Applied Mathematics and Computational Intelligence
Volume 15, No. 3, 2026 [142-158]

A Machine Learning Framework for Multi-Class Lung Disease Classification Using Lung Sounds

Wan Nurul Aiffah Ismail[1*], Najah Ghazali[2], Ruzelita Ngadiran[3]

[1,2,3]Faculty of Intelligent Computing, Universiti Malaysia Perlis, Kampus Alam UniMAP, Pauh Putra, 02600 Arau, Perlis, Malaysia

[*] Corresponding author : aiffah@unimap.edu.my

Received: 20 November 2025
Revised: 6 August 2026
Accepted: 10 August 2026

ABSTRACT

Diagnostic ambiguity between chronic pulmonary diseases like asthma and Chronic Obstructive Pulmonary Disease (COPD) presents significant diagnostic challenges due to their overlapping clinical symptoms. This overlap often leads to misdiagnosis and increased mortality, highlighting the urgent need for effective diagnostic methods. To address this issue, this study proposes a computerized lung sound-based method. The methodology is evaluated using the publicly available ICBHI and KAUH lung sound datasets. Lung sound denoising is performed using a Butterworth Bandpass Filter. A combination of acoustic features such as Zero-Crossing Rate (ZCR), Mel spectrogram, and Mel-Frequency Cepstral Coefficients (MFCC) are extracted, followed by Z-score normalization to ensure zero mean and unit variance. Four machine learning classifiers, namely Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and Multi-Layer Perceptron (MLP), are compared. Additionally, the system implements hyperparameter tuning strategies, such as Grid Search (GS), to distinguish among healthy, asthma, COPD, and other lung disease cases. After hyperparameter tuning, the RF classifier achieved the highest validation accuracy of 96.72% and a test accuracy of 96.59%, while k-NN achieved the largest performance gain, improving accuracy by 4.1% compared to its baseline. The high accuracy of the proposed framework, which employs a computationally efficient feature set, highlights its potential as a reliable decision-support tool for lung disease classification, offering considerable utility in resource-constrained clinical settings.

Keywords: lung disease classification, acoustic feature extraction, machine learning, hyperparameter tuning.

1 INTRODUCTION

Respiratory diseases constitute a major global health burden and rank among the leading causes of morbidity and mortality worldwide. Among these, asthma and chronic obstructive pulmonary disease (COPD) are the two most prevalent chronic respiratory illnesses with significant impact on quality of life and healthcare systems. Early and accurate diagnosis is imperative to initiate timely management and improve patient outcomes. The traditional auscultatory examination using stethoscopes, though widespread, is subjective and highly dependent on the clinician's expertise.

Despite their distinct pathophysiological mechanisms, asthma and COPD often share overlapping clinical symptoms such as chronic cough, wheezing, and dyspnea, which complicate differential diagnosis and leads to frequent misclassification in clinical practice [1]. Misdiagnosis not only delays appropriate treatment but may also increase the risk of exacerbation, and hospitalization.

Technological advances in digital stethophones and signal processing have enabled high-fidelity lung sound acquisition and digital storage, paving the way for computational analysis to augment clinical decision-making. Digital lung sound analysis provides an objective, non-invasive, and reproducible approach for respiratory disease screening and monitoring, particularly for asthma and COPD [1], [2]. It facilitates automated detection of pathological lung sounds, potentially supporting earlier diagnosis and intervention, which are pivotal in preventing disease progression and reducing exacerbations [1], [3].

Advances in digital auscultation, signal processing, and machine learning offer promising avenues for objective lung sound analysis, enabling more robust disease classification. Lung sounds provide non-invasive biomarkers for respiratory disease identification, recognizing adventitious sounds like wheezes and crackles characteristic of obstructive airway diseases [4]. Crucially, these approaches encompass meticulously designed feature extraction pipelines, robust classification frameworks, and data augmentation strategies to address data scarcity and class imbalance inherent in lung sound datasets. Mel-Frequency Cepstral Coefficients (MFCCs) remain the most widely adopted feature extraction method due to their effectiveness in capturing the perceptual characteristics of lung sounds [2]. Similarly, Tunable Q-factor Wavelet Transform (TQWT)-based feature extraction combined with dense wavelet packet energies significantly improved discrimination between healthy and diseased lung sounds [5]. A significant challenge in this field is data scarcity and class imbalance, often addressed through data augmentation strategies [6], [7].

Several works focus exclusively on asthma and COPD classification through lung sound analysis. Şen et al. (2021) employed multivariate pulmonary sound analysis based on a Bayesian framework and Linear SVM, achieving up to 98% accuracy in differentiating asthma from COPD. Their findings also highlighted the importance of specific respiratory cycle phases for disease discrimination [1]. Khanaghavalle et al. (2024) developed a COPD severity classification system based on spectrogram and chromogram features, trained on deep residual networks (ResNet50), indicating promising early-stage diagnosis potential [3].

However, the performance of these models is highly sensitive to their hyperparameter configurations. Hyperparameter tuning is the process of systematically searching for the most appropriate configuration settings of a machine learning model such as learning rate, number of layers, or regularization strength in order to maximize its performance. The importance of hyperparameter tuning in lung disease classification lies in its ability to improve predictive performance, reduce diagnostic errors, and enhance model adaptability to complex medical data. Furthermore, tuning reduces false positives and false negatives, which is crucial in medical diagnosis where misclassification can lead to serious consequences [8], [9], [10]. Recent research highlights multiple applications of hyperparameter tuning in lung disease classification. For SVM models, tuning the regularization parameter and gamma parameters through random Grid Search has produced notable gains in classification accuracy, precision, and specificity [8]. While Bayesian optimization provides a more efficient search strategy by modeling the objective function [11], [12]. More advanced approaches, including metaheuristic techniques such as Harmony Search, Grey Wolf Optimization, Dung Beetle Optimization, and Honey Badger Optimization, as well as surrogateassisted evolutionary algorithms, have been increasingly employed for deep learning models to handle high-dimensional optimization challenges [9], [13], [14]. Studies demonstrate that hyperparameter tuning can dramatically increase model accuracy, precision, recall, F1-score, and specificity.

