International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 4, August 2025, pp. 3737~3747
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i4.pp3737-3747

Integrating time-frequency features with deep learning for lung sound classification

Su Yuan Chang[1,2], Marni Azira Markom[1,3], Zhi Sheng Choong[1,4], Arni Munira Markom[5], Latifah Munirah Kamaruddin[1,3], Erdy Sulino Mohd Muslim Tan[3,6]
[1]Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, Arau Perlis, Malaysia
[2]AdoreSys Pte. Ltd., Penang, Malaysia
[3]Center of Excellence for Advanced Sensor Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, Arau Perlis, Malaysia
[4]SkyeChip Sdn Bhd., Penang, Malaysia
[5]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Selangor, Malaysia
[6]Faculty of Electrical Engineering and Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, Arau Perlis, Malaysia

Article Info

Article history:

Received Aug 5, 2024
Revised Apr 15, 2025
Accepted May 24, 2025

Keywords:

Deep learning
Lung sound
Lung sound classification
Medical signal processing
Time-frequency feature

ABSTRACT

Deep learning has transformed medical diagnostics, especially in analyzing lung sounds to assess respiratory conditions. Traditional methods like computed tomography scans (CT scans) and X-rays are impractical in resource-limited settings due to radiation exposure and time consumption, while conventional stethoscopes often lead to misdiagnosis due to subjective interpretation and environmental noise. This study evaluates deep learning models for lung sound classification using the International Conference on Biomedical Health Informatics 2017 dataset, comprising 920 annotated samples from 126 subjects. Pre-processing includes down sampling, segmentation, normalization, and audio clipping, with feature extraction techniques like spectrogram and Mel-frequency cepstral coefficients (MFCC). The adopted automatic lung sound diagnosis network (ASLD-Net) model with triple feature input (time domain, spectrogram, and MFCC) achieved the highest accuracy at 97.25%, followed by the dual feature model (spectrogram and MFCC) at 95.65%. Single-input models with spectrogram and MFCC performed well, while the time domain input alone had the lowest accuracy.

Corresponding Author:

Marni Azira Markom
Biomedical Electronic Engineering Programme, Department of Electronic Engineering, Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis
Kampus Pauh Putra, 02600 Arau Perlis, Malaysia
Email: marni@unimap.edu.my

1. INTRODUCTION

The necessity for enhanced diagnostics is underscored by the fact that lung diseases, which are further exacerbated by COVID-19, continue to be the primary cause of death on a global scale [1], [2]. Over 14% of fatalities worldwide are caused by pulmonary disorders, including asthma, chronic obstructive pulmonary disease (COPD), pneumonia, lung cancer, and tuberculosis [3]. Although traditional tools like CT scans and X-rays are useful, they are often time-consuming, entail radiation exposure, and may be inaccessible in resource-limited settings [4]. Concerns regarding diagnostic accuracy are raised by the imprecise results that conventional stethoscopes frequently produce as a result of subjective interpretation and environmental noise [5]. For improved diagnostic accuracy, digital stethoscopes with deep learning algorithms convert audio signals to digital format, to provide non-invasive, real-time monitoring [6], [7]. By precisely extracting and categorizing pulmonary sounds, they enable remote patient monitoring and medical education [7]. This study evaluates the efficacy in the classification of lung sounds using deep learning with a variety of feature inputs, with the objective of enhancing diagnostic accuracy.

Precise identification of lung sounds is essential for disease diagnosis and patient care. However, conventional stethoscopes frequently produce indistinct outcomes due to ambient noise and subjective interpretation. The objective of this work is to improve the analysis of lung sounds by employing pre-processing techniques, extracting relevant features, utilizing deep learning for classification of normal and abnormal lung sounds by employing distinct feature inputs. The raw data of lung sounds exhibit irregularities, which poses a challenge for direct classification. Pre-processing guarantees consistency, facilitating precise feature extraction and classification. In order to enhance the classification efficiency, it is necessary to optimize the feature extraction methods due to the intricate nature of respiratory sounds. Classifying a wide range of respiratory abnormalities remains challenging due to the quality of the data and the importance of the features, despite the existence of numerous deep learning models. This work investigates the effectiveness of various types of features utilized in the classification of lung sounds by using deep learning. Thorough assessment of these models is crucial in order to identify the most efficient feature sets for identifying respiratory problems and incorporating them into clinical practice.

This paper significantly contributes to lung disease diagnosis by providing an advanced method for capturing and analyzing lung sounds using deep learning and digital stethoscopes. It introduces a robust pre-processing framework to standardize lung sound data and compares feature extraction methods, including Mel-frequency cepstral coefficient (MFCC), spectrograms, and their combinations. A key contribution is the development of the automatic lung sound diagnosis network (ALSD-Net), which combines two-dimensional and one-dimensional convolutional layers to capture intricate lung sound patterns. Comprehensive evaluation highlights the superiority of integrated features, with the triple feature model achieving the highest accuracy. These findings demonstrate how deep learning-based digital stethoscopes can enhance non-invasive, real-time monitoring and remote patient care, especially in resource-limited settings.

