Sensors and Materials, Vol. 38, No. 6 (2026) 3497–3508
MYU Tokyo

S & M 4516

Auscultation Support System for Chronic Obstructive Pulmonary Disease Prediction Using Convolutional Neural Network and Long Short-term Memory Models

Zong-Jie Wu,[1] Lun-Ping Hung,[2*] Hsiang-Tsung Yeh,[2] and Shu-Han Liao[3**]

[1]Department of Industrial Engineering and Management, National Yang Ming Chiao Tung University, No. 1001, Daxue Rd., East Dist., Hsinchu City 300093, Taiwan, R.O.C.
[2]Department of Information Management, National Taipei University of Nursing and Health Sciences, No. 365, Ming-te Rd., Peitou Dist., Taipei City 112303, Taiwan, R.O.C.
[3]Department of Electrical and Computer Engineering, Tamkang University, No. 151, Yingzhuan Rd., Danshuei Dist., New Taipei City 251301, Taiwan, R.O.C.

(Received August 21, 2025; accepted June 18, 2026)

Keywords: chronic obstructive pulmonary disease, Mel spectrum, Mel-frequency cepstral coefficients, convolutional neural network, long short-term memory

Chronic obstructive pulmonary disease (COPD) is a leading cause of global mortality, often remaining undetected until irreversible lung damage occurs. Leveraging advancements in AI and acoustic diagnostics, in this study, we compare the performance of deep learning models for COPD detection using respiratory sound data. Mel spectrogram and Mel-frequency cepstral coefficients were extracted from a publicly available dataset comprising crackle sounds from COPD patients and normal breath sounds. We evaluated standalone convolutional neural network (CNN) models (Residual Network, InceptionV3, and VGG16), a long short-term memory (LSTM) model, and hybrid CNN-LSTM and LSTM-CNN architectures. The LSTM outperformed standalone CNNs, achieving 94% accuracy, 93% precision, 99% recall, and an F1-score of 0.96, demonstrating its effectiveness in modeling temporal dependencies. The VGG16-LSTM achieved the highest performance, with 97.1% accuracy and 99% recall, highlighting the advantage of combining spatial feature extraction with temporal learning. However, several limitations should be acknowledged. The study relies on a single publicly available dataset, lacks real-world clinical validation, and adopts a binary classification framework that does not account for COPD severity staging. Future work will focus on multi-dataset validation, severity-graded classification, and the integration of edge-deployable models into wearable acoustic sensing platforms for real-time clinical application. These results underscore the potential of advanced deep learning models for accurate and accessible COPD diagnosis and support the development of next-generation acoustic sensors with enhanced sensitivity, improved signal-to-noise ratios, and integrated processing capabilities.

*Corresponding author: e-mail: lunping@ntunhs.edu.tw
**Corresponding author: e-mail: shliao@gms.tku.edu.tw
https://doi.org/10.18494/SAM5883

ISSN 0914-4935
https://myukk.org/

1. Introduction

Chronic obstructive pulmonary disease (COPD) is one of the top four causes of death worldwide and represents a significant public health challenge owing to its impact on the respiratory system.[1] Currently, the diagnosis of COPD depends on indicators such as respiratory sounds, breathing rate, respiratory efficiency, and pulmonary function tests.[2] However, these physiological changes are often difficult to perceive in daily life, and by the time significant symptoms appear, the lungs might already have sustained severe and irreversible damage.[3]

Recently, AI technology has been widely applied in healthcare, driven by rapid advancements in computing power, data availability, and machine learning (ML) algorithms. Since the outbreak of COVID-19 in 2019, research on the respiratory system has been extensively conducted. Most physicians use computed tomography and chest X-ray image analysis for COPD diagnosis.[4–6] Although imaging-based diagnostic techniques are available, healthcare institutions continue to face challenges such as data privacy concerns, information security issues, and inconsistent storage formats.[7] Therefore, acoustic diagnosis has been given much attention for respiratory disease diagnosis and treatment owing to its low cost and easy accessibility to equipment.

