Lung Disease Identification System Based on Cough Sound Using Machine Learning Without Signal Segmentation

1st Audrey Ndjukatana Hamaratu Tidah
School of Biomedical Engineering
Telkom University
Bandung, Indonesia
audreytidah@student.telkomuniversity.ac.id

2nd Hilman Fauzi T.S.P
School of Biomedical Engineering
Telkom University
Bandung, Indonesia
hilmanfauzitsp@telkomuniversity.ac.id

3rd Thomhert S. Siadari
School of Biomedical Engineering
Telkom University
Bandung, Indonesia
thomhert@telkomuniversity.ac.id

***Abstract*—Respiratory diseases, including tuberculosis (TB) and COVID-19, continue to impose a substantial burden on public health, especially in regions where access to specialized respiratory care remains limited. This work presents an automated lung disease identification framework based on cough sound analysis using deep learning techniques without applying signal segmentation. The proposed system categorizes audio recordings into four classes: cough, non-cough, tuberculosis, and COVID-19. Prior to classification, the audio signals undergo silence removal, DC offset correction, amplitude normalization, and duration standardization. Acoustic representations are then extracted using Mel-Frequency Cepstral Coefficients (MFCC), delta coefficients, delta-delta coefficients, Root Mean Square (RMS), and Zero Crossing Rate (ZCR). Two sequential learning models, namely Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), are evaluated and compared. Experimental findings show that the LSTM model achieved superior performance, obtaining an accuracy of 92% and an F1-score of 56%, whereas the RNN model attained an accuracy of 78% and an F1-score of 43%. The results suggest that preserving complete cough recordings allows important temporal patterns to be retained, leading to more effective disease classification. These findings demonstrate the potential of cough-based analysis as a practical, low-cost, and non-invasive approach for preliminary respiratory disease screening.**

***Index Terms*—cough sound, lung disease, MFCC, LSTM, RNN**

I. INTRODUCTION

Lung diseases remain one of the leading causes of death worldwide. Tuberculosis, pneumonia, and COVID-19 are respiratory diseases with high transmission rates and significant impacts on public health. According to the Global Tuberculosis Report 2023 published by the World Health Organization (WHO), Indonesia ranks second in the world with approximately 1,090,000 tuberculosis cases and 131,200 deaths [1]. The high prevalence of lung diseases highlights the importance of early detection to reduce transmission rates and improve treatment effectiveness.

The high prevalence of lung diseases in Indonesia is not supported by an equal distribution of pulmonary specialists. Based on data from the Indonesian Society of Respirology (PDPI), the number of active pulmonary specialists in Indonesia in 2022 was only 1,426 physicians [2]. Most pulmonary specialists are concentrated in large cities, while remote areas still face limited access to respiratory healthcare services. This condition often leads to delayed diagnosis and treatment.

Previous studies on lung disease identification commonly used image processing methods based on chest X-ray images or cough sound analysis specifically for COVID-19 detection. X-ray-based diagnostic methods require relatively high examination costs and specialized medical facilities [3]. Meanwhile, cough sound-based systems are still generally limited to identifying a single disease, such as COVID-19 [4], and have not been widely applied for broader lung disease screening.

Along with the advancement of artificial intelligence technology, cough sound analysis has emerged as an alternative approach for preliminary respiratory disease screening. Cough sounds contain unique acoustic characteristics that can be used to distinguish normal and abnormal respiratory conditions. Deep learning methods enable automatic cough sound classification based on temporal patterns and frequency characteristics of audio signals.

Several previous studies applied signal segmentation techniques to isolate certain parts of cough sounds before classification. However, segmentation may remove important temporal information from the audio signal. Since cough sounds are sequential signals with temporal dependencies between signal segments, the use of complete cough signals is expected to improve classification performance.

To address these issues, this research develops a lung disease classification framework based on cough audio analysis using deep learning techniques without applying signal segmentation. Acoustic features, including MFCC, delta, delta-delta, RMS, and ZCR, are extracted from the recordings and utilized as model inputs. The study further evaluates and compares the effectiveness of RNN and LSTM architectures for respiratory disease classification.

II. METHODOLOGY

The research methodology consists of several main stages, namely dataset collection, audio preprocessing, feature extraction, classification using deep learning models, and model performance evaluation. The developed system aims to identify lung diseases based on cough sounds without using signal segmentation.

A. System Design

The system design describes the workflow of the proposed research system to ensure that the system operates according to the designed process. The overview of the proposed system design is illustrated in Fig. 1.