2. RELATED PREVIOUS STUDIES

Effective pre-processing is crucial for accurately classifying lung sounds using deep learning techniques. Resampling, as done by [8], ensures uniformity across datasets by choosing a sample rate of 44.1 kHz. In [9], the audio frequency was adjusted to 22 kHz, whereas [10] resampled all recordings at 4,000 Hz, considering that the primary signal of interest is mostly below 2,000 Hz. Filtering methods, such as Gaussian Butterworth filters used by [11], second-order Butterworth bandpass filters utilized by [12], and wavelet denoising with high-pass filtering employed by [13], successfully remove unwanted noise and retain important signals. For audio clipping techniques, study [10] fit cycles into duration segments of 2.7 s, ensuring data consistency and compatibility. Data augmentation methods, such as time stretching and vocal tract length perturbation (VTLP), are employed by [9] to increase data quality and diversity.

Mel-spectrograms and spectrograms obtained using short-time Fourier transform (STFT) are essential for examining lung sounds. Mel-spectrograms, as employed by [13] and [14], convert frequency components to the Mel scale, providing perceptually significant information for sound analysis. STFT spectrograms, as demonstrated by [9] and [10], enable the simultaneous collection of temporal and frequency features, which is crucial for accurately classifying respiratory cycles. Spectrogram clipping improves processing efficiency by emphasizing important sound features. Mel-frequency cepstral coefficients (MFCCs), obtained using FFT and discrete cosine transform (DCT), accurately characterize spectrum attributes. Researchers such as [15] and [16], have utilized MFCCs to improve classification accuracy by reducing feature correlation and emulating human auditory perception. These techniques enhance the quality and comprehensibility of feature representations vital for the analysis and categorization of lung sounds.

Deep learning algorithms have demonstrated encouraging outcomes in categorizing lung sounds. Researchers [14] achieved 94% accuracy using a VGGish-stacked bidirectional gated recurrent unit (BiGRU) model, focusing on precision, recall, and F1-score. In [11], the ALSD-Net, a convolutional neural network (CNN), achieved 94.24% accuracy. Another study [12] integrated a CNN with best discrepancy forest (BDF), attaining remarkable performance with 99.94% accuracy and impressive precision, specificity, sensitivity, and F1-score metrics. Traditional machine learning methods have also been successful: [8] achieved 99% accuracy with fine Gaussian support vector machine (SVM), and [15] obtained 97.45% accuracy with gradient boosting. Hybrid models, like the CNN bi-directional long short-term memory (BDLSTM) used by [17], achieved 98.26% accuracy, effectively integrating temporal and spatial variables. This research highlights significant advancements in AI algorithms for the accurate classification and diagnosis of respiratory disorders.

3. METHOD

This study adopts a structured approach to develop a deep learning model for lung sound classification. It involves multiple stages, beginning with signal acquisition and followed by pre-processing, feature extraction, model development, training, and evaluation. Each stage plays a critical role in ensuring the accuracy and reliability of the proposed system, and they are detailed in the following subsections.

3.1. Lung sound signal acquisition

The dataset from the International Conference on Biomedical and Health Informatics (ICHBI) 2017 Challenge is utilized in this study. It is a scientific challenge that took place in 2017, which offers a respiratory database and an official score system. A total of 5.5 hours of recordings with a collection of 920 annotated audio samples from 126 subjects are included in this database [18]. With a total of 6,898 cycles, these cycles are further categorized into 3,642 normal cycles, 1,864 containing crackles, 886 containing wheezes, and 506 containing both crackles and wheezes. Table 1 shows the lung sound characteristics.

Table 1. Lung sound characteristics

Lung sound	Cause	Acoustic characteristic	Associated disease
Normal	Smooth and continuous airflow through the respiratory passages without any abnormal disruptions.	Frequency range: 60–600 Hz	-
Crackle	Explosive opening of small airways or the alveoli, air bubble in larger airways.	Non-musical and explosive in nature Frequency range: 350–650 Hz Duration: <20 ms	Interstitial lung fibrosis Pneumonia COPD Bronchiectasis Asthma
Wheeze	Airflow limitation and airway narrowing.	Frequency range: 100–1,000 Hz Duration: >100 ms	COPD Asthma Tumor blocking airway

3.2. Pre-processing

This study utilizes various pre-processing techniques to prepare lung sound data for classification. Initially, down sampling is employed to decrease the sampling rate to 4,000 Hz in order to simplify the processing, while continuing to adhere to the guidelines given by the Nyquist theorem [10], [15]. Segmentation is the process of dividing recordings into respiratory cycles using annotated start and stop times. This results in distinct categories such as normal, crackles, wheezes, and combination of both crackles and wheezes. The audio clipping process ensures that all cycle durations are standardized to 2.7 seconds by either cropping excess duration from longer cycles or employing zero padding to shorter ones [10]. Amplitude normalization ensures consistency by rescaling all signals to the range of (-1, 1) in order to reduce variations caused by noise and physiological differences among patients [11]. The purpose of these pre-processing methods is to improve the consistency and excellence of lung sound data, making it more suitable for subsequent deep learning classification models.

3.3. Feature extraction

Extracting features from lung sound audio recordings is crucial for generating more manageable and useful information. The process involves the careful selection of significant features, streamlining the data, and identifying patterns associated with respiratory abnormalities. The selected features for this study consist of MFCC and spectrograms. Figure 1 illustrates three signal features plots that show the raw lung sound in time domain in Figure 1(a), the MFCC feature in Figure 1(b) and the spectrogram in Figure 1(c).