Acoustic sensor technology plays a crucial role in capturing subtle sound variations produced during respiration. These sensors, including highly sensitive microphones (condenser or piezoelectric microphones), are integrated into digital stethoscopes or wearable devices to detect abnormal lung sounds in normal breath sounds or speech characteristics. The advantages of acoustic diagnosis include its noninvasiveness, low cost, and ease of accessibility, making it a viable alternative to more expensive and less accessible imaging techniques. The development of AI models for acoustic COPD diagnosis significantly impacts the evolution of these sensors. AI algorithms, particularly deep learning algorithms, require large datasets of high-quality acoustic data to identify patterns indicative of COPD. This demand for reliable sound inputs ensures the development of sensors with enhanced sensitivity, improved signal-to-noise ratios (SNRs), and robust capabilities to filter out ambient noise and motion artifacts. Furthermore, AI's ability to analyze complex acoustic features, such as Mel-frequency cepstral coefficients (MFCCs), is closely related to sensor development to capture and transmit detailed sound characteristics with enhanced fidelity, ultimately enabling more accurate and automated diagnoses.[8–12]

MFCCs emulate the nonlinear frequency sensitivity of the human auditory system to extract perceptually relevant acoustic features.[13,14] Kim *et al.* integrated MFCCs with a convolutional neural network (CNN) to differentiate between normal and dysphagic patients using voice data. Their method showed an area under the curve of 95%, indicating the effectiveness of using acoustic features in disease diagnosis and classification.[15] Ye *et al.* applied a CNN with a long short-term memory (LSTM) model trained with MFCCs to diagnose post-stroke dysarthria. The hybrid model achieved 97.4% accuracy, outperforming standalone CNN or LSTM models.[16] Zhang *et al.*'s dual-channel CNN-LSTM model was used for lung disease classification with MFCCs as input. The model achieved an accuracy of 99.01%, emphasizing the value of hybrid deep learning models in respiratory sound analysis.[17]

COPD patients experience airway narrowing or mucus obstruction due to infections or lung tissue damage, resulting in characteristic pathological sounds such as crackles.[18] These features can be captured as diagnostic indicators using sensitive microphones. Current COPD sound

analysis mainly relies on traditional ML or standalone CNN models. For instance, Aykanat *et al.* extracted MFCCs from crackles, snoring, and normal speech and compared the performance characteristics of support vector machine (SVM) and CNN models, all of which achieved 80% accuracy.[19] Stasiakiewicz *et al.* used an SVM model to detect crackle sounds and identify different pulmonary diseases.[20] The model achieved 92.8% accuracy, demonstrating the effectiveness of crackle sound detection in disease diagnosis.[20]

In this study, we aim to construct a diagnostic model of COPD on the basis of a publicly available respiratory sound dataset. Mel spectrogram and MFCC features were extracted and input into standalone CNN and LSTM models and hybrid CNN-LSTM and LSTM-CNN models. The models' performance characteristics were compared to enhance the model's capability for COPD diagnosis.

By identifying which acoustic features and deep-learning models are effective for COPD diagnosis, requirements and targets for sensor development can be identified for the development of next-generation acoustic sensors with enhanced SNR to minimize ambient noise and capture specific features tailored to the characteristics of respiratory sounds. The smart sensors with embedded processing capabilities or standardized data output formats can be developed to facilitate direct data input into deep-learning models, reducing the complexity of data preparation.

2. Data, Materials, and Methods

We integrated CNN models and LSTM to develop a diagnosis method of COPD using auscultation. The developed methodology involves data preprocessing and the training of an acoustic model. In the preprocessing stage, five steps are conducted, namely, data screening, acoustic feature extraction, data augmentation, feature transformation, and dataset splitting. In the model training stage, deep learning models are trained and evaluated using metrics such as accuracy and recall. The outcomes of the evaluation provide evidence-based support for physicians in clinical decision-making processes. The overall process is illustrated in Fig. 1.