Despite this, a direct and systematic comparison of well-tuned, standard machine learning classifiers using a comprehensive set of acoustic features on a common framework is lacking. Therefore, this paper aims to:

- Identify current challenges including noise interference, and data imbalance
- Highlight prominent feature extraction techniques, especially those leveraging time-frequency.
- Examine state-of-the-art machine learning architectures tailored to respiratory disease classification.
- To systematically examine and quantify the impact of hyperparameter tuning using Grid Search on the performance of different classifiers for distinguishing between asthma, COPD, and healthy subjects.

Experimental results demonstrate that the proposed framework achieves high accuracy and robustness in distinguishing between asthma, COPD, healthy individuals, and other lung disease cases. By incorporating hyperparameter tuning strategies, the system outperforms baseline models, underscoring its potential as a reliable decision-support tool for clinical applications. The rest of the paper is organized as follows: Section 2 describes the datasets, preprocessing techniques, feature extraction methods and four classifiers, along with the hyperparameter tuning strategy. Section 3 reports the experimental setup and results, followed by a discussion. Finally, Section 4 concludes the paper and outlines directions for future research.

2 METHODOLOGY

Figure 1 presents the conceptual framework for automated lung sound classification which integrate signal pre-processing, feature extraction, and machine learning classifiers to identify respiratory diseases. The pre-processing stage was carried out in three sequential steps to ensure the quality of the lung sound recordings. First, all recordings were resampled to 4kHz sampling rate to ensure consistency across the dataset. Second, a 5th-order Butterworth band-pass filter was applied to suppress background noise and preserve the primary frequency components associated with lung sounds, specifically within the 100–2000 Hz range. Finally, the filtered signals were segmented into fixed intervals of 5 seconds. If the duration of a recording was shorter than the designated window, zero padding was employed to extend the signal to the required length. Next, feature extraction is performed to obtain relevant characteristics from the lung signal, which are then normalized to ensure consistency in scale and enhance classifier performance. Various supervised learning algorithms namely Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and k-Nearest Neighbors (k-NN) were employed to classify lung conditions. To optimize their performance, Grid Search (GS) was applied for systematic hyperparameter tuning, allowing the selection of the most suitable parameter configurations such as the number of estimators in RF, the regularization parameter and gamma values in SVM, regularization strength in MLP, and the number of neighbors in k-NN. As a result, the models achieved improvements in accuracy, precision, recall, and reliability in classifying COPD, asthma, healthy subjects, and other lung diseases.

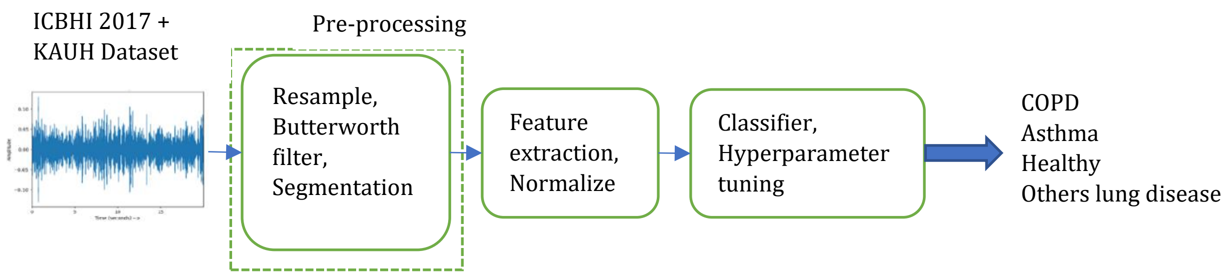

Figure 1: Automated Lung Sound Analysis Framework

2.1 Data Acquisition and Datasets

This automated respiratory disease classification research utilizes two publicly available datasets such as the International Conference on Biomedical and Health Informatics (ICBHI) lung sound database and King Abdullah University Hospital (KAUH) datasets [15], [16]. These databases contain a wide spectrum of lung sounds recorded under varying conditions, including crackles, wheezes, normal breath sounds, and combinations of them. The ICBHI and KAUH datasets differ in their sampling rate frequencies and recording length as shown in Table 1. To ensure consistency in ICBHI dataset analysis, researchers commonly resample all recordings to 4 kHz during preprocessing [17], [18], [19]. Unlike ICBHI, all signals in KAUH are consistently acquired at 4 kHz rate and are segmented into shorter frames for subsequent analysis [16], [17]. A lower sampling frequency, such as that in ICBHI, is sufficient to capture the fundamental frequency range of lung sounds (100–2000 Hz) but provides relatively limited detail in higher frequency components. In contrast, the higher sampling frequency of KAUH ensures enhance resolution and preserves more detailed acoustic information, including characteristics of wheezes and crackles. Generally, raw lung sounds often suffer contamination from ambient noise, cardiac sounds, and patient movement artifacts. Therefore, signal denoising is a critical process, typically achieved by employing filters such as the Butterworth Bandpass to retain relevant respiratory frequencies while removing out-of-band noise [20].