3.3.1. Mel-frequency cepstral coefficient

The process of extracting MFCC begins by applying the FFT on signals that have been windowed. A Mel-scale filter bank, specifically a triangle bandpass filter bank, is employed to convert the linear frequency spectrum into the Mel-frequency scale, replicating the auditory perception of humans. The formula employed as (1) [19]:

$$f_{mel} = 2596 * log(1 + \frac{f_{linear}}{700}) \quad (1)$$

where f_{mel} represents frequency in Mel scale and f_{linear} represents frequency in linear scale. Applying a logarithmic transformation after filtering decreases the impact of changes in amplitude. Afterwards, the logarithmic-scale signal is applied to the DCT in order to calculate the MFCCs. These coefficients quantify the amplitude of the spectrum in the time domain. This method improves the representation of essential acoustic features for classification tasks.

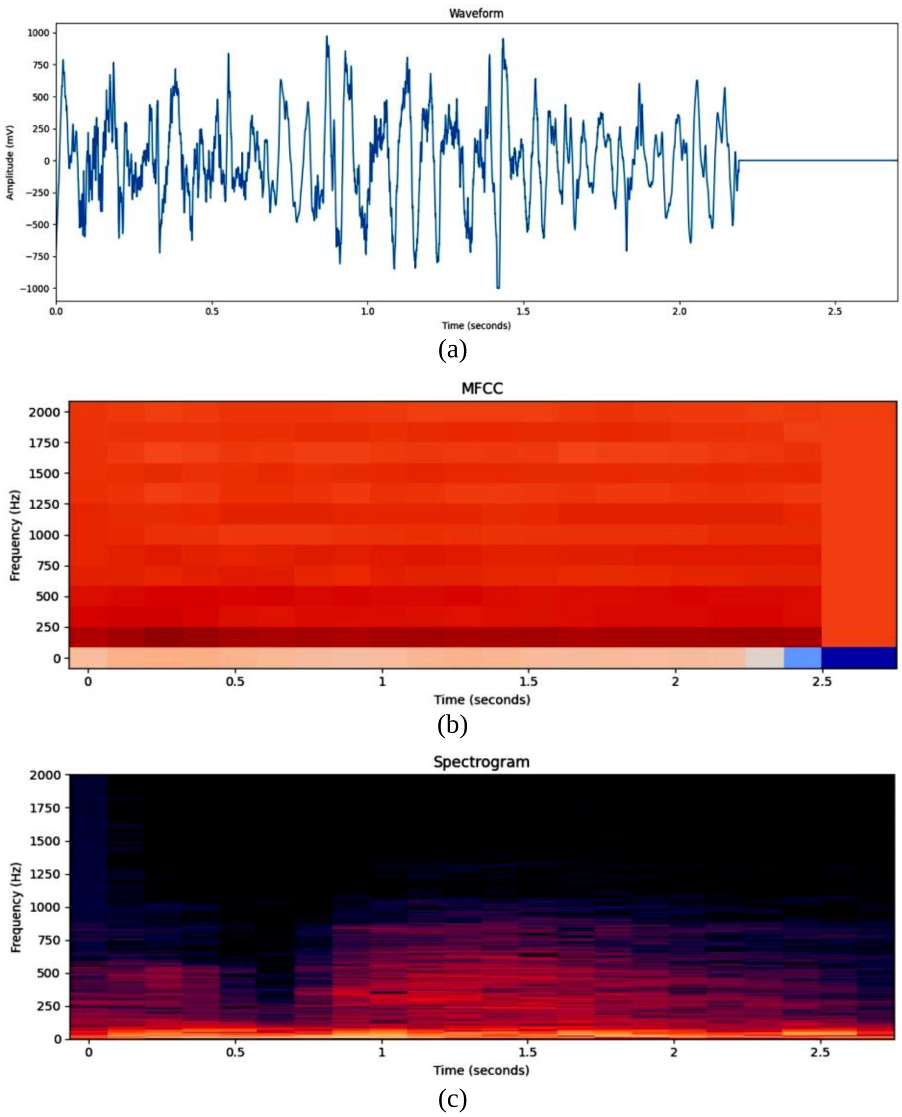


Figure 1. The lung sound signal, (a) the pre-processed signal in time domain, (b) MFCC feature, and (c) spectrogram feature

3.3.2. Spectrogram

Time-varying frequency components of audio signals can be identified by spectrograms, which are generated when the STFT transforms windowed portions of the signals into the frequency [20]. It is calculated by (2):

$$X(m,\omega) = \sum_{n=-\infty}^{\infty} x(n)\omega(n-m)e^{-j\omega n} \quad (2)$$

The discrete-time signal is represented by $x(n)$, the window function, which is usually Gaussian or Hanning, is represented by $\omega(n)$, the time index is m, and the angular frequency is indicated by ω. The spectrogram is a crucial tool for analyzing complicated, time-varying signals like respiratory cycles because it records how the frequency content changes over time by moving the window across the signal and computing the STFT at each place.

3.4. Deep learning classification

Using sequential CNN architecture designed for both audio and visual data processing, namely, the adopted ALSD-Net model offers an advanced method for automated lung sound diagnostics. This model involves two-dimensional convolutional layers to extract features from picture representations and one-dimensional convolutional layers to analyze audio recordings. Seven convolutional layers make up the model architecture. The first four layers each have 16 filters and rectified linear unit (ReLU) activation. To improve training efficiency and stability, max-pooling with a pool size of 2 and batch normalization are applied after each convolutional layer. With 32 filters apiece, subsequent convolutional layers preserve reliable feature extraction at various abstraction levels. In order to avoid overfitting during model training, dropout layers are carefully placed after the convolutional and dense layers, with rates of 0.2 and 0.4, respectively. In order to generate probabilistic predictions for the 4 types of lungs sounds that the model is targeting, the final layer uses a dense layer with SoftMax activation [11]. This architecture highlights the model's ability to identify complex patterns in lung sound data. Table 2 and Figure 2 show the hyperparameters set and the architecture model for ALSD-Net, respectively.