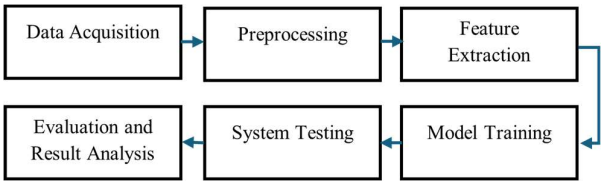

FIGURE 1
Research system flowchart

B. Data Acquisition

The dataset used in this study was obtained from the Kaggle platform and consists of four audio classes, namely cough, non-cough, tuberculosis, and COVID-19 [5]–[8]. The cough dataset was used to represent normal cough sounds, while the non-cough dataset contains sounds other than coughs such as conversations, environmental sounds, and other activity sounds. The tuberculosis and COVID-19 datasets were used to represent cough sounds originating from patients with specific lung diseases. The dataset distribution is shown in Table I.

TABLE 1
DATASET DISTRIBUTION

No	Class Name	Total Data
1	Cough	1,981
2	Non-cough	82,507
3	Tuberculosis	29,768
4	COVID-19	1,241
Total		115,497

All datasets were stored in WAV format and utilized throughout the model development process, including training, validation, and testing stages. The dataset was partitioned into training, validation, and testing subsets with proportions of 70%, 15%, and 15%, respectively.

To minimize potential sampling bias and improve model generalization, the data allocation was performed randomly. Ethical clearance for the use of the datasets was obtained from Institut Teknologi Bandung (ITB) under approval number KEP/II/2025/X/M271025ANH-PSIP.

C. Preprocessing

The preprocessing stage was carried out to improve the quality of audio signals before feature extraction. The preprocessing stages include trimming, DC offset removal, normalization, and truncation. Trimming was performed to remove silence segments from the audio signal, while DC offset removal was used to center the signal amplitude around zero. Furthermore, normalization was applied to standardize the audio amplitude range, and truncation was used to equalize the duration of audio signals before the feature extraction process.

Unlike conventional methods that partition audio recordings into multiple segments, this study employs unsegmented cough signals. Retaining the full signal structure allows temporal relationships across the recording to remain intact, facilitating improved learning of cough-related acoustic patterns.

D. Feature Extraction

Following preprocessing, acoustic features were generated to transform cough recordings into numerical representations suitable for deep learning-based classification. Feature extraction was implemented using the Librosa library. The extracted feature set consisted of MFCC, delta, delta-delta, RMS, ZCR.

MFCC was utilized to represent the spectral characteristics of cough sounds. The audio signal was segmented into short frames and transformed into the frequency domain using FFT. The resulting spectrum was then processed using a Mel filter bank to obtain a perceptually relevant frequency representation. Finally, DCT was applied to generate compact cepstral coefficients that summarize the acoustic information contained in the signal. [9]–[13].

$$c_n = \sum_{m=1}^{M} S_m \cos\left[\frac{\pi n}{M}(m-0.5)\right] (1)$$

Besides MFCC, temporal information was incorporated through delta and delta-delta coefficients. These features describe the dynamic behavior of cepstral representations across consecutive frames, enabling the model to capture both short-term variations and higher-order temporal dependencies within cough recordings.

The RMS feature was used to measure the energy or amplitude strength of audio signals. RMS helps the model distinguish the intensity of the cough among different classes of lung disease [10].

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n]^2} \quad (2)$$

Furthermore, the ZCR feature was used to calculate the number of signal sign changes from positive to negative or vice versa within an audio frame. ZCR represents the frequency variation level of cough sound signals [12], [14].

$$ZCR = \frac{1}{2N}\sum_{n=1}^{N} |\text{sign}(x[n]) - \text{sign}(x[n-1])| \ (3)$$

All extracted features were then combined into numerical representations used as input for the RNN and LSTM classification models. The feature extraction configuration used in this study is shown in Table 2.

TABLE 2
FEATURE EXTRACTION CONFIGURATION

Parameter	Value
Sampling Rate	16,000 Hz
Number of MFCC	40
Frame Length	1024
Hop Length	512

E. Classification Model

This study employs two deep learning models, namely Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), to classify lung diseases based on cough sounds. Both models were selected because of their capability to process sequential data and learn temporal relationships in audio signals [15]–[17].

1) Recurrent Neural Network (RNN): Machine Learning architecture specifically developed for sequential data analysis. Through recurrent connections, the network can incorporate information from previous time steps when generating predictions. In this study, the RNN model was employed to capture temporal patterns embedded in the extracted cough sound features [4], [11], [15].

2) Long Short-Term Memory (LSTM): a variant of

RNN that introduces memory units and gating mechanisms to address limitations associated with conventional RNN architectures, particularly the vanishing gradient problem. These components enable the network to retain relevant information over longer sequences, making LSTM suitable for modeling temporal dependencies in cough audio recordings [16], [17].