The ICBHI dataset provides longer recordings, often containing multiple respiratory cycles, whereas the KAUH dataset provides shorter recordings. The ICBHI dataset was initially divided into annotated intervals that specify the presence of distinct lung sounds, including crackles, wheezes, a combination of these adventitious sounds, or normal lung sounds. These annotated segments provide precise temporal that facilitate the identification of clinically relevant acoustic patterns. Next, the ICBHI recordings were subsequently segmented into fixed 5 seconds length windows. KAUH datasets also are frequently segmented into shorter frames during preprocessing. This process is important to standardize analysis and improve computational efficiency. A 5-second window is generally sufficient to capture at least one full breathing cycle, ensuring that clinically relevant information is preserved [21].

In this research, the two datasets are combined to form a more efficient and comprehensive dataset, capturing a broader spectrum of lung sound characteristics. This technique not only increases data diversity and reduces potential bias from individual datasets but also enhances the generalizability and robustness of the classification model across varied clinical scenarios.

Table 1: Key Features of ICBHI and KAUH Datasets

Dataset	Subject	Recording	Sample rate (kHz)	Individual recording lengths (second)	Distribution of Classes
ICBHI 2017	126	920	44.1, 4	10 - 90	Healthy: 35, Asthma: 1, COPD: 793, Others: 95
KAUH	112	336	4	5 – 30	Healthy: 105, Asthma: 99, COPD: 27, Others: 105

2.2 Feature Extraction

Mel-frequency cepstral coefficients (MFCCs) are extracted from denoised signals due to their robustness in capturing lung sound characteristics. MFCCs are particularly effective in capturing perceptually relevant spectral properties of lung sounds and have been widely employed in lung sound analysis [2], [22], [23], [24], [25]. ZCR, which quantifies the frequency at which a signal changes sign, provides valuable information about the periodicity and noisiness of the signal, and when combined with MFCC, can improve feature discrimination [26]. Similarly, mel spectrograms provide a time–frequency representation aligned with human auditory perception, making them effective for identifying abnormal lung sounds. This combination of multiple auditory features has been shown to significantly enhance classification accuracy.

Z-score normalization, which standardizes data to zero mean and unit variance, is widely applied in lung sound classification to ensure comparability across features, reduce bias, and improve model convergence [26], [27]. Normalization using Z-score standardization ensures input features have zero mean and unit variance, reducing bias during model learning. This transformation of normalized value, is mathematically expressed as Equation 1:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents the feature value, μ and σ denote the sample mean and standard deviation respectively. This approach is particularly important when combining heterogeneous features, which exist on different numerical scales. Without normalization, features with larger values may dominate the classifier, leading to suboptimal learning outcomes [26], [27]. Several studies have demonstrated that stacking MFCC, ZCR, and mel-spectrogram features, followed by z-score normalization, significantly enhances classification accuracy and robustness in both machine learning and deep learning models [2], [28], [29], [30]. Thus, feature stacking combined with z-score normalization represents a validated and effective strategy for improving lung sound classification performance across diverse modelling approaches.

2.3 Hyperparameter Tuning

After feature extraction and normalization, the dataset was randomly divided into three mutually exclusive subsets comprising 70% training, 15% validation, and 15% testing using a stratified data splitting strategy. Since the framework was developed using segmented lung sound samples, the partitioning was performed at the segment level, ensuring that feature representations generated from individual lung sound segments were assigned exclusively to a single subset. This approach minimized the risk of data leakage during model development. A fixed random seed was applied to ensure reproducibility of the experimental results. The normalized feature set was subsequently classified using SVM, RF, k-NN, and MLP, with hyperparameters optimized through Grid Search combined with 5-fold cross-validation performed exclusively on the training subset. During this process, the training data were iteratively divided into $k-1$folds for model training and one-fold for internal validation, allowing the optimal hyperparameter configuration to be selected based on the average cross-validation performance. The best estimator obtained from Grid Search was then evaluated on the independent validation and test subsets, where the validation results provided an intermediate performance assessment and the test results represented the final estimation of model generalization capability.

The key hyperparameters optimized for each classifier are summarized in Table 2, including the number of estimators and tree depth for RF, the regularization parameter $C$and kernel settings for SVM, the number of neighbors and distance metrics for k-NN, and the hidden layer sizes and learning rates for MLP. Based on the Grid Search results, the optimal hyperparameter configurations were selected according to the highest average cross-validation performance. Subsequently, the final models were retrained using the complete training subset and evaluated on the independent validation and test sets to assess their generalization capability.

Model performance is assessed using various evaluation metrics, including accuracy, balanced accuracy, per-class sensitivity or recall, per-class specificity, macro F1-score or average F1-score per-class, and confusion matrices. These metrics provide a comprehensive evaluation of overall correctness, class-wise performance, false positive and false negative trade-offs, and model robustness. Employing appropriate dataset split strategies and cross-validation safeguards against overfitting and ensures generalizability. Additionally, the total tuning time was recorded to assess the computational cost of GS for each classifier.

Table 2: Optimized Hyperparameters Across Supervised Learning Models.

Classifier	**Hyperparameter Tune**	**Range**
RF	Number of trees	[100, 200, 300]
	Maximum Depth of Trees	[10, 20, None]
	Minimum Samples to Split a Node	[2, 5, 10]
	Minimum Samples per Leaf Node	[1, 2, 4]
	Bootstrap Sampling	[True, False]
SVM	Regularization Parameter, C	[0.1, 1, 10]
	Kernel	['linear', 'rbf']
	Gamma	['scale', 'auto']
	Class weight	['balanced', None]

MLP	Number of layers and neurons at hidden layer Activation function Optimizer algorithm Starting learning rate	[(64,), (128,64), (128,64,32)] (relu, tanh) (adam, sgd) (0.001, 0.01)
k-NN	Number of nearest neighbors Weights Distance metric	[3, 5, 7, 9, 11] ['uniform', 'distance'] ['euclidean', 'manhattan', 'minkowski']