Table 2. Hyperparameters for ALSD-Net model

Hyperparameter	Value
Loss function	Categorical cross-entropy
Batch size	32
Epochs	50
Early stopping patience	10
Optimizer	Adam
Learning rate	1×10^{-5}
Activation function	ReLU

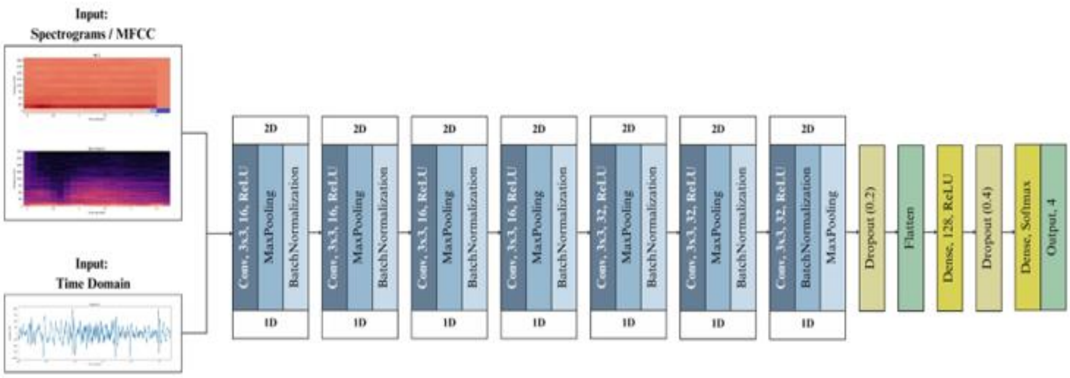


Figure 2. Structure of ALSD-Net model

3.5. Performance evaluation

The performance of the ALSD-Net model with distinct sets of feature input is evaluated by using important metrics and graphical tools such as learning curves and performance evaluation metrics. The first evaluation is the learning curves of the deep learning training and validation process, which illustrate how well a model predicts performance over time as a function of training effort, are crucial visual aids in deep learning. These curves primarily show the model's accuracy or loss based on the number of training epochs and both training and validation datasets. By analyzing these curves, significant details about the model's learning and generalization behaviors can be identified [21]. Then, once completed the training and validation, performance evaluation metrics that involve accuracy, precision, recall, and F1-score, are considered. These metrics provide insights into the strengths and weaknesses of different feature sets. Weighted averages of these metrics are computed to fairly compare models across balanced and imbalanced datasets. According to [22], assigning smaller weights to classes with more instances and larger weights to minority classes ensures a balanced evaluation, considering the influence of instances from all classes.

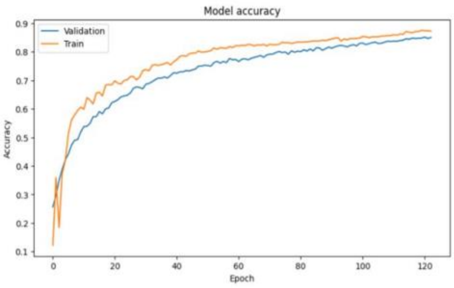
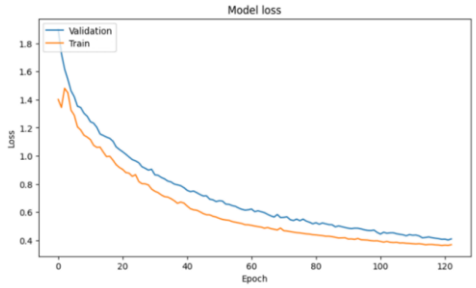
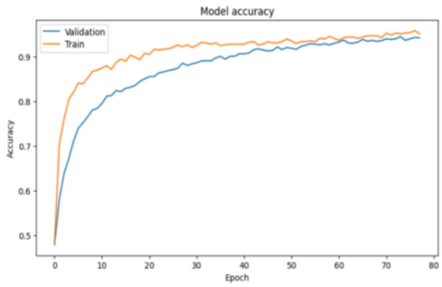
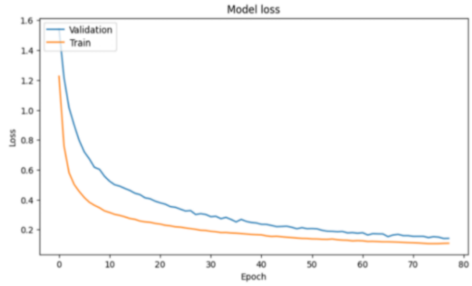
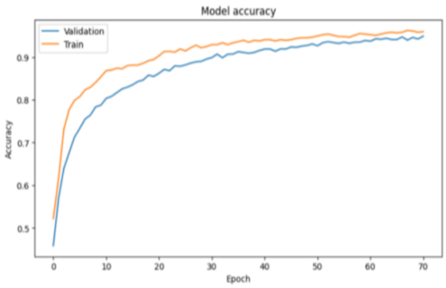
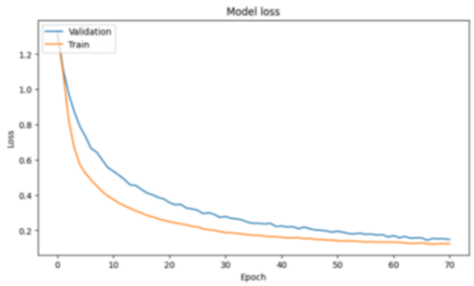
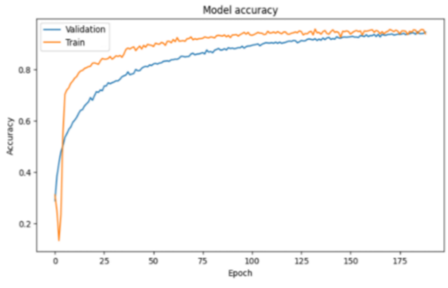
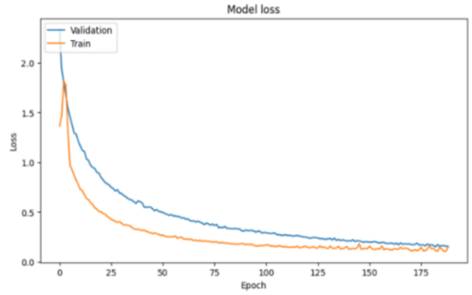
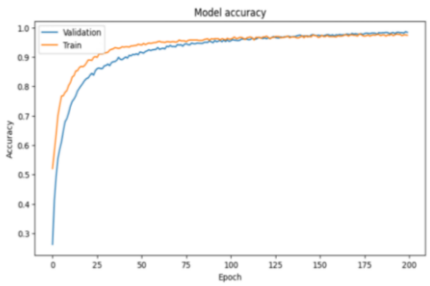
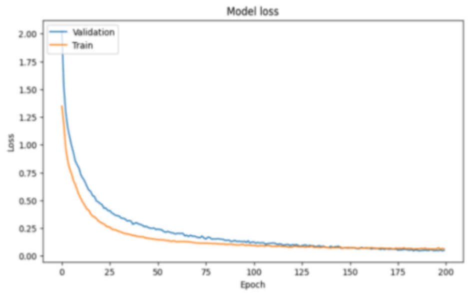
4. RESULTS AND DISCUSSION

