoes a Transfer Learning phase through the Full Fine-Tuning method. The system performs initial pre-training on public datasets and then conducts full calibration on 45 primary sound samples from smartphone microphones. This study aims to evaluate the model's potential as a preliminary home-screening tool that demonstrates enhanced computational adaptability against smartphone device noise. This system is designed to maintain high sensitivity on the evaluated localized cohort, offering a promising step toward accessible mobile respiratory monitoring.

LITERATURE REVIEW

Previous researchers have explored various artificial intelligence methods to analyze lung sounds. Comprehensive systematic evaluations and meta-analyses further validate the diagnostic efficacy of deep learning models across clinical respiratory sound benchmarks (Park et al., 2025). (Petmezas et al., 2022) designed a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture to classify respiratory sound abnormalities. The CNN architecture extracts spatial features comprehensively from spectrogram representations. Meanwhile, the LSTM component models the time sequence of each respiratory cycle. This hybrid model proved superior to single-algorithm implementations in detecting abnormal sound patterns within clinical medical records. Additionally, CNN-RNN fusion architectures have been effectively employed to capture both spatial spectrogram features and temporal sequence dynamics in respiratory sound data (Bikku et al., 2025).

Abnormal respiratory sounds often possess a very short duration and are easily masked by background noise. (Salor-b, 2022) applied an Attention Mechanism to a memory-based neural network model (LSTM-based) to handle these short-duration characteristics. This mechanism intelligently assigns greater mathematical computational weight to audio segments containing acoustic anomalies. The use of the Attention Mechanism helps the AI model suppress environmental noise interference and significantly improves the detection sensitivity rate.

The implementation of artificial intelligence models in medical environments is often hindered by recording device discrepancies between laboratory spaces and real-world conditions. (Lal, 2023) used Transfer Learning techniques to overcome the data distribution discrepancy (domain shift) problem in lung sound recognition systems. This method transfers baseline knowledge parameters from large-scale source datasets to small-scale target datasets. The implementation of Transfer Learning allows AI models to adapt rapidly to the frequency characteristics of new hardware without experiencing overfitting conditions.

These various studies have not comprehensively tested the combination of all three algorithmic elements (CNN-BiLSTM, Attention, and Transfer Learning) on primary data captured by smartphone microphones. This research fills that research gap by assembling the three methods simultaneously. The CNN-BiLSTM architecture acts as the primary feature extractor. The Attention Mechanism functions as a clinical target sound selection filter. Finally, Transfer Learning with the Full Fine-Tuning method calibrates the entire system network to accurately and precisely recognize 45 primary data samples from smart cellular phones.

RESEARCH METHOD

This research develops a lung sound abnormality detection system using a hybrid artificial intelligence architecture. The researcher specifically designed this methodology to overcome sound distortion issues in smartphone devices. The system processes 45 primary data samples from field recordings using standardized audio preprocessing methods. The system trims the sound waves into a fixed duration of six seconds to standardize the data dimensions (Saldanha et al., 2022). Furthermore, the system converts the raw audio signals into a two-dimensional visual representation called a Mel-Spectrogram, which effectively captures spatial-frequency characteristics (Hana et al., 2026; Wang et al., 2024; Yabani et al., 2026). This visual representation serves as the primary input data for the computational model architecture.

To evaluate the system's robustness in a true mobile health (mHealth) scenario, the primary dataset collection was conducted in real-world residential environments (uncontrolled home settings) rather than isolated clinical laboratories. The 22 unique subjects consisted of adult patients aged between 18 and 60 years old. During the acquisition process, subjects were instructed to position their smartphone approximately 5 cm away from their upper chest and perform exactly two full respiratory cycles (two complete inhale-exhale sequences), naturally yielding a fixed 6-second recording. The audio files were captured using various consumer smartphone built-in microphones and subsequently transmitted via a widely used mobile messaging application (WhatsApp). Initial screening labels were established via a self-reported questionnaire regarding the subject's current physiological