3 RESULTS AND DISCUSSION

Comparative performance across multiple classifiers indicates that machine learning algorithms can accurately distinguish asthma from COPD based on lung sounds with promising metrics, subject to quality data acquisition and preprocessing procedures. Table 3 presents the comparative performance of the proposed classifiers after hyperparameter optimization. Balanced accuracy and macro F1-score evaluate the impact of class imbalance, whereas per-class sensitivity or recall, specificity, and F1-score provide a detailed assessment of each classifier's performance for individual lung disease classes. The classification accuracy ranged from 93.7% to 96.6%, with RF achieving the highest balanced accuracy, macro F1-score, and macro–Area Under the Curve (AUC). In contrast, k-NN required the shortest training time but achieved the lowest classification accuracy. These results indicate that RF provides the best balance between predictive performance and computational efficiency, whereas SVM achieved competitive accuracy at the expense of substantially higher computational time. Table 4 presents the per-class evaluation of all classifiers, demonstrating excellent performance in identifying the COPD class, with F1-score values ranging from 99% to 100% and AUC values reaching 100%. This indicates that COPD features are highly distinguishable from those of other classes. In contrast, the Asthma class remained the most challenging category, particularly for the MLP classifier, due to its lower recall and F1-score values. Among the evaluated classifiers, RF consistently achieved the best per-class performance, obtaining high F1-scores across all classes, ranging from 79% to 100%, while maintaining high specificity. These findings are further supported by the confusion matrix heatmaps in Figure 2, which demonstrate strong class discrimination, with most samples correctly classified and only limited misclassifications observed, mainly involving the Asthma and other respiratory disease categories.

Table 3: Overall performance comparison

Classifier	Accuracy (%)	Balanced Accuracy (%)	Macro F1 (%)	Macro AUC (%)	Training Time (s)
RF	**96.59**	**89.5**	**88.14**	**99.5**	14.76
SVM	95.06	84.5	85.6	98.5	1439.96
MLP	94.76	80.75	81.3	98.75	707.14
k-NN	93.72	85	85.9	95.75	**3.9**

Table 4: Per-class classification performance of different classifiers

Class	Performance Metric	RF (%)	SVM (%)	MLP (%)	k-NN (%)
Asthma	Precision	90	72	71	73
	Recall	70	65	50	65
	F1-score	79	68	59	69
	Specificity	99	98	98	99
	AUC	99	97	97	90
COPD	Precision	100	99	99	99
	Recall	99	99	99	99
	F1-score	100	99	99	99
	Specificity	100	97	99	97
	AUC	100	100	100	100
Healthy	Precision	78	77	66	74
	Recall	93	87	91	87
	F1-score	85	82	76	80
	Specificity	98	98	98	97
	AUC	99	98	99	97
Others	Precision	91	91	96	96
	Recall	91	87	83	89
	F1-score	91	89	89	92
	Specificity	98	99	97	99
	AUC	100	99	99	96