This section presents the results and emphasizes the evaluation of the performance and efficiency of the proposed ALSD-Net model in classifying lung sounds. The assessment involves multiple types of lung sound features to understand their impact on the model's classification accuracy. These features include time domain signals, spectrogram representations, MFCC, dual feature input (a combination of spectrogram and MFCC), and triple feature input (a combination of time domain, spectrogram, and MFCC). The analysis aims to identify which feature set or combination provides the most effective input for enhancing the model's classification capabilities.

4.1. The adopted ALSD-Net training performance

The analysis focuses on accuracy and loss curves derived from training and validation phases as shown in Table 3. Rapid gains in accuracy are observed in the time domain input, which stabilizes at roughly 0.90 for training and 0.85 for validation. The associated loss values also show convergence below 0.4 and 0.5, respectively.

Table 3. The training performance

Feature	Accuracy curve	Loss curve
Time domain		
Spectrogram		
MFCC		
Spectrogram and MFCC		
Time domain, spectrogram and MFCC		

The spectrogram input also shows a faster accuracy increase, stabilizing at values close to 0.95 for training and 0.93 for validation, with just small losses between 0.1 and 0.18. Comparable patterns can be seen in the MFCC input, with loss values between 0.1 and 0.2 and accuracy values between 0.95 and 0.93. The integration of MFCC and spectrogram inputs results in high accuracy, approximately 0.94 for validation and almost 0.95 for training, with losses decreasing to about 0.20 and 0.25 for each situation. Specifically, when time domain, spectrogram, and MFCC inputs are combined, training and validation peak accuracies are approximately 0.98, while loss values are close to 0.10.

4.2. The adopted ALSD-Net testing performance

Each feature set was assessed using metrics such as accuracy, precision, recall, and F1-score for classifying normal lung sounds, crackles, wheezes, and combination of crackles and wheezes, as shown in Table 4. The results show significant differences in performance between various feature inputs. For example, the time domain feature demonstrated outstanding precision and recall metrics, reaching its maximum accuracy of 95.69% in categorizing normal lung sounds. By comparison, the spectrogram input demonstrated exceptional performance in identifying crackles, with a 99.44% accuracy rate, as well as good precision and recall ratings for this class. In a similar vein, the MFCC feature set performed admirably, achieving a 97.20% classification accuracy for crackles.

Table 4. Performance evaluation metrics of the adopted ALSD-Net based on multiple features

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Time domain feature				
Normal	95.69	88.78	95.70	92.11
Crackles	94.97	89.47	94.97	92.14
Wheezes	44.73	85.00	44.74	58.62
Both (Crackles + Wheezes)	84.12	89.83	84.13	86.89
Overall Accuracy				88.84
Spectrogram				
Normal	92.47	98.01	92.47	95.16
Crackles	99.44	95.15	99.44	97.27
Wheezes	90.79	76.67	90.79	83.13
Both (Crackles + Wheezes)	96.83	98.39	96.83	97.6
Overall Accuracy				94.49
MFCC				
Normal	92.47	97.45	92.47	94.9
Crackles	97.20	96.67	97.21	96.94
Wheezes	90.78	71.13	90.79	79.77
Both (Crackles + Wheezes)	92.06	96.67	92.06	94.31
Overall Accuracy				93.48
Spectrogram + MFCC				
Normal	95.84	95.42	96.73	96.07
Crackles	99.17	98.89	98.89	98.89
Wheezes	97.92	83.33	85.53	84.42
Both (Crackles + Wheezes)	99.58	93.94	98.41	96.12
Overall Accuracy				95.65
Time domain + Spectrogram + MFCC				
Normal	97.67	99.17	96.51	97.82
Crackles	99.57	98.90	99.44	99.17
Wheezes	97.67	85.71	94.74	90.00
Both (Crackles + Wheezes)	99.41	96.83	96.83	96.83
Overall Accuracy				97.25