3) Model Training: The extracted feature vectors were used as inputs for the classification models during the training stage. Model optimization was performed using the Adam optimizer in combination with the categorical crossentropy loss function, which is appropriate for multi-class classification tasks. To improve model robustness and reduce the risk of overfitting, Early Stopping and Model Checkpoint strategies were incorporated throughout the training process. The detailed training parameters are summarized in Table III.

TABLE 3
TRAINING MODEL CONFIGURATION

Parameter	Value
Epoch	50
Batch Size	16
Optimizer	Adam
Learning Rate	Default Adam
Loss Function	Categorical Crossentropy
Metrics	Accuracy
Callback	Early Stopping / Model Checkpoint

F. Performance Metrics

Classification performance was evaluated by analyzing the prediction results generated by the models across four audio categories. A confusion matrix was utilized as the primary evaluation tool to quantify correct and incorrect predictions through TP, TN, FP, and FN values, as illustrated in Table IV.

TABLE 4
CONFUSION MATRIX

	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Based on the values of the confusion matrix, several evaluation metrics were calculated, including precision, precision, recall, specificity, and F1-score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

$$F1\text{-}score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (8)$$

Accuracy was used to measure the overall classification correctness. Precision was used to evaluate the accuracy of positive predictions, while sensitivity (recall) was used to measure the model's ability to detect positive classes. Specificity was used to evaluate the model's capability to correctly recognize negative classes. Furthermore, the F1-score was used to assess the balance between precision and sensitivity. Higher evaluation metric values indicate better performance of the model in classifying cough sounds for the lung disease identification system.

III. RESULTS

Experimental evaluation indicates that cough-based lung disease classification can be effectively performed without applying signal segmentation. Among the investigated models, LSTM achieved superior classification performance. The assessment was carried out using accuracy, precision, recall, and F1-score as evaluation criteria.

Table V summarizes the performance of both RNN and LSTM classifiers. The results demonstrate that the LSTM model yielded higher scores for all evaluation metrics compared with the RNN model.

TABLE 5
CLASSIFICATION RESULTS

Model	Accuracy	Precision	Recall	F1-Score
RNN	0.78	0.47	0.55	0.43
LSTM	0.92	0.54	0.76	0.56

As shown in Table V, the LSTM model achieved superior performance across all evaluation metrics. The model obtained an accuracy of 0.92, a recall of 0.76, and an F1-score of 0.56, whereas the RNN model achieved an accuracy of 0.78, a recall of 0.55, and an F1-score of 0.43. The higher F1-score obtained by the LSTM model indicates a better trade-off between precision and recall, highlighting its effectiveness in classifying cough sounds related to lung diseases.

Furthermore, the LSTM model demonstrated a more stable training behavior, characterized by lower loss values and more consistent validation performance. These results suggest that LSTM possesses stronger generalization capability than RNN, which exhibited larger performance fluctuations during training.

A. Discussion

The results indicate that the non-segmentation approach was able to preserve the temporal and contextual characteristics of cough sound signals. By using complete audio signals, the model can learn cough frequency patterns, duration, and signal dynamics more effectively compared to segmentation-based approaches.

The superior performance of the LSTM model compared to the RNN model indicates that temporal information plays an important role in cough sound classification. LSTM contains memory cells and gating mechanisms that enable the model to retain important long-term information, making it more effective in learning sequential patterns in audio signals.

Although the LSTM model achieved high accuracy, the precision values of both models were still relatively low. This condition indicates that false positive predictions were still present during the classification process. One possible reason is the use of non-segmented audio signals, where noise, background sounds, and non-cough components were still included in the input data.

Furthermore, the LSTM model demonstrated a more stable training process compared to the RNN model. The smaller gap between training and validation performance indicates that the LSTM model did not experience significant overfitting. In contrast, the RNN model showed larger training fluctuations, resulting in lower generalization capability.

Overall, the results demonstrate that the LSTM model with a non-segmentation approach provides the

best performance for cough sound-based lung disease identification systems.

B. Limitations

This study still has several limitations. One of the main limitations is the use of complete audio signals without segmentation, causing noise, background sounds, and non-cough segments to still be processed by the model. This condition resulted in relatively low precision values for both models due to false positive predictions.

In addition, the dataset distribution among classes was imbalanced, particularly for the COVID-19 and cough classes, which contained fewer samples compared to other classes. Dataset imbalance may affect the generalization capability of the model in recognizing all classes optimally.

This study is also limited to audio-based features and has not incorporated other data sources such as medical
imaging or clinical data to improve the performance of the lung disease identification system.

C. Future Work

Future research can be improved by incorporating noise reduction methods and adaptive segmentation techniques to enhance audio signal quality before classification.

In addition, data augmentation techniques and dataset balancing methods can be applied to improve model generalization capability. Future studies may also develop more advanced deep learning architectures to improve cough sound classification performance.