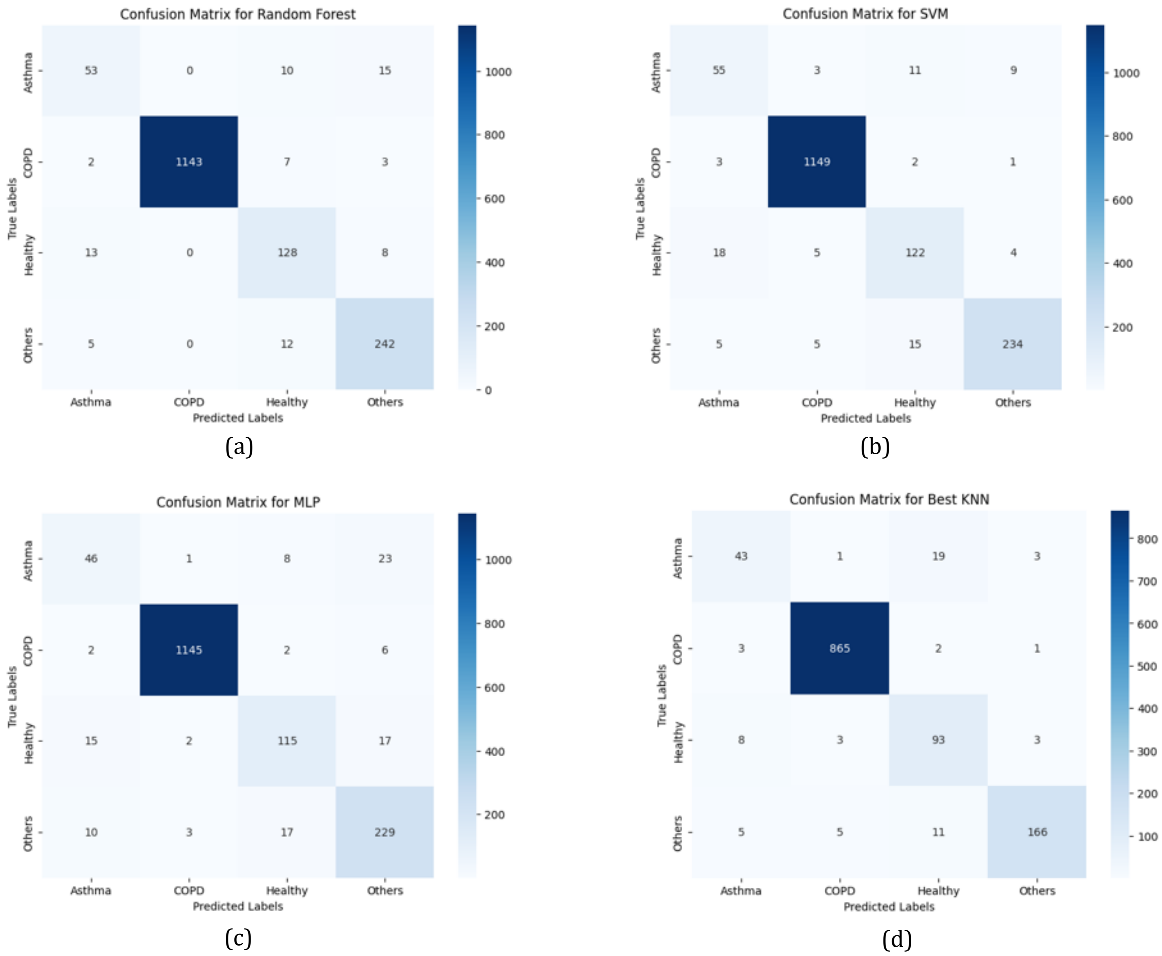

Figure 2: Confusion matrix heatmaps for: (a) Random Forest; (b) Support Vector Machine; (c) Multi-layer Perceptron; (d) k-Nearest Neighbors.

Moreover, integrating a hyperparameter optimization process such as Grid Search ensures that each classifier is fine-tuned to achieve its optimal performance. Figure 3 presents visual performance differences among four classification algorithms on lung sound data after tuning parameter process. After hyperparameter optimization, all classifiers demonstrated improved validation and test accuracies. RF achieved the highest accuracy after tuning, demonstrating that parameter adjustments, such as the number of trees and maximum depth, enhance its ability to capture complex decision boundaries. For the SVM, validation accuracy improved from 0.9512 to 0.9596, while the test accuracy remained relatively stable from 0.9488 to 0.9506. In the case of the RF classifier, both validation and test accuracies increased, from 0.9619 to 0.9672 and from 0.9543 to 0.9659, respectively, highlighting the effectiveness of parameter tuning. The k-NN classifier showed the most significant improvement, with accuracy increasing from 0.8964 to 0.9372 afterward, representing an enhancement of 4.08%. In contrast, the accuracy improvements for MLP, RF, and SVM were comparatively smaller, at around 1.22%, 1.16%, and 0.18%, respectively. This bar chart showed that the hyperparameter optimization process does not only improve accuracy but also tailors each classifier to better handle the class imbalance present in the dataset.

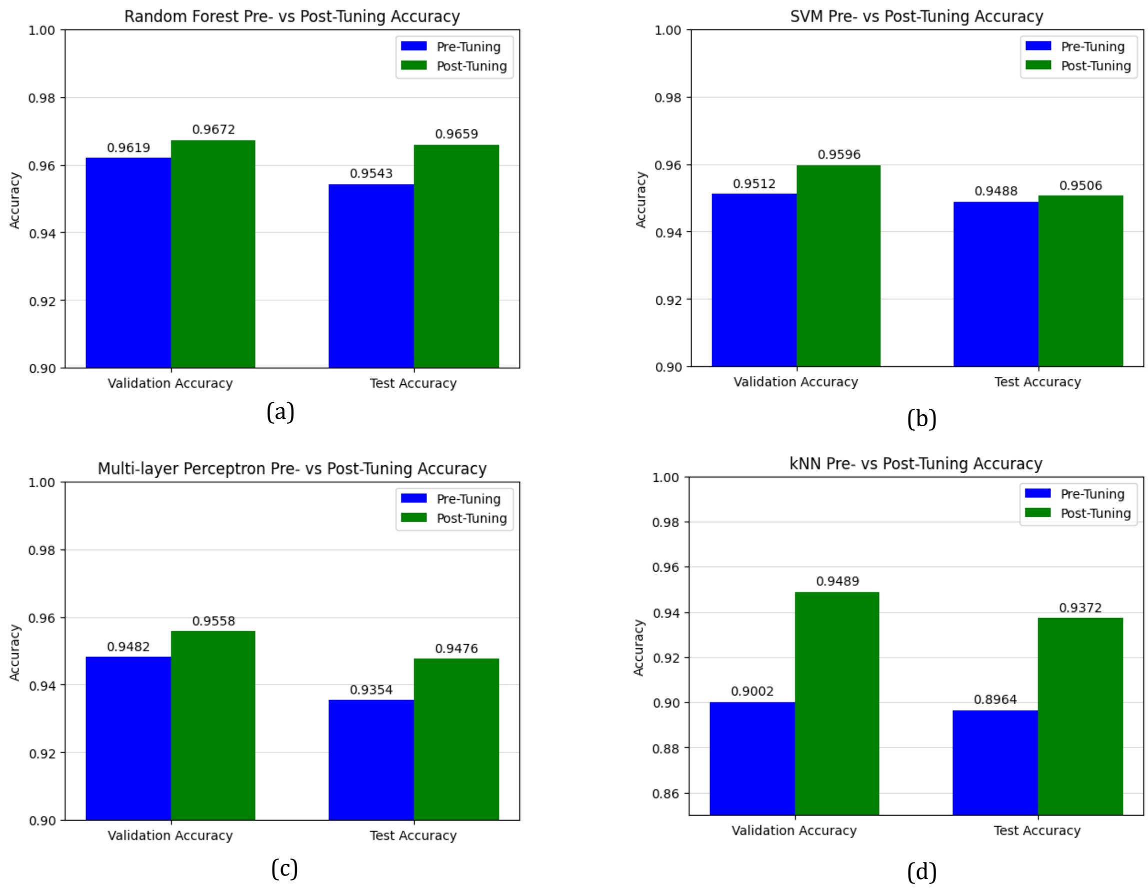

Figure 3: Comparison of validation and test accuracy before and after hyperparameter optimization using Grid Search for (a) Random Forest; (b) Support Vector Machine;(c) Multilayer Perceptron; (d) k-Nearest Neighbors.