state; subjects reporting active symptoms such as cough, common cold, or other respiratory distress were provisionally assigned to the 'Abnormal' class, while asymptomatic subjects were assigned to the 'Normal' class. However, to account for subclinical acoustic variations—such as natural age-related respiratory tract changes in older subjects (e.g., individuals aged 50+ who self-reported as healthy but exhibited altered breathing patterns)—all audio samples underwent mandatory cross-verification by independent expert medical practitioners to establish the absolute ground truth annotations, ensuring strict diagnostic reliability while keeping all personal identifiers anonymized. Ethical Considerations: As this study focuses on the computational development of a diagnostic support algorithm involving strictly non-invasive audio recordings, all participants provided explicit informed consent prior to data collection. Participation was entirely voluntary, and comprehensive data anonymization protocols were enforced to ensure that no personally identifiable information (PII) was stored, processed, or published, guaranteeing full compliance with subject privacy protection.



Figure 1. The proposed Research Methodology Flowchart For Respiratory Sound Detection

To ensure methodological rigor and prevent any risk of data leakage or subject leakage, the dataset partitioning was evaluated using a Subject-Wise 5-Fold Cross-Validation strategy. The 45 primary samples collected in this study represent exactly 45 unique audio recordings (fixed at a 6-second duration per sample) gathered from 22 distinct individual subjects, where subjects contributed two recordings each, and one subject provided three recordings. To eliminate subject leakage entirely during the validation phase, the dataset was strictly partitioned using a Subject-Wise Splitting strategy. This approach guarantees that multiple recordings belonging to the exact same individual were always kept together within the same fold, ensuring the model was strictly evaluated on unseen subjects during each iteration. Furthermore, to mitigate overfitting risks during the full fine-tuning phase on this primary dataset, spatial data augmentation (pitch shifting) and a dropout regularization rate of 0.3 were integrated into the network layers.

The researcher builds the primary model by sequentially combining three algorithmic components. First, the Convolutional Neural Network (CNN) extracts spatial features and static frequency patterns from the Mel-Spectrogram images. Second, the Bidirectional Long Short-Term Memory (Bi-LSTM) processes these spatial features to map the respiratory time sequence dependencies forward and backward. Finally, the Attention Mechanism assigns the highest computational weight to the crucial milliseconds detected containing sound anomalies. The combination of these three components ensures the system can ignore background noise and focus its evaluation on disease indicators.

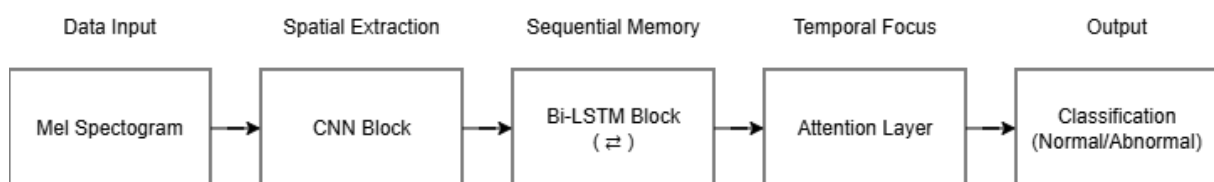


Figure 2. The detailed block diagram of the hybrid CNN-BiLSTM-Attention network architecture

In the training phase, the researcher applies a Transfer Learning strategy with a Full Fine-Tuning technique (Zhang et al., 2022). The system conducts initial pre-training using the public Respiratory Sound Database (ICBHI 2017) accessed via Kaggle (<https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database>). This baseline dataset consists of 920 annotated recordings from 126 patients captured exclusively using conventional clinical stethoscopes. This phase allows the network to recognize the foundational acoustic basis of lung sounds. After completing the pre-training, the system unlocks all freezing constraints on the CNN and Bi-LSTM algorithm layers to perform a thorough calibration using the primary smartphone field recordings.