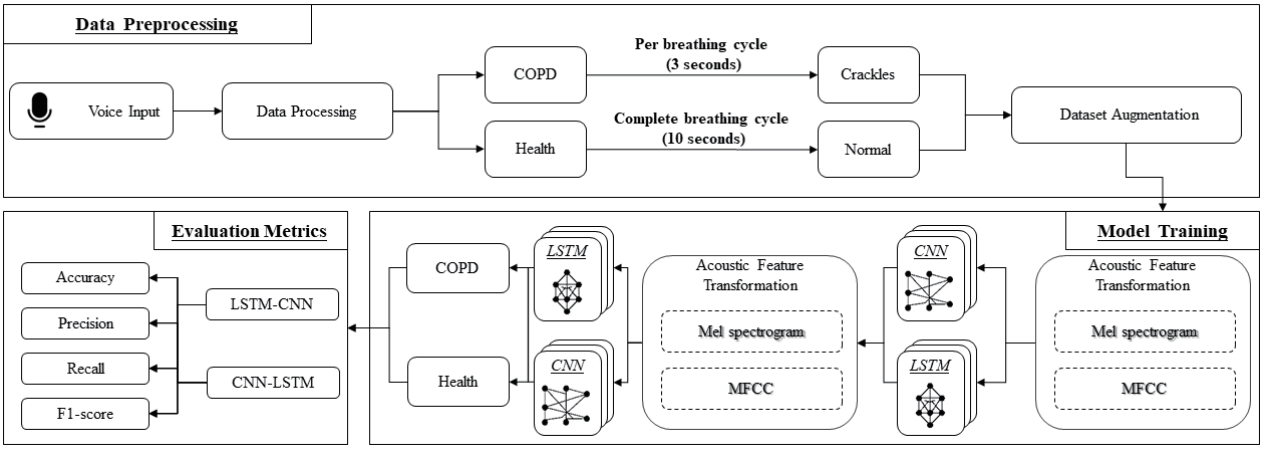


Fig. 1. COPD acoustic diagnosis process.

2.1 Data preprocessing

The dataset utilized in this study was provided by the International Conference on Biomedical and Health Informatics (ICBHI). It is a publicly available respiratory sound dataset that includes recordings from 126 participants. Pathological respiratory sounds were recorded from diverse anatomical regions, such as the anterior chest wall, back, lateral aspects of the thorax, and trachea.[21]

The sound samples were collected from 64 COPD patients and 26 healthy individuals. In total, 828 respiratory sound segments were obtained. During data preprocessing, two distinct datasets were constructed by extracting full respiratory cycles: one containing crackles from COPD patients and the other comprising normal breath sounds from healthy individuals. To address the challenges of limited data and class imbalance, audio time-stretching was employed for data augmentation, resulting in 1865 crackle samples and 1864 normal cycles.

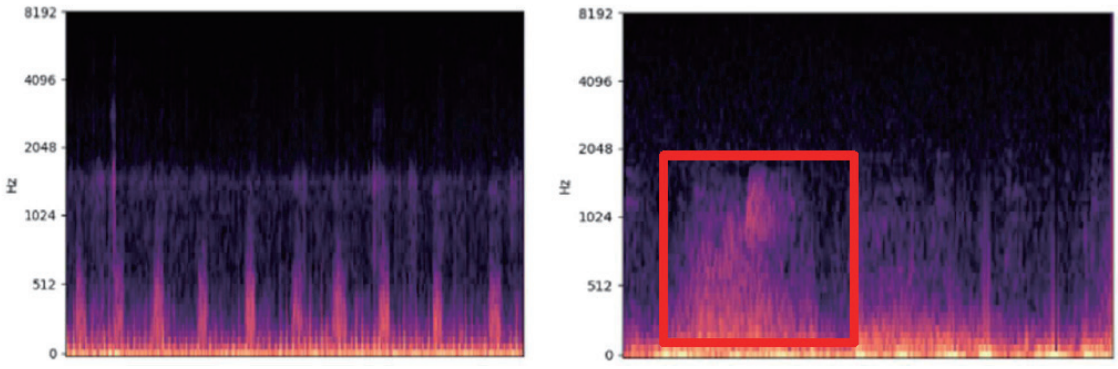
After augmentation, features were extracted from each respiratory cycle and converted into Mel spectrograms for CNN and LSTM models. The Mel spectrograms were drawn to illustrate the distribution of frequency over time. The spectrogram of a healthy subject [Fig. 2(a)] presents a regular frequency pattern, whereas that of a COPD case [Fig. 2(b)] shows less defined periodicity and a blurred region highlighted in the red box, which may be indicative of interference from crackles.