The accuracy was improved overall by combining spectrogram and MFCC inputs, especially when identifying combinations of crackles and wheezes, which resulted in a 99.58% accuracy rate. The combination of spectrogram, MFCC, and time domain characteristics produced the best overall accuracy of 97.25%, demonstrating remarkable performance in the classification of wheezes and crackles. These results highlight the usefulness of feature integration in improving the lung sound classification accuracy of the adopted ALSD-Net, providing important new information for the development of reliable diagnostic instruments in respiratory healthcare applications.

Table 4 also compares the performance of different feature input sets for lung sound classification. The time domain input model shows the lowest performance, with an accuracy of 88.84%, precision of 88.64%, recall of 88.84%, and an F1-score of 87.91%. This indicates its limited ability to capture complex patterns, making it the least reliable approach.

In contrast, the triple feature input model, which combines time domain, spectrogram, and MFCC features, achieves the highest performance. It records an accuracy of 97.25%, a precision of 97.40%, recall of

97.10%, and an F1-score of 97.22%. This model's superior performance is due to the integration of varied features, which enhances its prediction accuracy and reliability.

The spectrogram and MFCC input models also perform well but not as highly as the triple feature model. The spectrogram model achieves an accuracy of 94.49%, precision of 94.95%, recall of 94.49%, and an F1-score of 94.61%. Similarly, the MFCC model records an accuracy of 93.48%, precision of 94.28%, recall of 93.48%, and an F1-score of 93.71%. These models outperform the time domain model significantly, highlighting the advantage of capturing frequency domain information.

The dual feature input model, combining spectrogram and MFCC features, shows impressive performance with an accuracy of 95.65%, precision of 94.85%, recall of 96.21%, and an F1-score of 95.52%. This model outperforms the individual spectrogram and MFCC models and closely matches the performance of the triple feature model. Integrating different feature types allows for the utilization of complementary information, resulting in improved overall performance.

4.3. Overall results

Table 5 compares the performance of different feature input sets for lung sound classification. The time domain input model shows the lowest performance, with an accuracy of 88.84%, precision of 88.64%, recall of 88.84%, and an F1-score of 87.91%. This indicates its limited ability to capture complex patterns, making it the least reliable approach.

In contrast, the triple feature input model, which combines time domain, spectrogram, and MFCC features, achieves the highest performance. It records an accuracy of 97.25%, a precision of 97.40%, recall of 97.10%, and an F1-score of 97.22%. This model's superior performance is due to the integration of varied features, which enhances its prediction accuracy and reliability.

The spectrogram and MFCC input models also perform well but not as highly as the triple feature model. The spectrogram model achieves an accuracy of 94.49%, precision of 94.95%, recall of 94.49%, and an F1-score of 94.61%. Similarly, the MFCC model records an accuracy of 93.48%, precision of 94.28%, recall of 93.48%, and an F1-score of 93.71%. These models outperform the time domain model significantly, highlighting the advantage of capturing frequency domain information.

The dual feature input model, combining spectrogram and MFCC features, shows impressive performance with an accuracy of 95.65%, precision of 94.85%, recall of 96.21%, and an F1-score of 95.52%. This model outperforms the individual spectrogram and MFCC models and closely matches the performance of the triple feature model. Integrating different feature types allows for the utilization of complementary information, resulting in improved overall performance.

The triple feature model incorporating time domain, spectrogram, and MFCC emerged as the top performer with an accuracy of 97.25% in this study that assesses various feature inputs for lung sound classification. This model shows strong precision and recall metrics and performs well in all categories ('Normal', 'Crackles', 'Wheezes', and combinations). The time domain model, on the other hand, has the lowest accuracy of any model, which is 88.84%, and finds it difficult to distinguish between various lung sound abnormalities due to its limited ability to capture intricate frequency and temporal fluctuations.

In comparison, single-input models such as spectrogram (94.49% accuracy) and MFCC (93.48% accuracy) perform better in certain areas. For example, spectrogram is better at capturing frequency data over time, while MFCC makes use of spectral features, but it might be less successful at capturing subtle time-frequency resolutions. These results are confirmed by earlier research, including works by [23] and [24], which demonstrate how well spectrogram performs in comparison to MFCC in tasks similar to classification.

The integration of multiple features in the dual feature model (combining spectrogram and MFCC) achieves an accuracy of 95.65%, surpassing single-feature models by leveraging their complementary strengths. This method is consistent with research by [25], which showed that merging spectrogram and MFCC characteristics increased accuracy when compared to utilizing them separately. All things considered, the model's capacity to precisely categorize intricate lung sound patterns is improved by the integration of time domain, spectrogram, and MFCC features, which makes it a promising development for respiratory healthcare applications.