Further development can also be carried out by implementing the cough sound-based lung disease identification system into mobile or web-based applications so that it can be used as a more accessible early screening system for the public.

IV. CONCLUSION

This study successfully developed a lung disease identification system based on cough sounds using deep learning methods without signal segmentation. The system classified four audio classes, namely cough, non-cough, tuberculosis, and COVID-19, using MFCC, delta, delta-delta, RMS, and ZCR features as input for the classification models.

The results indicate that the LSTM model consistently achieved higher performance than the RNN model, reaching an accuracy of 0.92 and an F1-score of 0.56. By contrast, the RNN model achieved an accuracy of 0.78 and an F1-score of 0.43. These findings demonstrate that LSTM is more effective in learning temporal information from cough signals, leading to improved lung disease classification performance.

The findings also suggest that utilizing complete cough recordings without segmentation helps retain important temporal and contextual information contained in the audio signal. Furthermore, the LSTM model exhibited more stable training behavior and stronger generalization capability than the RNN model.

Despite limitations related to background noise and dataset imbalance, the proposed approach demonstrated promising results for cough-based lung disease classification. Therefore, the developed framework may provide a foundation for future respiratory screening systems that are affordable, non-invasive, and widely accessible.

REFERENCES

[1] World Health Organization, *Global Tuberculosis Report 2023*, 2023.

[2] Perhimpunan Dokter Paru Indonesia, "Distribusi Spesialis Paru di Indonesia," 2022.

[3] D. R. Ningtias, T. Suhartono, M. Rofi'i, and B. Wahyudi, "Pengembangan Artificial Intelligence untuk Deteksi Penyakit Pneumonia Berdasarkan Citra X-Ray," in *Prosiding Seminar Nasional Fisika*, vol. 35, 2023.

[4] K. A. Nasab, J. Mirzaei, A. Zali, S. Gholizadeh, and M. Akhlaghdoust, "Coronavirus diagnosis using cough sounds: Artificial intelligence approaches," *Frontiers in Artificial Intelligence*, vol. 6, Feb. 2023, doi: 10.3389/frai.2023.1100112.

[5] C. Filo, "UrbanSound8K," Kaggle Dataset.

[6] H. Sharma, "Cough Audio Dataset," Kaggle Dataset.

[7] H. Sharma, "CoughClassifier Trial," Kaggle Dataset.

[8] R. Shirsath, "TB Audio," Kaggle Dataset.

[9] S. Hidayat, A. S. Anas, S. A. A. Yusuf, and M. Tajuddin, "Sistem Pengenalan Pembicara dengan Metode Wavelet-MFCC dan Pengklasifikasi Hidden Markov Models (HMM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 2, pp. 290–297, 2022.

[10] B. Soewito, "Algoritma Mel Frequency Cepstral Coefficient (MFCC) untuk Ekstraksi Suara," *Journal of Information Technology*, 2022.

[11] S. B. Mulia, N. W. Nugraha, and M. H. Robbani, "Implementasi Machine Learning untuk Identifikasi Orang Batuk/Bersin," *eProceedings of Engineering*, vol. 9, no. 5, pp. 2725–2730, 2022.

[12] H. N. P. Adiarso, A. B. Osmond, and A. L. Prasasti, "Penentuan Dialek Jawa Malang, Jember, Solo, dan Banyumas Menggunakan Metode Recurrent Neural Network," *eProceedings of Engineering*, vol. 6, no. 2, pp. 5633–5638, Aug. 2019.

[13] R. Pratama, J. Raharjo, and B. Hidayat, "Deteksi Klasifikasi Ruangan Berdasarkan Reverberation Time Dengan Metode Mel-Frequency Cepstral Coefficients (MFCC) Dan Learning Vector Quantization (LVQ)," *eProceedings of Engineering*, 2023.

[14] G. F. Ajrin, R. Magdalena, and B. Hidayat, "Identifikasi Dialek Suku Bangsa Menggunakan Metode Mel-Frequency Cepstral Coefficient dan Zero Crossing Rate dengan Deep Neural Network Classifier," *eProceedings of Engineering*, vol. 9, no. 6, pp. 2958–2963, Dec. 2022.

[15] S. Nurmaini *et al.*, *Pengenalan Deep Learning dan Implementasinya*. Palembang, Indonesia: UPT Penerbit dan Percetakan Universitas Sriwijaya, 2021.

[16] MathWorks, "Recurrent Neural Network (RNN)."

[17] S. Larasati, S. Saadah, and P. E. Yunanto, *Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) Methods to Forecast Daily Turnover at BM Motor Ngawi*. Bandung, Indonesia: Universitas Telkom, 2024.