The MFCC feature extraction involved the following process:[22,23] The original speech signal $s(n)$ undergoes pre-emphasis using a high-pass filter to amplify high-frequency formant energies using Eq. (1).

$$y(n) = s(n) - \alpha s(n-1) \tag{1}$$

Here, $y(n)$ is the pre-emphasized signal and α is the pre-emphasis coefficient with $0 \le \alpha \le 1$.

Next, frame blocking was performed to divide the continuous speech signal into fixed-length short-time segments. In this study, the sampling rate was set to f_s = 22050 Hz, with a frame length of N = 2048 samples and a hop size of 512 samples to preserve temporal continuity. Each frame was multiplied by a Hamming window to reduce spectral leakage caused by boundary effects [Eq. (2)].



(a) (b)

Fig. 2. (Color online) Spectrograms below 4096 Hz of (a) healthy subject and (b) COPD case.

$$s_w(n) = \left[0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)\right] s(n),\ n = 0,1,2,\ldots,N-1. \tag{2}$$

Here, $s_w(n)$ is the windowed signal and w is the Hamming window at the n-th sample.

The windowed signal was transformed into the frequency domain using the fast Fourier transform (FFT) to obtain the spectrum $S(k)$, as shown in Eq. (3).

$$S(k) = \sum_{n=0}^{N-1} s_w(n) e^{-j2\pi kn/N},\ k = 0,1,\ldots,N-1. \tag{3}$$

Next, the spectrum was passed through a Mel-scale filter bank to simulate the nonlinear frequency perception of the human ear. The frequency-to-Mel scale was converted using Eq. (4).

$$f_{mel} = 2595\log_{10}\left(1 + \frac{f}{700}\right) \tag{4}$$

Each Mel filter outputs the logarithmic energy E_m, which is computed as

$$E_m = \log\left(\sum_{k=k_m-1}^{k_m+1} |S(k)|^2 H_m(k)\right),\ m = 1,2,\ldots,M. \tag{5}$$

Here, $H_m(k)$ is the response of the mth Mel filter and M is the total number of filters.

Subsequently, a discrete cosine transform (DCT) was applied to the Mel energies to extract MFCC c_n as.

$$c_n = \sum_{m=1}^{M} E_m \cos\left(\frac{\pi n}{M}(m - 0.5)\right),\ n = 1,2,\ldots,N_c. \tag{6}$$

Here, N_c is the number of cepstral coefficients. To capture the temporal dynamics of the sound features, first- and second-order delta coefficients were also computed, forming an MFCC feature set. To construct and evaluate the models, the dataset was split into sub-datasets for training and validation, and testing in a ratio of 9:1. The training-validation sub-dataset was split into training and validation sets in a ratio of 8:2. The testing set was reserved for performance evaluation.

2.2 Acoustic model training

In model design, we implemented a reciprocal prediction mechanism. Spatial features were extracted by three CNN models, namely, a residual network with 50 layers (ResNet50),[24] InceptionV3,[25] and a visual geometry group 16-layer network (VGG16).[26] The extracted

features were processed by LSTM.[27] Conversely, the features extracted by LSTM were processed by the CNN models. This integration of CNN-based models and LSTM yielded the synergistic effects of CNN and LSTM models. By combining CNN and LSTM architectures, the developed system in this study captured the spatial and temporal characteristics of respiratory sound signals, and improved the classification accuracy and stability for COPD detection. The models' performance characteristics were evaluated using precision, recall, and F1-score.

2.3 Performance evaluation metrics

In this study, we employed several widely used classification performance metrics to evaluate the effectiveness of the proposed ML models, including accuracy, precision, recall, and F1-score.

As shown in Table 1, a confusion matrix was first constructed for each classification model to show the numbers of true positives (*TP*), false positives (*FP*), true negatives (*TN*), and false negatives (*FN*). These values served as the basis for calculating the subsequent evaluation metrics. Accuracy is defined as the proportion of correctly classified samples among all samples, reflecting the overall predictive correctness of the model. Precision represents the proportion of true positive instances among all the instances predicted as positive, indicating the reliability of positive predictions. Recall is defined as the proportion of correctly identified positive instances among all actual positive instances, reflecting the model's ability to detect the positive class. The F1-score, calculated as the harmonic mean of precision and recall, provides a composite measure that balances these two metrics. It is particularly useful in class-imbalanced scenarios or when both detection performance and the risk of misclassification are important considerations.