Table 5. Comparison of weighted average across distinct feature input sets

Feature input	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Time domain	88.84	88.64	88.84	87.91
Spectrogram	94.49	94.95	94.49	94.61
MFCC	93.48	94.28	93.48	93.71
Dual feature input	95.65	94.85	96.21	95.52
Triple feature input	97.25	97.40	97.10	97.22

5. CONCLUSION

In conclusion, this study systematically evaluated the performance of different feature inputs within the adopted ALSD-Net for classifying lung sounds. The triple feature model, combining time domain, spectrogram, and MFCC, performed the best with 97.25% accuracy. The time domain model, on the other hand, struggled with the intricate frequency and timing patterns in lung sounds, yielding the lowest accuracy of 88.84%. While MFCC used spectrum characteristics but struggled with detailed timing, spectrogram (94.49% accuracy) and MFCC (93.48% accuracy) were two examples of single-feature models that demonstrated strengths in particular areas. Spectrogram was particularly good at capturing time-frequency dynamics.

Furthermore, the research highlights the noteworthy benefit of combining several features, as demonstrated by the dual feature model (combining spectrogram and MFCC) that attains a 95.65% accuracy rate, utilizing complementary advantages to improve classification capability. The best feature combinations for enhancing lung sound classification accuracy are shown by these studies, which advance the field. They are consistent with earlier studies showing spectrograms to be more effective than MFCC in comparable tasks, confirming the strategy of using feature richness to improve diagnostic precision in respiratory care. To further improve and validate these findings for wider clinical applications, future efforts could investigate hybrid feature combinations and larger datasets, ultimately enhancing the capabilities of automated lung sound analysis systems.

ACKNOWLEDGMENTS

Authors have no acknowledgments to declare.

FUNDING INFORMATION

Authors state no funding involved.

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
Chang Su Yuan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Marni Azira Markom	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Choong Zhi Sheng	✓		✓				✓			✓	✓		✓	
Arni Munira Markom		✓			✓		✓							✓
Latifah Munirah Kamarudin	✓			✓						✓				
Erdy Sulino Mohd Muslim Tan		✓										✓		✓

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.

INFORMED CONSENT

No informed consent was required, as the data used in this study were obtained from publicly available databases.

ETHICAL APPROVAL

No ethical approval was required for this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the ICBHI Respiratory Sound Database at https://doi.org/10.1088/1361-6579/ab03ea, as described in Rocha *et al.* (2019) [18].

REFERENCES

[1] W. Guan *et al.*, "Clinical characteristics of Coronavirus Disease 2019 in China," *New England Journal of Medicine*, vol. 382, no. 18, pp. 1708–1720, Apr. 2020, doi: 10.1056/NEJMoa2002032.

[2] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[3] D. Chamberlain, J. Mofor, R. Fletcher, and R. Kodgule, "Mobile stethoscope and signal processing algorithms for pulmonary screening and diagnostics," in *2015 IEEE Global Humanitarian Technology Conference (GHTC)*, Oct. 2015, pp. 385–392, doi: 10.1109/GHTC.2015.7344001.

[4] H. Choi *et al.*, "Extension of Coronavirus Disease 2019 on chest CT and implications for chest radiographic interpretation," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200107, Apr. 2020, doi: 10.1148/ryct.2020200107.

[5] Y.-C. Wu *et al.*, "Development of an Electronic Stethoscope and a classification algorithm for cardiopulmonary sounds," *Sensors*, vol. 22, no. 11, p. 4263, Jun. 2022, doi: 10.3390/s22114263.

[6] J. J. Seah, J. Zhao, D. Y. Wang, and H. P. Lee, "Review on the advancements of stethoscope types in chest auscultation," *Diagnostics*, vol. 13, no. 9, p. 1545, Apr. 2023, doi: 10.3390/diagnostics13091545.

[7] C. Koning and A. Lock, "A systematic review and utilization study of digital stethoscopes for cardiopulmonary assessments," *Journal of Medical Research and Innovation*, vol. 5, no. 2, pp. 4–14, Feb. 2022, doi: 10.25259/JMRI_2_2021.

[8] B. Abera Tessema, H. Nemomssa, and G. Lamesgin Simegn, "Acquisition and classification of Lung sounds for improving the efficacy of auscultation diagnosis of pulmonary diseases," *Medical Devices: Evidence and Research*, vol. Volume 15, pp. 89–102, Apr. 2022, doi: 10.2147/MDER.S362407.

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

[10] G. Petmezas *et al.*, "Automated Lung sound classification using a hybrid CNN-LSTM network and focal loss function," *Sensors*, vol. 22, no. 3, p. 1232, Feb. 2022, doi: 10.3390/s22031232.

[11] N. Baghel, V. Nangia, and M. K. Dutta, "ALSD-Net: Automatic lung sounds diagnosis network from pulmonary signals," *Neural Computing and Applications*, vol. 33, no. 24, pp. 17103–17118, Dec. 2021, doi: 10.1007/s00521-021-06302-1.

[12] M. Zhang, M. Li, L. Guo, and J. Liu, "A low-cost AI-empowered stethoscope and a lightweight model for detecting cardiac and respiratory diseases from lung and heart auscultation sounds," *Sensors*, vol. 23, no. 5, p. 2591, Feb. 2023, doi: 10.3390/s23052591.

[13] Y. Shi, Y. Li, M. Cai, and X. D. Zhang, "A Lung sound category recognition method based on wavelet decomposition and BP neural network," *International Journal of Biological Sciences*, vol. 15, no. 1, pp. 195–207, 2019, doi: 10.7150/ijbs.29863.