Table 5 highlights the best parameter and training time for each classifier after hyperparameter optimization. RF achieved stability with 100 estimators, unrestricted depth, and controlled splitting to avoid overfitting. SVM performed optimally with an RBF kernel, a higher penalty parameter, and automatic gamma adjustment. The MLP performed best with a single hidden layer of 64 neurons, ReLU activation, and SGD optimizer with a learning rate of 0.01, balancing learning efficiency and accuracy. Lastly, the k-NN classifier showed the optimal setting was three neighbors using distance-based weighting and the Manhattan metric, which improved classification performance on the dataset. The recorded grid search times highlight the computational cost of optimization, with SVM requiring the longest tuning time, followed by MLP, RF and k-NN. These results indicate that simpler models such as k-NN demand substantially less computational effort compared to more complex models like SVM and MLP, underscoring the trade-off between model complexity and tuning efficiency.

The findings highlight that hyperparameter tuning enhances classification performance across all models, with RF and k-NN showing the most substantial improvements. Furthermore, while complex models like SVM and MLP demanded longer optimization times, simpler models such as k-NN achieved strong performance with comparatively lower computational cost, emphasizing the balance between model complexity, efficiency, and performance.

Table 5: Optimal Hyperparameters for Each Classifier After Grid Search

Classifier	Best Parameter	Grid Search time (s)
RF	'n_estimators': 100 'max_depth': None 'min_samples_split': 2 'min_samples_leaf': 2 'bootstrap': False	7728.27
SVM	'C': 10 'kernel': 'rbf' 'gamma': 'auto' 'class_weight': None	20840.01
MLP	'hidden_layer_sizes': (64,) 'activation': 'relu' 'solver': 'sgd' 'learning_rate_init': 0.01	16876.77
k-NN	'n_neighbors': 3 'weights': 'distance' 'metric': 'manhattan'	3995.53

When compared with findings in the existing literature, the observed improvements are consistent with prior studies reporting that hyperparameter tuning can significantly impact model reliability in biomedical signal classification. Specifically, the notable improvement in k-NN performance aligns with evidence that distance-based classifiers are highly dependent on carefully chosen parameters. The consistent gains across RF and MLP emphasize the benefits of tuning methods and neural networks in healthcare-related predictive tasks.

A comparative evaluation against prior literature in Table 6 highlights the robustness and versatility of the proposed framework. For binary classification tasks, studies such as Şen et al. reported 98% accuracy and Mukherjee et al. reported 99.22% accuracy. However, their scopes were strictly limited to two-class distinctions such as asthma versus COPD or healthy versus non-healthy cases, often using private or non-fused datasets. In contrast, the proposed framework addresses a significantly more complex multi-class problem involving four distinct respiratory categories, where overlapping acoustic patterns and data imbalance present greater diagnostic challenges. Furthermore, when evaluating performance within multi-class paradigms, the proposed approach demonstrates clear superiority over existing methods. Khanaghavalle et al. achieved 94.57% accuracy for COPD severity stages using ResNet50. While Aykanat et al. achieved only 64% accuracy in multi-class classification using unoptimized single-domain MFCCs, the Grid Search-optimized Random Forest in this study achieved a higher accuracy of 96.59% while expanding the scope to multi-disease classification across fused datasets. In summary, our hybrid feature-stacking strategy combined with hyperparameter optimization proves crucial for achieving high accuracy across lung conditions.

Table 6: Comparison of Proposed Framework with Prior Benchmark Studies

Study	Dataset	Task	Feature Extraction	Model	Optimization Method	Reported Metrics
Sen et. al [1]	Private 14-channel respiratory sounds	Binary (Asthma vs COPD)	Vector Autoregressive (VAR) Model	Bayesian + Linear SVM	None	Accuracy : 98%
Petmezas et al. [18]	ICBHI	Multi-class lung sound	STFT Spectrogram	Hybrid CNN-LSTM	Focal Loss	Accuracy : 76.39%, Sensitivity : 52.78%
Khanaghavalle et al. [31]	Respiratory Database @ TR	COPD severity	Spectrogram + Mel-spectrogram + Chromogram	ResNet50	Data augmentation + Hyperparameter fine-tuning	Accuracy : 94.57%
Huang et al. [32]	ICBHI	Multi-class lung disease	Mel-spectrogram	CNN	None	Accuracy: 84.7% F1-score: 86.6%
Mukherjee et al. [33]	ICBHI	Binary (Healthy vs Non-healthy)	Linear Predictive Cepstral Coefficient (LPCC)	MLP classifier	Manual hyperparameter tuning	Accuracy: 99.22%, F1-score: 99.17%
Aykanat et al. [34]	Private	Multi-class (12 lung disease)	MFCC	SVM, k-NN, Gaussian Bayes	None	Best Accuracy k-NN: 64%
Haider et al. [35]	Private	Binary (COPD vs Healthy)	Lung sound features (median frequency, Linear Predictive Coding)	SVM, k-NN, Logistic Regression, Decision Tree, Discriminant Analysis	Feature selection	Best Accuracy SVM: 83.6%
Proposed Work	ICBHI + KAUH	Multi-class (Asthm, COPD, Healthy, Others)	MFCC + Mel-Spectrogram + Zero Crossing Rate	RF, MLP, SVM, k-NN	Grid Search + Stratified CV	Accuracy: 96.59%, Balanced Accuracy: 89.50%, Macro-F1: 88.14%, AUC: 99.5%

Despite these positive outcomes, some limitations should be acknowledged. The proposed framework was evaluated using segmented respiratory sound samples, where multiple segments derived from the same recording probably share similar acoustic characteristics. Therefore, future work should adopt patient-level or recording-level data partitioning to reduce the risk of information leakage and provide a more reliable evaluation of model generalization. The proposed framework was evaluated on an imbalanced dataset, particularly with limited asthma samples, which remains a limitation of this study. Future work will explore advanced data balancing techniques, such as Variational Autoencoders (VAEs) or SMOTE, and cost-sensitive learning approaches to enhance minority class recognition. Finally, the grid search-based hyperparameter tuning process is computationally intensive, which limits its scalability for larger datasets. More efficient optimization approaches could be explored to reduce computational costs while maintaining classification performance.