3. Results

The ResNet50 model achieved an accuracy of 80%, a precision of 84%, a recall of 86%, and an F1-score of 0.85. The VGG16 model reached an accuracy of 81%, a precision of 86%, a recall of 85%, and an F1-score of 0.85. Although the InceptionV3 model achieved the highest accuracy of 84%, its precision, recall, and F1-score were 72%, 69%, and 0.71, respectively, which were lower than those of the ResNet50 and VGG16 models. This indicated that the InceptionV3 model showed a higher rate of misclassification than the other CNN models.

Table 1
Confusion matrix for the binary classification mode.

		Predicted class		
		Positive	Negative	
Actual class	Positive	*TP*	*FN*	Sensitivity $\frac{TP}{TP+FN}$
	Negative	*FP*	*TN*	Specificity $\frac{TN}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	Negative predictive value $\frac{TN}{TN+FN}$	Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$

In contrast, the LSTM model outperformed the CNN models, with an accuracy of 94%, a precision of 93%, a recall of 99%, and an F1-score of 0.96, as shown in Table 2. This demonstrated the model's robust and effective classification ability for sequential respiratory sound data.

The performance of the three CNN models combined with LSTM is show in Table 3. The VGG16-LSTM yielded the best performance in COPD diagnosis, achieving an accuracy of 97.1% and a recall of 99%. While its precision was slightly lower than that of the InceptionV3-LSTM, it outperformed in the other metrics.

The results of this study demonstrate the potential of deep learning models for the accurate diagnosis of COPD using respiratory sound data. The analysis of the results of standalone CNN models and LSTM, as hybrid combinations, provides valuable information for the application of deep-learning models.

Among the standalone models, LSTM outperformed the CNN models, which underscores the LSTM's capability to effectively capture temporal dependencies and sequential patterns from respiratory sound data. Since respiratory sounds are time-series data with dynamic variations indicative of physiological states, the ability of LSTM to process and learn from sequential information is critical. In contrast, while the ResNet50 and VGG16 models showed comparable performance with the hybrid models, the InceptionV3 model demonstrated a lower recall and F1-score, suggesting a higher rate of misclassification, especially concerning true positive identification. While CNN models are effective at extracting spatial features, their application might be less robust for sequential respiratory sound analysis than LSTM.

The integration of CNN models and LSTM enhanced diagnostic performance. The VGG16-LSTM model performed the best, indicating that the spatial feature extraction capabilities of CNN models are enhanced with the temporal learning strengths of LSTM. This provides an accurate understanding of respiratory sound characteristics for COPD diagnosis. The high recall achieved by the VGG16-LSTM model is significant in a clinical context, as it implies a very low

Table 2
Performance metrics of CNN and LSTM.

Model		Accuracy	Precision	Recall	F1-score
CNN	ResNet50	0.80	0.84	0.86	0.85
	InceptionV3	0.84	0.72	0.69	0.71
	VGG16	0.81	0.86	0.85	0.85
LSTM	LSTM	**0.94**	**0.93**	**0.99**	**0.96**

Table 3
Performance metrics of combined CNN–LSTM and LSTM–CNN models.

Model		Accuracy	Precision	Recall	F1-score
CNN-LSTM	ResNet50-LSTM	0.964	0.972	0.986	0.979
	InceptionV3-LSTM	0.968	0.977	0.986	0.982
	VGG16-LSTM	0.971	0.976	0.990	0.983
LSTM-CNN	LSTM- ResNet50	0.846	0.972	0.859	0.912
	LSTM- InceptionV3	0.877	0.977	0.888	0.930
	LSTM- VGG16	0.842	0.976	0.851	0.909

rate of false negatives for accurate early detection and timely intervention in COPD treatment. While other hybrid models also showed strong performance, the better metrics of the VGG16-LSTM suggest that the features learned by the VGG16 model and fed into LSTM for temporal analysis yield the most effective diagnostic capability.

