# European Journal of Clinical Pharmacy

**Print ISSN:** 2385-409X **Online ISSN**: 3105-0409

Website: https://farmclin.com/


*Original Research Article*

# Comparative Analysis of Time-Frequency and Cepstral Features for Lung Disease Classification Using Deep and Traditional Machine Learning Models

*Article History:*


**Name of Author:**
Nisha. P[1*], Dr. J. Vijaya Kumar[1] and Dr. Nandhirajan[2]

***Affiliation:*** *[1]Department of Electronics and Instrumentation, Bharathiar University, Coimbatore, India.*
*[2]Raja Clinic, Oddanchatram, Dindigul, India.*

***Corresponding Author:***
Nisha. P



***Received***: 29-11-2025
***Revised***: 10-12-2025
***Accepted***: 20-12-2025
***Published***: 28-12-2025





***Abstract***: The early detection of respiratory disorders depends on the accurate classification of lung sounds. Mel-Frequency Cepstral Coefficients (MFCC) and six time-frequency representations, Spectrogram, Mel-Spectrogram, Log-Mel Spectrogram, Scalogram, Constant-Q Transform (CQT), and Gammatone are examined in this study. While a custom 5-layer Convolutional Neural Network (CNN) is used for image-based features, a Support Vector Machine (SVM) is used for MFCC features. A multi-class dataset including normal lung sounds, pneumonia, asthma, and chronic obstructive pulmonary disease (COPD) is used for the experiments. The results indicate that CNN models with CQT and Log-Mel Spectrogram features have the highest accuracy (94% and 95.4% respectively), while MFCC with SVM serves as a baseline of 77.96%. These results lay the groundwork for creating accurate computer-aided diagnostic systems and highlight the significance of feature selection in automated respiratory disease analysis based on lung sounds.



**Keywords**: Lung disease classification, time–frequency features, respiratory disorders, CNN, SVM.


## INTRODUCTION

Accurate extraction and classification are still essential for a thorough assessment of the respiratory system, despite the fact that lung sound signals are naturally non-stationary and unpredictable, making reliable feature extraction difficult [1]. The fast diagnosis of diseases like pneumonia, asthma, and chronic obstructive pulmonary disease (COPD) can be aided by the early detection of abnormal patterns of sound in breathing. Computerized lung sound analysis reduces subjective variation that frequently occurs during standard auscultation and provides an accurate and objective diagnostic technique. These systems help doctors make clinical decisions by offering correct interpretation, which also makes it possible for them to treat more patients more effectively [2]. Therefore, the accuracy and accessibility of healthcare delivery can be improved by integrating automation into the diagnosis of respiratory diseases.

### 1.1 Importance of Lung Sound Analysis

Accurate diagnosis depends on identifying irregular characteristics in lung sounds, but these signals are frequently non-linear and difficult to understand. As such, standard lung sound tests continue to be essential in clinical and public health contexts [3]. Lung sounds, both normal and abnormal, reveal important details about the condition of the respiratory system [4]. When irregular sounds are immediately and precisely identified, the number of deaths and disability rates associated with respiratory illnesses can be reduced [5]. In order to facilitate clinical evaluation, computerized respiratory sound analysis (CRSA) automatically captures lung sounds and stores them in a structured database. By making it easier to compare respiratory sounds, these algorithms help meet the increasing demand for accurate automatic diagnosis [6].

### 1.2 Time-Frequency and Cepstral Features for Lung Diseases

The classification of lung disease requires the ability to represent audio signals effectively. Comprehensive information about the spectral and temporal characteristics of signals can be obtained through time-frequency representations [7,8] like Spectrograms, Mel-Spectrograms, Log-Mel Spectrograms, Scalograms, Constant-Q Transform (CQT), and Gammatone filterbanks. Since they enable the network to take advantage of spatial patterns in the data, these 2D representations are especially well-suited for convolutional neural networks (CNNs) [9].

Speech and audio processing frequently use Mel-Frequency Cepstral Coefficients (MFCCs) [3], which provide a condensed representation of the spectral envelope of signals [10]. MFCCs are an excellent starting point for machine learning-based lung disease classification because of their effectiveness in feature extraction and capacity to capture temporal dynamics through delta and delta-delta coefficients.

### 1.3 Challenges in Automated Lung Sound Classification

There are various difficulties in automatically classifying lung diseases:

- Differences in recording circumstances: Signal quality is impacted by variations in stethoscope types, patient posture, and background noise.
- Subtle variations in pathological sounds: Accurate feature representation is necessary to distinguish similar aberrant sounds (such as crackles versus wheezes).
- Insufficiently labeled databases: It is difficult to train and generalize models due to the scarcity of clinical datasets with annotated lung sounds.