Overall, the findings underscore the importance of systematic hyperparameter optimization in developing reliable machine learning frameworks for respiratory disease classification. By improving classifier performance, particularly in k-NN and RF, tuning contributes to building more robust diagnostic tools, which could enhance the accuracy of computer-aided systems for asthma and COPD detection.

4 CONCLUSION

This study addressed key challenges in lung sound-based disease classification by systematically evaluating feature extraction methods, machine learning classifiers, and the role of hyperparameter optimization. Current limitations, including noise interference, class imbalance, and variability in dataset, were identified as critical barriers to robust model generalization. To overcome these issues, the study highlighted the importance of preprocessing techniques for noise reduction and normalization, which ensure more reliable feature representation. Furthermore, a comprehensive examination of state-of-the-art machine learning architecture demonstrated their suitability for multi-class lung disease. Among the evaluated classifiers, RF was identified as a suitable baseline model for respiratory disease classification due to its strong predictive capability and computational efficiency. Systematic optimization consistently improved accuracy, with RF and k-NN exhibiting the most pronounced gains. The analysis further revealed that while complex models such as SVM and MLP required longer tuning times compared to simpler models like k-NN, which offered a more efficient balance between performance and computational cost.

These findings underscore the importance of selecting appropriate optimization strategies and model architectures to ensure both reliability and efficiency in medical diagnostic applications. However, challenges such as environmental noise, dataset limitations, variability in recording conditions, and lack of model interpretability remain substantial barriers to clinical implementation. Future efforts should focus on integrating multimodal data sources, developing explainable AI models to foster clinical trust, and creating standardized, large-scale annotated datasets to enable robust model validation. Moreover, the development of real-time, portable devices could enhance continuous monitoring and early diagnosis in practical healthcare settings. Collectively, these directions promise to bridge the gap between research advancements and clinical deployment, ultimately improving respiratory disease diagnosis and patient outcomes.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Universiti Malaysia Perlis for providing the facilities and support for this research. The authors also thank all supervisors and research members for their valuable guidance and acknowledge the contributors of the lung sound datasets used in this study.

REFERENCES

[1] I. Sen, M. Saraclar, and Y. P. Kahya, "Differential diagnosis of asthma and COPD based on multivariate pulmonary sounds analysis," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1601–1610, 2021.

[2] K. Rishabh, D. Singh, and V. Gupta, "Deep learning approaches for lung sound classification: A comprehensive review," *Artif. Intell. Med.*, vol. 150, p. 102785, 2025. doi: 10.1016/j.artmed.2024.102785.

[3] G. R. Khanaghavalle, G. Rahul, S. R. Senajith, T. S. Vishnuvasan, and S. Keerthana, "Chronic obstructive pulmonary disease severity classification using lung sound," in *Proc. 10th Int. Conf. Commun. Signal Process. (ICCSP)*, 2024, pp. 428–432.

[4] J. Sanjana, P. P. Naik, M. A. Padukudru, S. G. Koolagudi, and J. Rajan, "Attention-based CRNN models for identification of respiratory diseases from lung sounds," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2023, pp. 1–7.

[5] Z. Neili and K. Sundaraj, "Addressing varying lengths in PCG signal classification with BiLSTM model and MFCC features," in *Proc. 8th Int. Conf. Image Signal Process. Appl. (ISPA)*, 2024, pp. 1–5.

[6] S. Z. H. Naqvi and M. A. Choudhry, "An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis," *Sensors*, vol. 20, no. 22, p. 6512, 2020.

[7] J. Saldanha et al., "Data augmentation using variational autoencoders for improvement of respiratory disease classification," *PLoS One*, vol. 17, no. 8, p. e0266467, 2022.

[8] F. Gomiasti, W. Warto, E. Kartikadarma, J. Gondohanindijo, and D. Setiadi, "Enhancing lung cancer classification effectiveness through hyperparameter-tuned support vector machine," *J. Comput. Theor. Appl.*, 2024. doi: 10.62411/jcta.10106.

[9] H. Saini and D. Saini, "Enhancing classification of lung diseases by optimizing training hyperparameters of the deep learning network," *Multimed. Tools Appl.*, 2024. doi: 10.1007/s11042-024-20085-2.

[10] Y. Zamzam, T. Saragih, R. Herteno, D. Nugrahadi, and P. Huynh, "Comparison of CatBoost and Random Forest methods for lung cancer classification using hyperparameter tuning Bayesian optimization-based," *J. Electron. Electromed. Eng. Med. Informat.*, vol. 6, no. 2, p. 382, 2024. doi: 10.35882/jeeemi.v6i2.382.

[11] S. Nabeel et al., “Optimizing lung cancer classification through hyperparameter tuning,” *Digital Health*, vol. 10, p. 20552076241249661, 2024. doi: 10.1177/20552076241249661.

[12] S. Ashwini, J. Arunkumar, R. Prabu, and N. Singh, “Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network,” *Soft Comput.*, vol. 27, pp. 1–15, 2023. doi: 10.1007/s00500-023-09480-3.

[13] M. Alamgeer et al., “Dung beetle optimization with deep feature fusion model for lung cancer detection and classification,” *Cancers*, vol. 15, no. 15, p. 3928, 2023. doi: 10.3390/cancers15153982.