The high performance of deep-learning models, particularly hybrid models, contributes to the development of next-generation acoustic sensors optimized for high-fidelity, continuous, and sequential data capture. First, the reliance on features, such as Mel spectrograms, necessitates sensors with enhanced sensitivity and a wider frequency response across the entire range of human respiratory sounds (often optimized for roughly 100–1000 Hz), since subtle acoustic markers of COPD are often missed by conventional sensors (e.g., approximately 50–2000 Hz, depending on the target acoustic markers). Second, the performance of deep learning models largely depends on clean data. Therefore, sensors with superior SNR and improved ambient noise cancellation capabilities are required to minimize interference from the environment and patient movement, allowing the models to accurately analyze physiologically relevant acoustic patterns.

The sequential nature of respiratory sound analysis and the high processing demands of LSTM and hybrid models require acoustic sensors with enhanced performance. This leads to the development of smart acoustic sensors with embedded edge computing capabilities for preliminary feature extraction. This reduces data transmission bandwidth requirements and enables near real-time diagnosis. The validation of robust deep-learning models for COPD diagnosis provides a reference for effective sensor calibration and validation. Sensor manufacturers need to design and test sensors considering the data characteristics required for optimal performance with advanced algorithms, accelerating the translation of research results into clinically viable diagnostic tools.

4. Discussion

Previous studies have demonstrated the effectiveness of standalone CNN[19] and SVM-based[20] methods for respiratory sound classification, while hybrid CNN–LSTM[17] architectures have also been applied to multi-class lung disease detection. Despite these advances, several methodological and evaluation-related limitations persist in the existing literature.

A notable limitation lies in the predominant focus on a unidirectional hybrid architecture. Prior work has largely confined its investigation to the CNN–LSTM configuration, without examining whether an inverse arrangement, namely, LSTM–CNN, yields a comparable or systematically different performance. The absence of such bidirectional architectural analysis restricts a more complete understanding of how temporal and spatial feature extraction interact within sequential deep learning frameworks.

In parallel, the performance improvements reported in the literature are frequently associated with auxiliary enhancement strategies, including data augmentation and multi-modal feature integration based on the generative adversarial network. While effective, these approaches introduce additional layers of complexity that may obscure the intrinsic contribution of the

model architecture itself. Consequently, the extent to which performance gains can be attributed to architectural design remains insufficiently isolated.

From an evaluation perspective, the emphasis on overall accuracy in multi-class classification settings presents further limitations. Aggregate performance metrics may mask clinically relevant disparities, particularly in relation to COPD, where recall is of primary importance for early detection. This misalignment between evaluation practices and clinical priorities underscores the need for more targeted performance assessment.

In response to these limitations, in this study, we introduce a structured investigation that disentangles architectural, feature, and clinical evaluation factors within a unified framework. A reciprocal prediction paradigm is employed to systematically compare CNN–LSTM and LSTM–CNN configurations across multiple backbone architectures, enabling the direct assessment of architectural ordering effects. Model performance is further examined under controlled conditions using dual acoustic feature representations, specifically Mel spectrograms and MFCCs, within a binary COPD classification task. Finally, the findings are extended beyond model-level evaluation to guide the design considerations of next-generation acoustic sensing systems, establishing a linkage between algorithmic requirements and sensor-level specifications.

5. Conclusions

We validated the effectiveness of deep learning models, particularly those incorporating LSTM, for the accurate diagnosis of COPD using respiratory sound data. This study is positioned within the domain of ML techniques, combining signal-level feature engineering with advanced deep learning architectures. Feature extraction methods, including MFCCs and Mel spectrograms, were used to transform raw respiratory signals into structured representations. Among the evaluated models, the standalone LSTM demonstrated strong capability in capturing temporal dependencies, while the hybrid VGG16-LSTM model achieved superior diagnostic performance, with an accuracy of 97.1% and a recall of 99%. This high recall is particularly important for minimizing false negatives and enabling early clinical intervention.

These findings highlight the effectiveness of supervised deep learning architectures, including CNNs for spatial feature extraction and LSTM networks for temporal modeling, as state-of-the-art approaches for time-series classification. The integration of hybrid CNN-LSTM frameworks further enables the complementary learning of spatial and temporal acoustic patterns. Combined with data augmentation techniques and rigorous evaluation using standard metrics (accuracy, precision, recall, and F1-score), we demonstrate the potential of ML-driven approaches in supporting noninvasive clinical decision-making for COPD diagnosis.