Table 1. Proposed Hybrid CNN-BiLSTM-Attention Hyperparameter Configuration

Architectural Layer / Block	Component Specification	Configuration Value
Spatial Feature Extractor (CNN)	Convolutional Layers	3 Layers (Filters: 32, 64, 128)
	Kernel Size	3 x 3

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

	Activation Function	Rectified Linear Unit (ReLU)
	Pooling Layer	Max Pooling (2 x 2)
Temporal Sequence Mapper	Bi-LSTM Hidden Units	128 Units
	Bi-LSTM Layer	1 Layer
	Regularization Technique	Dropout Rate (0.3)
Focus & Context Aggregator	Attention Mechanism	Bahdanau Attention
	Output Layer & Activation	Dense Layer with Softmax
	Classification Threshold	0.5 (Binary Threshold)
Training Hyperparameters	Pretrained Weights	ICBHI 2017 Dataset (Kaggle)
	Optimization Algorithm	Adam Optimizer
	Learning Rate	0.0001
	Batch Size	Batch Size: 16
	Epochs	40 Epochs
	Total Trainable Parameters	~1.24 Million Parameters

The hyperparameter configurations detailed in Table 1 are carefully selected to balance the computational complexity of the hybrid network and the small scale of the target primary dataset. The implementation of a moderate dropout rate of 0.3 within the temporal sequence mapper block serves as a crucial regularizer to actively penalize network weight exploding, effectively minimizing the risk of model overfitting during the intensive full fine-tuning phase. Furthermore, employing a conservative learning rate of 0.0001 with the Adam optimizer ensures a stable and gradual gradient convergence over the 40 training epochs, allowing the model to smoothly bridge the acoustic discrepancies stemming from the consumer-grade smartphone hardware shift.

To mathematically counteract the risk of overfitting during the 40 epochs of full fine-tuning, several explicit regularization techniques were strictly applied. First, spatial data augmentation via pitch shifting was conducted on the primary audio samples to artificially double the training variance. Second, a Dropout rate of 0.3 was embedded within the Bi-LSTM layers to randomly deactivate network nodes, preventing co-adaptation of features. Third, an Early Stopping mechanism with a patience threshold of 5 epochs was configured based on the validation loss monitoring to halt the training immediately if no generalization improvement occurred. Finally, the integration of Subject-Wise 5-Fold Cross-Validation provides a robust cross-validation structure, ensuring that the low training loss of 0.0467 reflects genuine cross-entropy minimization rather than data memorization.

RESULT AND DISCUSSION

The calibration process through the Full Fine-Tuning method for 40 epochs provided a highly significant performance improvement. The system successfully adjusted the mathematical weights of the model to the acoustic characteristics of the smartphone microphone. Testing on 45 primary data samples produced an overall accuracy rate of 93.33%, correctly classifying 42 recordings. The model demonstrated a high level of stability at the end of the training phase, with the loss value decreasing to reach 0.0467. The researcher presents the detailed classification performance evaluation of the system in Table 2.

Table 2. Classification Report of the CNN-BiLSTM-Attention Model Performance

Class	Precision	Recall (Sensitivity)	F1-Score
Abnormal	0.86	1.00	0.93
Normal	1.00	0.88	0.94
Accuracy			93.33%

This hybrid model demonstrates a vital clinical advantage regarding the sensitivity (recall) value for the abnormal class. Despite utilizing data from smartphone devices that are susceptible to noise, the system successfully identified all abnormal respiratory sound samples (19 out of 19 recordings) without generating a single false negative prediction. Acquiring a recall value of 1.00 serves as a crucial indicator for an early screening system, as failing to detect a sick patient is far more dangerous than misdiagnosing a healthy individual. The researcher displays the visualization of this classification prediction accuracy through the Confusion Matrix in Figure 3.