### 1.4 Machine Learning Approaches in Lung Sound Analysis

When machine learning has advanced, additional approaches for analysing lung sounds are starting to classify data using machine learning methods [11]. CNNs are frequently used to capture both temporal and spectral patterns in image-like representations like spectrograms and scalograms. Traditional classifiers like support vector machines (SVMs) offer computationally effective substitutes for 1D feature vectors like MFCCs [7, 12]. By combining these methods, a comparative analysis can determine which feature-classifier combination is best for lung disease analysis.

### 1.5 Proposed System and Contributions

In order to classify asthma, COPD, pneumonia and normal lung sounds, this study suggests comparing six time-frequency features (Spectrogram, Mel-Spectrogram, Log-Mel Spectrogram, Scalogram, CQT, and Gammatone) using a 5-layer CNN in addition to MFCC features using SVM. This work's contributions include:

- MFCC and six time-frequency representations are compared in a single framework for a thorough feature evaluation.
- Effective classification framework: SVM for MFCC-based features and custom 5-layer CNN for image-based features.
- Finding the most discriminative feature sets for clinical lung sound analysis provides insight into feature effectiveness.
- Useful application: Establishing a baseline for automated systems that monitor lung sounds in medical environments.

This work clearly identifies the relative advantages of various feature extraction techniques and classifiers, directing future advancements in automated lung disease analysis system.

## METHODOLOGY

In order to facilitate the initial diagnosis of respiratory disorders, this study explores the automatic classification of lung sounds. The suggested method uses a custom 5-layer Convolutional Neural Network (CNN) to analyse six distinct time-frequency image representations and combines Mel-Frequency Cepstral Coefficients (MFCC) with Support Vector Machines (SVM). The framework's overall process is depicted in Figure. 1 and includes the following steps:

### 2.1 Dataset

The investigation uses a multi-class respiratory sound database that includes recordings from patients with pneumonia, COPD, and asthma in addition to recordings from healthy individuals. The dataset consists of 7,174 recordings. All of the data were made under supervision using a digital stethoscope, which ensured a high enough sampling frequency and minimal background noise. Each audio sample's respiratory state was individually labeled. The dataset was split between training and validation subgroups using an 80:20 ratio.

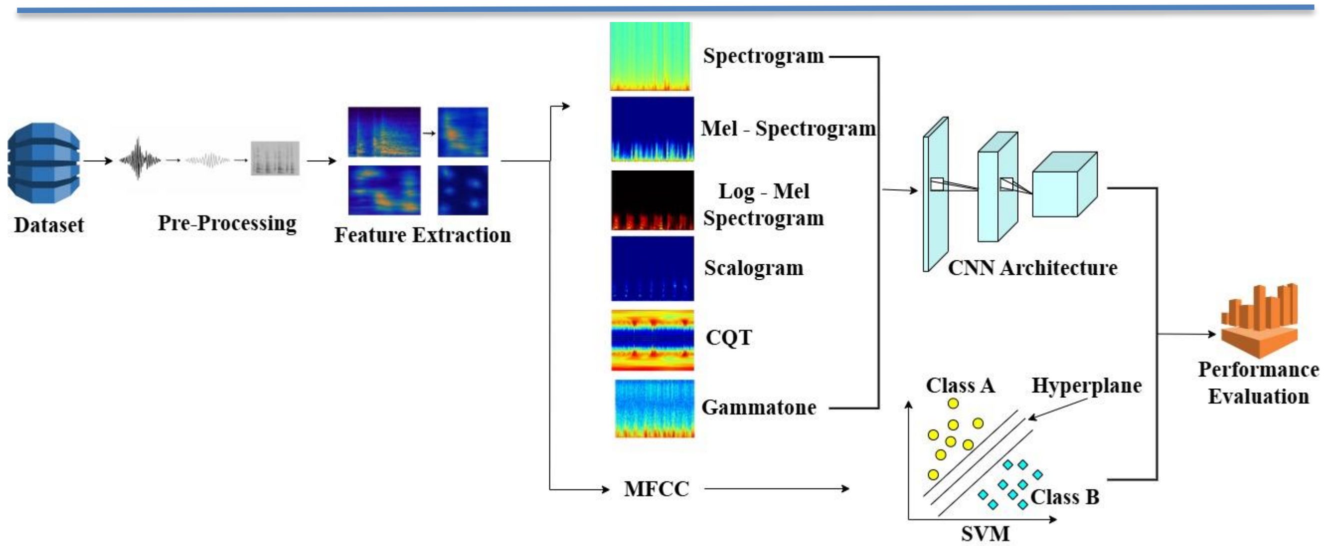



**Figure 1.** Proposed lung disease classification framework

### 2.2 Signal Preprocessing

Due to the recording environment and measurement timing, background noise and unwanted frequencies regularly contaminate raw lung sound recordings [13]. Preprocessing steps included:

- changing the sample rate to 16 kHz, which is constant.
- A bandpass filter was used in this study to remove undesired low- and high-frequency artifacts from the respiratory sound spectrum [14].
- Features are retrieved from continuous recordings by segmenting them into fixed-length frames.

### 2.3 Feature Extraction

Feature extraction, which transforms raw audio data into representations suitable for deep learning and machine learning models, is an essential step in automated lung disease classification. In this investigation, six distinct time-frequency representations were computed in order to capture the temporal and spectral characteristics of lung sounds, which are inherently non-stationary. Because each representation emphasizes a distinct characteristic of the signal, the model is able to learn a richer feature set.

Spectrogram: The spectrogram is created by applying the Short-Time Fourier Transform (STFT) to the segmented audio signal [15]. To balance time and frequency resolution, a 50% overlap Hamming window of 1024 samples (at 16 kHz) was used. To capture the frequency content over time, the signal is split up into overlapping frames, and each frame is subjected to the Fourier transform. The trade-off between frequency resolution and time resolution in the resulting spectrogram is directly controlled by the window size, which establishes the number of samples per frame [16]. This two-dimensional image, which is particularly helpful for detecting the presence of irregular breathing sounds like asthma, COPD, and pneumonia, represents energy distributions across time and frequency.

Mel-Spectrogram: The Mel-Spectrogram is created by applying a Mel-scale filter bank to the magnitude of the STFT [17]. The Mel spectrogram makes use of the Mel scale, which highlights frequency ranges that are pertinent to human perception and corresponds with the pitch sensitivity of the auditory system [18]. 64 Mel filterbanks covering the effective lung sound frequency range (50–2000 Hz) were used in this investigation [19]. This allows the model to focus on frequency components that are perceptually significant and often associated with pathological lung conditions.

Log-Mel Spectrogram: The logarithmic transformation of the Mel-Spectrogram enhances low-amplitude features, which are often subtle indicators of aberrant respiratory conditions. The logarithmic scaling was carried out as follows:

$$Log - Mel(f,t) = \log\big(1 + \alpha \cdot M(f,t)\big) \tag{1}$$

where $\alpha$ is a scaling factor set to 10 and $M(f,t)$ is the Mel-spectral magnitude. This ensures that both strong and weak frequency components make a significant contribution to feature learning while compressing the spectrogram's dynamic range. Logarithmic scaling ensures the preservation of subtle variations associated with abnormal lung sounds.

Scalogram: By using the Continuous Wavelet Transform (CWT), scalograms offer a high-resolution time-frequency representation of lung sounds [20]. In order to analyse transient phenomena like pneumonia and short-duration asthma in detail, a Morlet wavelet with 32 voices per octave was employed. By recording signal characteristics at various scales, scalograms enable the CNN to recognize hierarchical patterns.

As a reflection of auditory perception, the CQT's logarithmic frequency resolution is coarser at higher frequencies and finer at lower frequencies [21]. In order to cover clinically relevant respiratory sounds, a frequency range of 50–2000 Hz and a Q-factor of 32 were selected. CQT highlights these crucial components, which are frequently underrepresented in linear spectrograms, because many aberrant sounds (like asthma) occur in the low-frequency range.

Gammatone Filterbank: Modeled after the cochlear filtering mechanism in the human auditory system, the Gammatone filterbank focuses on perceptually important frequency bands [22]. This work produced multi-channel outputs that reflect auditory-like processing by using 64 channels spaced on an equivalent rectangular bandwidth (ERB) scale. This works especially well for catching subtle changes in abnormal lung sounds.