The demonstrated reliance of high-performing models on detailed acoustic features necessitates the development of advanced sensing systems with enhanced sensitivity, flat frequency response, and high SNRs. In this context, the development of an intelligent stethoscope integrating a high-sensitivity acoustic sensor and an edge computing module capable of executing a VGG16-LSTM model is technically feasible with current technologies. However, practical implementation requires that the sensor supports a frequency range of 20–

4000 Hz with an SNR of at least 70 dB, while the computational demands of VGG16 must be addressed through model optimization techniques such as pruning and INT8 quantization, enabling significant model size reduction and deployment on resource-constrained edge devices. In addition, efficient wireless communication and real-time signal processing are essential to ensure seamless data transmission and inference within clinical workflows.

Nevertheless, this study has several limitations that should be acknowledged. First, the models were trained and evaluated solely on the ICBHI publicly available dataset, which may limit the generalizability of findings across different patient populations, recording environments, and clinical settings. Second, the binary classification framework (COPD vs healthy) does not address COPD severity staging (e.g., GOLD grades I–IV), which is essential for clinical management and treatment planning. Third, the current models have not been validated in real-time clinical deployment scenarios, leaving a gap between laboratory performance and bedside applicability. Future studies will aim to address these limitations by (1) expanding validation to multi-center, multi-dataset cohorts to improve model robustness and generalizability, (2) developing multi-class severity classification models to support COPD staging and disease progression monitoring, and (3) integrating the optimized deep learning pipeline into wearable acoustic sensing platforms with edge computing capabilities for real-time, point-of-care COPD screening.

Acknowledgments

This research was supported by the National Science and Technology Council of Taiwan under grant no. NSTC 112-2221-E-227-002 -MY3.

Author contributions

Zong-Jie Wu: Formal analysis, Validation, Software. Lun-Ping Hung: Supervision, Visualization, Methodology, Project administration, Validation. Hsiang-Tsung Yeh: Software, Writing - original draft, Writing - review & editing. Shu-Han Liao: Writing - original draft, Writing - review & editing.

References

1 World Health Organization, The Top 10 Causes of Death: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed August 2025).
2 K. L. Bailey: Med. Clin. North Am. **96** (2012) 745. https://doi.org/10.1016/j.mcna.2012.04.011
3 D. Singh, M. Miravitlles, and C. Vogelmeier: Adv. Therapy **34** (2017) 281. https://doi.org/10.1007/s12325-016-0459-6
4 D. A. Lynch: Br. J. Radiol. **95** (2022) 20201005. https://doi.org/10.1259/bjr.20201005
5 M. H. Al-Sheikh, O. Al. Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim: Sci. Rep. **13** (2023) 19373. https://doi.org/10.1038/s41598-023-46147-3
6 G. R. Washko: Semin. Respir. Crit. Care Med. **31** (2010) 276. https://doi.org/10.1055/s-0030-1254068
7 C. S. Kruse, R. Goswamy, Y. J. Raval, and S. Marawi: JMIR Med. Inf. **4** (2016) e38. https://doi.org/10.2196/medinform.5359
8 A. H. Sfayyih, A. H. Sabry, S. M. Jameel, N. Sulaiman, S. M. Raafat, A. J. Humaidi, and Y. M. A. Kubaiaisi: Diagnostics **13** (2023) 1748. https://doi.org/10.3390/diagnostics13101748

