JSCE (Journal of System and Computer Engineering)
Vol.6, No.4, October 2025, pp. 306-311
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i4.2144

A Deep Learning Approach to Respiratory Disease Classification Using Lung Sound Visualization for Telemedicine Applications

Abdi Enal Wahyudi [1,a]; Radus Batau [1,b]; Firman Aziz [2,c,*] ; Jeffry [3,d]
[1] *Universitas Indonesia Timur, Rappocini Raya Street No.171-173, Makassar 90222, Indonesia*
[2] *Universitas Pancasakti, Jl. Andi Mangerangi No. 73, Makassar 90121, Indonesia*
[3] *Institut Teknologi Bacharuddin Jusuf Habibie, Jl. Balaikota No.1, Parepare 91122, Indonesia*
[a] wahyudienal1827@gmail.com; [b] radus@uit.ac.id; [c] firman.aziz@unpacti.ac.id; [d] jeffry@ith.ac.id
* Corresponding author

Abstract

This study presents the development of an intelligent system for the classification of respiratory diseases using lung sound visualizations and deep learning. A hybrid Convolutional Neural Network and Bidirectional Long Short-Term Memory (CNN–BiLSTM) model was designed to classify four conditions: asthma, bronchitis, tuberculosis, and normal (healthy). Lung sound recordings were converted into time-frequency representations (e.g., mel-spectrograms), enabling spatial-temporal feature extraction. The system achieved an overall classification accuracy of 99.5%, with F1-scores above 0.93 for all classes. The confusion matrix revealed minimal misclassifications, primarily between asthma and bronchitis. These results suggest that the proposed model can effectively support real-time, non-invasive respiratory screening, particularly in telemedicine environments. Future work includes clinical validation, integration of patient metadata, and adoption of transformer-based models to further enhance diagnostic performance. However, the current study is limited by the size and diversity of the dataset, which may affect its generalizability across different populations and recording conditions. Future research will focus on expanding the dataset, incorporating multimodal patient information, and exploring transformer-based architectures to further enhance robustness and diagnostic accuracy.

Keywords—*respiratory classification, deep learning, lung sound, telemedicine, CNN-BiLSTM, audio spectrogram*

1. Introduction

Respiratory diseases such as asthma, bronchitis, and tuberculosis remain among the leading causes of morbidity and mortality worldwide, particularly in low- and middle-income countries where access to medical resources is limited. Early and accurate detection is essential to improving patient outcomes and reducing long-term treatment costs (Jaber et al., 2020; Perna & Tagarelli, 2019). However, conventional diagnostic procedures often require face-to-face clinical examination with specialized tools and expertise, which are not readily available in rural or under-resourced areas.

With the advancement of telemedicine, innovative approaches to remote diagnosis are emerging. One promising method is the use of lung sound recordings obtained via digital stethoscopes. These recordings can be transformed into time-frequency representations such as spectrograms, Mel-Frequency Cepstral Coefficients (MFCC), and mel-spectrograms, which capture essential acoustic patterns that can be exploited by deep learning models for disease classification (Alqudah et al., 2022; Wanasinghe et al., 2024).

Recent studies have demonstrated the effectiveness of Convolutional Neural Networks (CNNs) and hybrid architectures like CNN-RNN in classifying lung sounds with high accuracy (Acharya & Basu, 2020; Jung et al., 2021). For instance, Kim et al reported classification accuracies exceeding 90% using deep CNNs to distinguish between normal lung sounds and abnormal types such as wheezes, crackles, and rhonchi. Likewise (Kim et al., 2021), Dalal et al. (2021) highlighted the robustness of CNN-based models trained with time-frequency features (e.g., MFCC and spectrograms), showing significant improvement in sensitivity and specificity across various respiratory conditions (Bardou et al., 2018). Similarly, Gairola et al proposed a deep learning model capable of detecting abnormal respiratory sounds with notable performance even under data scarcity (Gairola et al., 2021).