[14] K. N. Lal, "A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning," *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 36615–36631, Sep. 2023, doi: 10.1007/s11042-023-14727-0.

[15] B. Vidhya, M. Nikhil Madhav, M. Suresh Kumar, and S. Kalanandini, "AI based diagnosis of Pneumonia," *Wireless Personal Communications*, vol. 126, no. 4, pp. 3677–3692, Oct. 2022, doi: 10.1007/s11277-022-09885-7.

[16] J. Amose, P. Manimegalai, and R. Pon Selchiya, "Classification of adventitious Lung sounds: Wheeze, Crackle using machine learning techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3, pp. 1143–1152, 2023.

[17] 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, Oct. 2022, doi: 10.1007/s12652-021-03184-y.

[18] B. M. Rocha *et al.*, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, p. 035001, Mar. 2019, doi: 10.1088/1361-6579/ab03ea.

[19] B. Acar demirci, Y. Koçyiğit, D. Kizilirmak, and Y. Havlucu, "Adventitious and normal respiratory sound analysis with machine learning methods," *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*, vol. 18, no. 2, pp. 169–180, Jun. 2021, doi: 10.18466/cbayarfbe.1002917.

[20] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, no. 1, p. 4, Dec. 2020, doi: 10.1007/s13755-019-0091-3.

[21] C. Perlich, "Learning curves in machine learning," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 577–580.

[22] S. TAN, "Neighbor-weighted K-nearest neighbor for unbalanced text corpus," *Expert Systems with Applications*, vol. 28, no. 4, pp. 667–671, May 2005, doi: 10.1016/j.eswa.2004.12.023.

[23] Z. Wang and Z. Sun, "Performance evaluation of lung sounds classification using deep learning under variable parameters," *EURASIP Journal on Advances in Signal Processing*, vol. 2024, no. 1, p. 51, Apr. 2024, doi: 10.1186/s13634-024-01148-w.

[24] L. D. Mang, F. D. G. Martínez, D. M. Muñoz, S. G. Galán, and R. Cortina, "Classification of adventitious sounds combining cochleogram and vision transformers," *Sensors*, vol. 24, no. 2, p. 682, Jan. 2024, doi: 10.3390/s24020682.

[25] S.-Y. Jung, C.-H. Liao, Y.-S. Wu, S.-M. Yuan, and C.-T. Sun, "Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features," *Diagnostics*, vol. 11, no. 4, p. 732, Apr. 2021, doi: 10.3390/diagnostics11040732.

BIOGRAPHIES OF AUTHORS

Su Yuan Chang received her B.Eng. degree in biomedical electronic engineering from Universiti Malaysia Perlis, Malaysia in 2024. Currently, she is an engineer at AdoreSys Pte. Ltd, Penang, Malaysia. Her research interests include biomedical devices and systems, artificial intelligence, image and signal processing. She can be contacted at email: s201151370@studentmail.unimap.edu.my.

Marni Azira Markom holds a PhD in mechatronic engineering from Universiti Malaysia Perlis, Malaysia, in 2018, having previously graduated with a B.Eng (Computer Engineering) and an MSc (Computer Engineering) from the same university in 2006 and 2009, respectively. She is currently a senior lecturer at the Universiti Malaysia Perlis, Malaysia, and serves as a research fellow at the esteemed Centre of Excellence for Advanced Sensor Technology (CEASTech). Her expertise encompasses artificial intelligence systems and applications, medical signal and image processing, electronic and mechatronic systems. She has been elected as General Secreatary for Malaysia Society of Medical and Biological Enginering (MSMBE). She can be contacted at email: marni@unimap.edu.my.

Zhi Sheng Choong received his B.Eng. degree in biomedical electronic engineering from Universiti Malaysia Perlis, Malaysia in 2024. Currently, he is an engineer at SkyeChip Sdn Bhd, Penang, Malaysia. His research interests include biomedical devices and system, artificial intelligence, image and signal processing as well as electronic design. He can be contacted at email: s201151374@studentmail.unimap.edu.my.

Arni Munira Markom a senior lecturer at the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. She received her PhD in electronics (Photonics Engineering) from Universiti Malaya, Malaysia in 2016. She previously had a Master in microelectronics from Universiti Kebangsaan Malaysia and a Bachelor of Electronics (Computer Engineering) from Universiti Teknikal Malaysia Melaka, Malaysia. Her research areas are photonics technology, fiber lasers, fiber sensors, electrical engineering, and embedded systems including microcontrollers and IoT devices. She can be contacted at email: arnimunira@uitm.edu.my.

Latifah Munirah Kamarudin is an awarded Top Research Scientist in Malaysia with expertise in wireless sensing and artificial intelligence. She leads the Center of Excellence for Advanced Sensor Technology (CEASTech) and holds a position as associate professor at the Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis (UniMAP). In addition, she serves as a visiting professor at the University of Yamanashi, Japan. Her research interest includes signal and image processing, communication system, electronic system and artificial intelligence. She can be contacted at email: latifahmunirah@unimap.edu.my.

Erdy Sulino Mohd Muslim Tan is senior teaching engineer at the Faculty of Electrical Engineering Technology, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia. He received his MSc (Mechatronics engineering) and B.Eng (Mechatronics engineering) from Universiti Malaysia Perlis, Malaysia in 2008 and 2019, respectively. His research areas are artificial intelligence, agriculture technology, machine learning, signal processing, and embedded systems development. He can be contacted at email: erdysulino@unimap.edu.my.