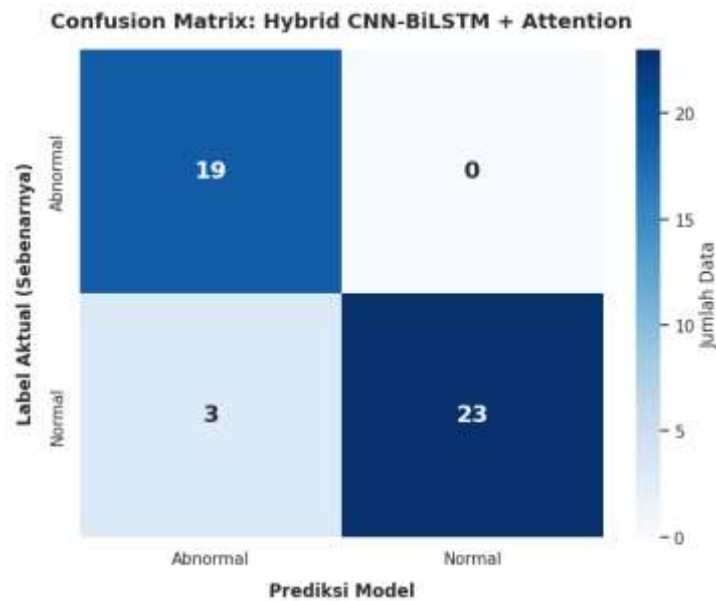


Figure 3. Confusion matrix of the hybrid model testing on smartphone primary data

To prove the significance of the proposed calibration strategy's contribution, the researcher conducted a comparative experiment in Table 3. The researcher tested the exact same hybrid architecture (Hybrid CNN-BiLSTM-Attention) under two different training scenarios to observe the effect of sound domain adaptation.

Table 3. Performance Comparison Based on Training Scenarios

Model Architecture	Training Strategy	Evaluation Dataset	Overall Accuracy
Hybrid CNN-BiLSTM-Attention	Trained exclusively on Kaggle ICBHI public data (Stethoscope)	Smartphone Field Data	57.78%
Hybrid CNN-BiLSTM-Attention	Pre-trained Kaggle + Full Fine-Tuning Primary Data	Smartphone Primary Data (45 Samples)	93.33%

The comparative results in Table 3 mathematically validate the existence of the acoustic domain shift phenomenon and demonstrate the efficacy of the proposed fine-tuning strategy. The baseline hybrid model trained exclusively on conventional stethoscope data (ICBHI) achieved an overall accuracy of only 57.78% (correctly predicting 26 out of 45 samples) when directly evaluated on unseen smartphone recordings without domain adaptation. In contrast, when evaluated using the Subject-Wise 5-Fold Cross-Validation fine-tuning scheme—where testing folds strictly contained unseen individual subjects—the performance rose to 93.33%. Because both evaluation scenarios strictly measured predictions on unseen target domain data, this comparison fairly and rigorously confirms that Transfer Learning via full fine-tuning successfully bridges the acoustic hardware discrepancy between stethoscopes and smartphone microphones.

The prediction errors in the proposed model occurred exclusively within the normal class, where the system misclassified three healthy sound samples as abnormal. This occurred because external environmental noise in the field mirrored the frequency patterns of wheezing. However, the integration of the Attention Mechanism and the Bidirectional LSTM layer generally proved successful in isolating static smartphone noise components and amplifying short-duration anomalous spectral signals, aligning with modern temporal modeling advancements (Im et al., 2023; Suma et al., 2023). This Transfer Learning approach successfully resolved the domain shift issue, thereby enabling the model to generalize field data exceptionally well despite utilizing consumer-grade recording devices.

While a standalone component-by-component ablation study (e.g., isolating CNN-only or CNN-LSTM without Attention) was not explicitly plotted in this study, the structural integration of Bi-LSTM and Attention mechanisms is strongly supported by foundational literature in biomedical signal processing. The Bi-LSTM layer captures bidirectional temporal dependencies across respiratory cycles, while the Attention mechanism assigns higher computational weights to millisecond-level acoustic anomalies masked by noise. Nevertheless, conducting a

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

comprehensive multi-stage ablation study remains an essential direction for future research on larger, multi-center cohorts to further quantify the isolated feature-attraction capacity of each architectural block.