[14] H. Askr, M. Moawad, A. Darwish, and A. Hassanien, “Multiclass deep learning model for predicting lung diseases based on honey badger algorithm,” *Int. J. Inf. Technol.*, 2024. doi: 10.1007/s41870-024-02046-y.

[15] B. M. Rocha et al., “A respiratory sound database for the development of automated classification,” in *Proc. Int. Conf. Biomed. Health Informat.*, 2017, pp. 33–37. doi: 10.1007/978-981-10-7419-6_6.

[16] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, “A dataset of lung sounds recorded from the chest wall using an electronic stethoscope,” *Data Brief*, vol. 35, p. 106913, 2021. doi: 10.1016/j.dib.2021.106913.

[17] U. Hassan, A. Singhal, and P. Chaudhary, “Lung disease detection using EasyNet,” *Biomed. Signal Process. Control*, vol. 91, p. 105944, 2024. doi: 10.1016/j.bspc.2024.105944.

[18] 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, 2022. doi: 10.3390/s22031232.

[19] S. Ali et al., “Towards the development of the clinical decision support system for the identification of respiration diseases via lung sound classification using 1D-CNN,” *Sensors*, vol. 24, no. 21, p. 6887, 2024. doi: 10.3390/s24216887.

[20] R. Xu, “Research on a pre-diagnostic system for lung diseases based on multi-classification model of support vector machine,” in *Proc. Int. Conf. Electron. Devices Comput. Sci. (ICEDCS)*, 2024, pp. 1071–1076. doi: 10.1109/ICEDCS.2024.1234567.

[21] T. Fernando, S. Sridharan, S. Denman, H. Ghaemmaghami, and C. Fookes, “Robust and interpretable temporal convolution network for event detection in lung sound recordings,” *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2898–2908, 2021. doi: 10.1109/JBHI.2022.3144314.

[22] S. Sengupta, M. Sahidullah, and G. Saha, “Lung sound classification using cepstral-based statistical features,” *Comput. Biol. Med.*, vol. 75, pp. 118–129, 2016. doi: 10.1016/j.compbiomed.2016.05.015.

[23] S. Y. Jung, D. Park, J. H. Lee, and H. J. Hwang, "Classification of adventitious respiratory sounds using a convolutional neural network," *Healthcare*, vol. 9, no. 9, p. 1117, 2021. doi: 10.3390/healthcare9091117.

[24] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 195, 2017. doi: 10.1186/s13640-017-0195-0.

[25] S. S. Dhavala, K. V. Rani, M. R. Kumar, and P. Kumar, "Automated detection of wheeze and crackle sounds using Mel-frequency cepstral coefficients and deep learning techniques," *Biomed. Signal Process. Control*, vol. 72, p. 103338, 2022. doi: 10.1016/j.bspc.2021.103338.

[26] M. Engin, I. Turkoglu, and E. Z. Engin, "Performance comparison of different feature extraction methods for lung sound classification," *Comput. Methods Programs Biomed.*, vol. 214, p. 106551, 2022. doi: 10.1016/j.cmpb.2021.106551.

[27] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artif. Intell. Med.*, vol. 88, pp. 58–69, 2018.

[28] R. Wanasinghe, D. Perera, and S. Jayawardena, "Multi-feature fusion and deep learning for robust lung sound classification," *Expert Syst. Appl.*, vol. 239, p. 122345, 2024. doi: 10.1016/j.eswa.2023.122345.

[29] Z. Tariq, U. Akram, M. Z. Shaukat, and S. Khan, "Automatic detection of abnormal lung sounds using spectrogram-based deep features and convolutional neural networks," *IEEE Access*, vol. 10, pp. 51847–51859, 2022. doi: 10.1109/ACCESS.2022.3172698.

[30] K. Uppin, C. Y. Patil, and A. Kulkarni, "Detection of adventitious lung sounds using deep learning and spectrogram analysis," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 3, pp. 1042–1052, 2021. doi: 10.1002/ima.22582.

[31] G. R. Khanaghavalle, G. Rahul, S. R. Senajith, T. S. Vishnuvasan, and S. Keerthana, "Chronic obstructive pulmonary disease severity classification using lung sound," in *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, India, 2024, pp. 428–432, doi: 10.1109/ICCSP60870.2024.10543344.

[32] D.-M. Huang, J. Huang, K. Qiao, N.-S. Zhong, H.-Z. Lu, and W.-J. Wang, "Deep learning-based lung sound analysis for intelligent stethoscope," *Military Medical Research*, vol. 10, no. 1, Art. no. 44, 2023, doi: 10.1186/s40779-023-00479-3.

[33] H. Mukherjee, P. Sreerama, A. Dhar, S. M. Obaidullah, K. Roy, M. Mahmud, and K. C. Santosh, "Automatic lung health screening using respiratory sounds," *Journal of Medical Systems*, vol. 45, no. 2, Art. no. 19, 2021, doi: 10.1007/s10916-020-01681-9.

[34] M. Aykanat, Ö. Kılıç, B. Kurt, and S. B. Saryal, "Lung disease classification using machine learning algorithms," *International Journal of Applied Mathematics, Electronics and Computers*, vol. 8, no. 4, pp. 125–132, 2020, doi: 10.18100/ijamec.799363.

[35] N. S. Haider, B. K. Singh, R. Periyasamy, and A. K. Behera, "Respiratory sound based classification of chronic obstructive pulmonary disease: A risk stratification approach in machine learning paradigm," *Journal of Medical Systems*, vol. 43, no. 8, Art. no. 255, Jun. 2019, doi: 10.1007/s10916-019-1388-0.