9 R. N. Saunders, X. G. Tan, S. M. Qidwai, and A. Bagchi: Ann. Biomed. Eng. **47** (2019) 2005. https://doi.org/10.1007/s10439-018-02157-1
10 P. Kapetanidis, F. Kalioras, C. Tsakonas, P. Tzamalis, G. Kontogiannis, T. Karamanidou, T. G. Stavropoulos, and S. Nikoletseas: Sensors **24** (2024) 1173. https://doi.org/10.3390/s24041173
11 F. Kong, Y. Zou, Z. Li, and Y. Deng: Sensors **24** (2024) 5354. https://doi.org/10.3390/s24165354
12 A. Rao, E. Huynh, T. J. Royston, A. Kornblith, and S. Roy: IEEE Rev. Biomed. Eng. **12** (2019) 221. https://doi.org/10.1109/RBME.2018.2874353
13 K. Nishikawa, K. Akihiro, R. Hirakawa, H. Kawano, and Y. Nakatoh: Cognit. Rob. **2** (2022) 21. https://doi.org/10.1016/j.cogr.2021.12.003
14 S. Davis and P. Mermelstein: IEEE Trans. Acoust. Speech Signal Process. **28** (1980) 357. https://doi.org/10.1109/TASSP.1980.1163420
15 H. Kim, H.-Y. Park, D. Park, S. Im, and S. Lee: Biomed. Signal Process. Control. **86** (2023) 105259. https://doi.org/10.1016/j.bspc.2023.105259
16 W. Ye, Z. Jiang, Q. Li, Y. Liu, and Z. Mou: Appl. Acoust. **197** (2022) 108934. https://doi.org/10.1016/j.apacoust.2022.108934
17 Y. Zhang, Q. Huang, W. Sun, F. Chen, D. Lin, and F. Chen: Biomed. Signal Process. Control **94** (2024) 106257. https://doi.org/10.1016/j.bspc.2024.106257
18 J. C. Hogg: The Lancet **364** (2004) 709. https://doi.org/10.1016/S0140-6736(04)16900-6
19 M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal: EURASIP J. Image Video Process. **2017** (2017) 65. https://doi.org/10.1186/s13640-017-0213-2
20 P. Stasiakiewicz, A. P. Dobrowolski, T. Targowski, N. Gałązka-Świderek, T. Sadura-Sieklucka, K. Majka, A. Skoczylas, W. Lejkowski, and R. Olszewski: Biomed. Signal Process. Control. **67** (2021) 102521. https://doi.org/10.1016/j.bspc.2021.102521
21 B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. P. Paiva, I. Chouvarda, and P. de Carvalho: Physiol. Meas. **40** (2019) 035001. https://doi.org/10.1088/1361-6579/ab03ea
22 A. Benba, A. Jilbab, and A. Hammouch: Int. J. Speech Technol. **19** (2016) 449. https://doi.org/10.1007/s10772-016-9338-4
23 A. Ilapakurti, S. Kedari, J. S. Vuppalapati, S. Kedari, and C. Vuppalapati: 2019 IEEE 5th Int. Conf. Big Data Computing Service and Applications (BigDataService) (IEEE, 2019) 340–345. https://doi.org/10.1109/BigDataService.2019.00060
24 K. He, X. Zhang, S. Ren, and J. Sun: 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (IEEE, 2016) 770–778. https://doi.org/10.1109/CVPR.2016.90
25 C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna: 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (IEEE, 2016) 2818–2826. https://doi.org/10.1109/CVPR.2016.308
26 K. Simonyan and A. Zisserman: arXiv:1409.1556 (2015). https://arxiv.org/abs/1409.1556
27 S. Hochreiter and J. Schmidhuber: Neural Comput. **9** (1997) 1735. https://doi.org/10.1162/neco.1997.9.8.1735

About the Authors

Zong-Jie Wu received his B.S. degree in 2019 and M.S. degree in 2021 from National Taipei University of Nursing and Health Sciences. Sinc e 2021, he has been pursuing a Ph.D. degree in the Department of Industrial Engineering and Management at National Yang Ming Chiao Tung University, where he is currently a fourth-year doctoral student. His research interests include IoT, mobile computing, healthcare technologies, data mining, networking, AI, and machine learning. (zongjie13@gmail.com)

Lun-Ping Hung received his Ph.D. degree from Tamkang University in 2003. Since 2009, he has been working for National Taipei University of Nursing and Health Sciences, where he currently serves as a Distinguished Professor. His research interests include IoT applications in healthcare, mobile health applications, wireless sensor networks, recommendation systems, medical informatics, and intelligent learning systems. (lunping@ntunhs.edu.tw)

Hsiang-Tsung Yeh is a master's degree student at National Taipei University of Nursing and Health Sciences. His research interests include IoT, mobile computing, healthcare, data mining, AI, and machine learning. (vul3ejo3805@gmail.com)

Shu-Han Liao received his Ph.D. degree from Tamkang University in 2013. Since 2022, he has been an assistant professor at Tamkang University. His research interests include wireless communication systems, IoT applications, 5G/Beyond 5G (B5G), next-generation communication technologies (6G), smart home technologies, emerging innovative technologies, smart healthcare, and ESG-oriented sustainable development. (shliao@gms.tku.edu.tw)