To highlight the methodological contribution of this study, Table 4 presents a direct comparison with previous baseline studies in respiratory sound analysis. While previous architectures heavily rely on clinical stethoscopes in controlled environments, the proposed architecture specifically targets the domain shift challenge toward consumer-grade mobile devices.

Table 4. Methodological Comparison Table

Reference	Architecture	Target Hardware	Target Dataset	Key Contribution & Limitation
Petmezas et al. (2022)	CNN-LSTM	Digital Stethoscope	ICBHI	High spatial-temporal accuracy but limited to controlled hospital devices.
Salor-b (2022)	LSTM+Attention	Digital Stethoscope	ICBHI	Resolves short-duration anomalies but lacks hardware transferability.
Wang et al. (2024)	CNN+Transfer Learning	Electronic Stethoscope	ICBHI+Private	Addresses domain shift but still relies on dedicated medical hardware.
Bikku et al. (2025)	CNN-RNN Fusion	Clinical Sources	ICBHI + Coswara	Proven spatial-temporal feature extraction, but lacks focus on mobile hardware shift.
Proposed Method	CNN-BiLSTM-Attention + Full Fine-Tuning	Consumer Smartphone	ICBHI + Primary Data	Resolves stethoscope-to-smartphone domain shift using an mHealth approach.

Limitations of the study Despite the promising screening potential, this study acknowledges several methodological limitations. First, the small sample size of 45 primary recordings restricts broad demographic generalization, necessitating future validation on large-scale, multi-center cohorts. Second, while Subject-Wise 5-Fold Cross-Validation effectively prevented subject leakage during this phase, a completely independent external test set from different smartphone brands is required to fully map out device biases. Lastly, the term 'robustness' utilized in this study refers strictly to the model's practical adaptability against real-world ambient residential noise (e.g., room echo or minor background sounds), rather than performance stability under mathematically controlled Signal-to-Noise Ratio (SNR) injection experiments.

CONCLUSION

This research successfully designed and implemented a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) architecture integrated with an Attention Mechanism to detect lung sound abnormalities. The application of the Transfer Learning method through the Full Fine-Tuning technique proved successful in resolving the acoustic domain shift problem between clinical stethoscopes and

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

smartphone microphones. The system achieved an overall accuracy rate of 93.33% when tested using 45 primary data samples. This model demonstrates a highly satisfying classification metric on the tested dataset. While external environmental noise occasionally poses minor classification challenges in field environments, the integration of full fine-tuning demonstrates a highly promising potential for consumer-grade smartphones to serve as an accessible, cost-effective early mobile screening support system.

This hybrid model demonstrates a satisfying classification performance, successfully detecting abnormal sound cases on the tested dataset with high sensitivity. The results indicate that consumer-grade devices such as smartphones hold promising potential as an alternative, cost-effective tool for early mobile screening support. However, the prediction errors that occurred in the normal class acknowledge that the system still possesses vulnerabilities to external environmental noise distortion. Therefore, it is highly recommended for future studies to integrate advanced noise-reduction filtering algorithms prior to the feature extraction stage to further stabilize the model's performance in unpredictable field conditions.