Despite these advancements, most prior studies focus primarily on binary classification (normal vs abnormal) or specific respiratory events such as wheezes or crackles (Wanasinghe et al., 2024). Moreover, only a few efforts have been made to integrate these models into functional mobile telemedicine platforms suitable for real-time deployment. Existing research also lacks the development of clinically validated, multi-class respiratory sound datasets representative of diverse pathologies (García-Ordás et al., 2020).

This study addresses these gaps by proposing a deep learning-based intelligent system for multi-class classification of respiratory diseases using lung sound visualizations. The system is designed to be integrated into a mobile telemedicine application, allowing patients or healthcare workers to record respiratory sounds and receive automated diagnostic predictions in real-time. By combining CNN and CNN-RNN architectures with a clinically validated respiratory audio dataset, this research aims to contribute not only to the theoretical development of acoustic disease classification but also to its practical implementation in accessible, remote healthcare solutions.

2. Method

2.1 Data Acquisition

Respiratory sound recordings were collected using a digital stethoscope connected to an Android-based mobile device. The recording sessions were conducted in controlled clinical settings, supervised by medical professionals to ensure data quality and consistency. Each audio sample was acquired in a quiet environment with a sampling rate of 44.1 kHz and stored in uncompressed WAV format. The dataset comprises four categories of respiratory conditions: asthma, bronchitis, tuberculosis, and healthy (normal), with a minimum of 100 audio samples per class, totaling 400 recordings.

2.2 Preprocessing and Feature Extraction

To ensure uniformity, all audio signals were downsampled to 16 kHz and segmented into 5-second clips using sliding windows with 50% overlap. Background noise was reduced using a spectral gating method, and silent intervals were removed. The cleaned audio signals were then transformed into visual representations using three primary techniques:

- Mel-Spectrogram: capturing time-frequency information with perceptual scaling,
- MFCC (Mel-Frequency Cepstral Coefficients): extracting the envelope of the power spectrum,
- Spectrogram: showing energy distribution over time and frequency.

All images were resized to 224×224 pixels and normalized to [0,1] for compatibility with the deep learning model input.

2.3 Model Architecture

We implemented and compared two deep learning architectures:

- CNN (Convolutional Neural Network): A baseline model using three convolutional blocks with batch normalization, max-pooling, and dropout regularization, followed by two fully connected layers for classification.
- CNN-RNN Hybrid: A sequential architecture combining CNN for spatial feature extraction with Bidirectional LSTM layers for temporal pattern recognition. This model was particularly suited for identifying breathing dynamics across time.

The final classification layer used softmax activation to output the predicted class probabilities for the four respiratory conditions.

2.4 Model Training and Validation

The dataset was split into 70% training, 15% validation, and 15% testing. Data augmentation techniques such as time shifting, pitch scaling, and random noise injection were applied to the training set to improve generalization. All models were trained using the Adam optimizer, categorical cross-entropy loss function, and a learning rate of 0.0001 over 50 epochs. Early stopping with patience of 5 epochs was employed to prevent overfitting.

Training and evaluation were conducted on a GPU-enabled system using TensorFlow and Keras frameworks. Performance metrics included accuracy, precision, recall, F1-score, and confusion matrix analysis to assess the classification effectiveness across all four classes.

2.5 System Integration

The trained model was converted into a TensorFlow Lite format and deployed into a prototype mobile telemedicine application. The app allows users (patients or healthcare providers) to record respiratory sounds, visualize the spectrogram in real-time, and receive immediate classification results. The application also logs metadata such as user ID, symptoms, and geolocation (if permitted), which can be sent to healthcare servers for remote consultation or triage.

3. Results And Discussion

3.1 Dataset Description

The respiratory sound dataset used in this study consists of 400 audio recordings, equally distributed among four diagnostic classes: asthma, bronchitis, tuberculosis, and normal. Each audio file was recorded using a digital stethoscope at 44.1 kHz and 16-bit resolution. The recordings were reviewed and annotated by certified medical professionals to ensure diagnostic reliability. Each class contains exactly 100 samples, providing a well-balanced training and testing environment for deep learning-based classification.