REFERENCES

- Bikku, T., V, S. S. K. P. N., Thota, S., Pujari, J. J., Batchu, R. K., Mortezaagha, P., Kumar, M. K., & Anitha, R. (2025). *Deep Learning-Driven Early Diagnosis of Respiratory Diseases using CNN-RNN Fusion on Lung Sound Data*. 1–19.
- Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. (2022). *applied sciences A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection*.
- Hana, E. N., Faisal, M. R., Kartini, D., Mazdadi, M. I., & Wahyu, S. (2026). *Detecting respiratory diseases using spectrogram-based deep features and machine learning algorithms*. 15(2), 1417–1428. <https://doi.org/10.11591/eei.v15i2.10585>
- Huang, D. M., Huang, J., Qiao, K., Zhong, N. S., Lu, H. Z., & Wang, W. J. (2023). Deep learning - based lung sound analysis for intelligent stethoscope. *Military Medical Research*, 1–23. <https://doi.org/10.1186/s40779-023-00479-3>
- Im, S., Kim, T., Min, C., Kang, S., Roh, Y., Kim, C., Kim, M., Kim, S. H., Shim, K., Koh, J., Id, S. H., Id, J. L., Kim, D., & Id, D. K. (2023). *Real-time counting of wheezing events from lung sounds using deep learning algorithms : Implications for disease prediction and early intervention*. 1–18. <https://doi.org/10.1371/journal.pone.0294447>
- Islam, M. T., & Nirjon, S. (2021). *Sound-Adapter: Multi-Source Domain Adaptation for Acoustic Classification Through Domain Discovery*. 1–15. <https://doi.org/10.1145/3412382.3458265>
- Kapetanidis, P., Kalioras, F., Tsakonas, C., & Tzamalidis, P. (2024). *Respiratory Diseases Diagnosis Using Audio Analysis and*. 1–37.
- Lal, K. N. (2023). *A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning*.
- Malviya, A., Dixit, R., Shukla, A., & Kushwaha, N. (2024). A Novel Approach to Detection of COVID-19 and Other Respiratory Diseases Using Autoencoder and LSTM. *SN Computer Science*, 6. <https://doi.org/10.1007/s42979-024-03546-1>
- Mufarah, G., Alshmrani, M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). ORIGINAL ARTICLE A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal*, 64, 923–935. <https://doi.org/10.1016/j.aej.2022.10.053>
- Park, J. S., Park, S., Moon, J. W., Kim, K., & In, D. (2025). *Artificial Intelligence Models for Pediatric Lung Sound Analysis : Systematic Review and Meta-Analysis Corresponding Author : 27*, 1–18. <https://doi.org/10.2196/66491>
- Petmezas, G., Cheimariotis, G., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., & Maglaveras, N. (2022). Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. *IEEE Journal of Biomedical and Health Informatics*, 26(3), 1242–1254.
- Saky, S. M. A. I., Arefin, S., Islam, R., Islam, M. S., & Muthanna, M. (2026). *Explainable Multi-Modal Deep Learning for Recording-Level Classification of Respiratory Audio Signals Under Internal and Domain-Shift Evaluation*. 1–30.
- Saldanha, J., Chakraborty, S., Id, S. P., & Id, K. K. (2022). *Data augmentation using Variational Autoencoders for improvement of respiratory disease classification*. <https://doi.org/10.1371/journal.pone.0266467>
- Salor-b, N. (2022). Respiratory Sound Classification Using an Attention LSTM Model with Mixup Data Augmentation. *Sensors*, 22(14)(November), 51–65.
- Suma, K. V., Koppad, D., & Kumar, P. (2023). Multi-Task Learning for Lung sound & Lung disease classification. *IEEE International Conference on Distributed Computing*, 1–22.
- Wang, F., Yuan, X., Liu, Y., & Lam, C. (2024). Journal of King Saud University - Computer and Information Sciences LungNeXt: A novel lightweight network utilizing enhanced mel-spectrogram for lung sound

- classification. *Journal of King Saud University - Computer and Information Sciences*, 36(8), 102200.
<https://doi.org/10.1016/j.jksuci.2024.102200>
- Yabani, M., Faisal, M. R., Indriani, F., & Turianto, D. (2026). *Lung Sound Classification Using Mel-Spectrogram Features for Early Detection of Respiratory Diseases*. 8(1), 168–184.
- Zhang, Q., Zhang, J., Yuan, J., Huang, H., Zhang, Y., Zhang, B., Lv, G., Lin, S., Wang, N., Liu, X., Tang, M., Wang, Y., Ma, H., Liu, L., & Yuan, S. (2022). *SPRSound : Open-Source SJTU Paediatric*. 99(99), 1–13.