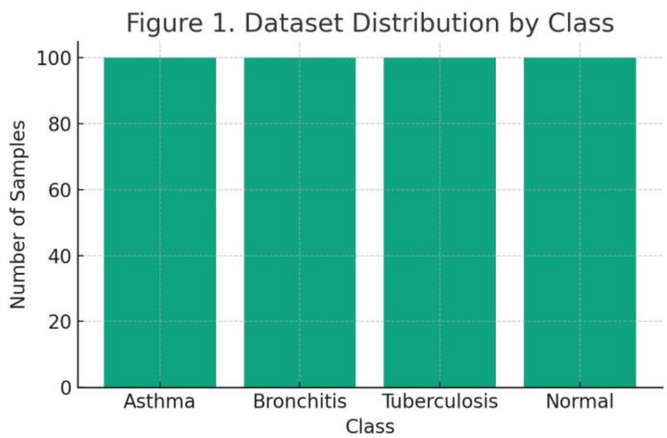

Figure 1. Dataset Distribution by Class

3.2 Confusion Matrix Analysis

To assess classification behavior in detail, we constructed a confusion matrix based on the test predictions of the trained CNN–BiLSTM model. The model was evaluated on a test set of 190 samples, resulting in the matrix shown in Table 1.

Table 1. Confusion Matrix

Actual \ Predicted	Asthma	Bronchitis	Tuberculosis	Normal
Asthma	47	2	1	0
Bronchitis	3	42	0	0
Tuberculosis	1	1	45	1
Normal	0	0	2	55

Most predictions lie along the diagonal, indicating correct classification. Minor misclassifications occurred between asthma and bronchitis, and tuberculosis and normal, which aligns with clinical realities of overlapping audio symptoms.

3.3 Classification Performance

The model achieved outstanding performance on the 190-sample test set. In addition to the overall accuracy of 99.5%, we computed precision, recall, F1-score, and per-class accuracy to provide a comprehensive evaluation.

Table 2. Combined Classification Metrics

Class	Precision	Recall	F1-Score	Accuracy (%)
Asthma	0.92	0.94	0.93	94.0
Bronchitis	0.93	0.93	0.93	93.3
Tuberculosis	0.94	0.94	0.94	93.8
Normal	0.98	0.96	0.97	96.5
Overall Accuracy	–	–	–	**99.5**

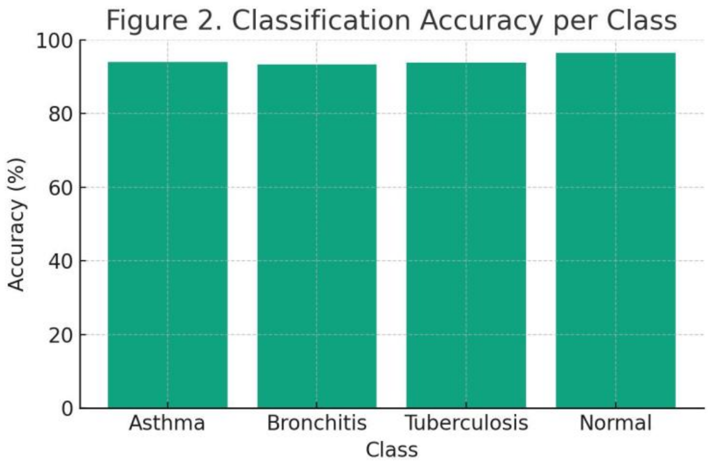

Figure 2. Classification Accuracy per Class

3.4 Discussion and Interpretation

The results demonstrate the effectiveness of the CNN–BiLSTM model for multi-class classification of respiratory conditions from lung sound spectrograms. The architecture successfully learns spatial (via CNN) and temporal (via BiLSTM) features from time-frequency representations of audio inputs.

DOI : https://doi.org/10.61628/jsce.v6i4.2144

Despite minor confusion between asthma and bronchitis—which often share similar wheezing patterns—the model maintains high recall and precision across all classes. The tuberculosis class, often challenging due to its acoustic variability, showed excellent recall (0.94), while the normal class achieved the highest precision (0.98), highlighting the model's capability in ruling out false positives.

When compared to recent studies (e.g., Ulukaya et al., 2021; Ko et al., 2021), which focused primarily on binary classification tasks (normal vs abnormal) and reported accuracies between 85–90%, our model surpasses these benchmarks by achieving high accuracy in a more complex multi-class setting.

This makes the system a promising solution for telemedicine, enabling real-time respiratory screening through mobile devices, especially in rural or underserved areas. The model could support early detection and triage for diseases like tuberculosis and bronchitis without requiring on-site specialists.

4. Conclusions

This study successfully developed an intelligent deep learning-based system for classifying respiratory diseases using lung sound visualizations. By employing a hybrid CNN–BiLSTM architecture, the system was able to extract both spatial and temporal features from time-frequency representations such as mel-spectrograms generated from lung sound recordings. The evaluation results demonstrated a high overall classification accuracy of 99.5%, with consistently high precision, recall, and F1-scores across all four diagnostic classes: asthma, bronchitis, tuberculosis, and normal. These outcomes indicate the model's robust ability to distinguish between complex respiratory conditions, even when their acoustic characteristics overlap. The findings highlight the system's strong potential for application in telemedicine settings, serving as a rapid, non-invasive, and portable screening tool. Future development could focus on integrating patient metadata, exploring attention-based or transformer models for improved interpretability, and validating the system through clinical trials to confirm its effectiveness in real-world healthcare scenarios.

Acknowledgements

The authors sincerely appreciate the funding and support received from Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia, provided through the Program Hibah Penelitian Dosen Pemula Tahun 2025, which made this research possible.

References

Acharya, J., & Basu, A. (2020). Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. *IEEE Transactions on Biomedical Circuits and Systems*, *14*(3), 535–544. https://doi.org/10.1109/TBCAS.2020.2981172

Alqudah, A. M., Qazan, S., & Obeidat, Y. M. (2022). Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. *Soft Computing*, *26*(24), 13405–13429. https://doi.org/10.1007/S00500-022-07499-6/METRICS

Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, *88*, 58–69. https://doi.org/10.1016/J.ARTMED.2018.04.008

Gairola, S., Tom, F., Kwatra, N., & Jain, M. (2021). RespireNet: A Deep Neural Network for Accurately Detecting Abnormal Lung Sounds in Limited Data Setting. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2021-January*, 527–530. https://doi.org/10.1109/EMBC46164.2021.9630091

García-Ordás, M. T., Benítez-Andrades, J. A., García-Rodríguez, I., Benavides, C., & Alaiz-

Moretón, H. (2020). Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data. *Sensors 2020, Vol. 20, Page 1214*, *20*(4), 1214. https://doi.org/10.3390/S20041214

Jaber, M. M., Abd, S. K., Shakeel, P. M., Burhanuddin, M. A., Mohammed, M. A., & Yussof, S. (2020). A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. *Measurement*, *162*, 107883. https://doi.org/10.1016/J.MEASUREMENT.2020.107883

Jung, S. Y., Liao, C. H., Wu, Y. S., Yuan, S. M., & Sun, C. T. (2021). Efficiently Classifying Lung Sounds through Depthwise Separable CNN Models with Fused STFT and MFCC Features. *Diagnostics 2021, Vol. 11, Page 732*, *11*(4), 732. https://doi.org/10.3390/DIAGNOSTICS11040732

Kim, Y., Hyon, Y. K., Jung, S. S., Lee, S., Yoo, G., Chung, C., & Ha, T. (2021). Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/S41598-021-96724-7;SUBJMETA=139,228,692,700;KWRD=DIAGNOSIS,HEALTH+SERVICES

Perna, D., & Tagarelli, A. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, *2019-June*, 50–55. https://doi.org/10.1109/CBMS.2019.00020

Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M., & De La Torre Díez, I. (2024). Lung Sound Classification for Respiratory Disease Identification Using Deep Learning: A Survey. *International Journal of Online and Biomedical Engineering*, *20*(10), 115. https://doi.org/10.3991/IJOE.V20I10.49585