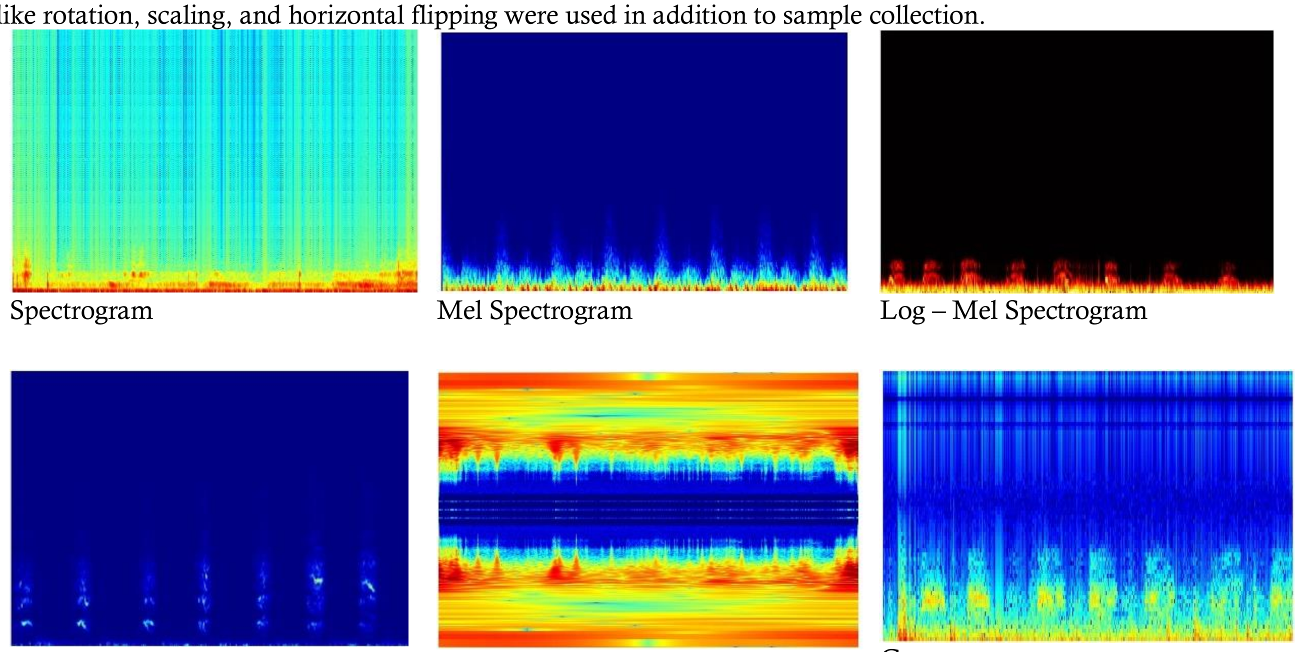
For comparison, Mel-Frequency Cepstral Coefficients (MFCCs) were also extracted. Because it offers a simplified but detailed representation of the spectral envelope of sound, this method is especially useful for respiratory disease analysis in cases of chronic obstructive pulmonary disease (COPD) [23]. 13 MFCCs per frame were calculated using a 25 ms Hamming window with a 10 ms frame shift (60% overlap) to create a 13-dimensional feature vector. The MFCC representation allows for direct performance comparison with deep learning-based methods and acts as a traditional machine learning baseline. For each model, sample images are displayed in Figure 2.

**2.4 Feature Selection and Optimization**

Preprocessing and feature optimization are still essential for improving model performance, two complementary feature extraction techniques are used after pre-processing. First, the audio signals are directly converted into Mel-Frequency Cepstral Coefficients (MFCCs), a small, manually constructed feature set that has been demonstrated to work especially well with more conventional machine learning classifiers like Support Vector Machines (SVMs). Second, Convolutional Neural Networks (CNNs), which can automatically learn hierarchical and discriminative feature representations from raw input data, are fed spectrogram representations of the pre-processed signals. An accurate comparison of deep learning-based feature learning and traditional feature extraction is made possible by this combined method [24].

Features Based on Images (CNN Input): To guarantee consistent input to the CNN, the time-frequency representations Spectrogram, Mel, Log-Mel, Scalogram, CQT, and Gammatone were transformed into grayscale pictures and adjusted to a set size of 224 × 224 pixels. Training stability and convergence are enhanced by the normalization process, which converts pixel values to a uniform range.

To increase model generalization [25], decrease overfitting, and diversify the training set, data augmentation methods like rotation, scaling, and horizontal flipping were used in addition to sample collection.



Spectrogram Mel Spectrogram Log – Mel Spectrogram

Scalogram Gammatone

CQT

**Figure 2.** Six Feature Extraction Images

MFCC-based Features (SVM Input): Thirteen coefficients per frame, a frame length of 25 ms, and 50% overlap were used to extract MFCCs. Since Mel-Frequency Cepstral Coefficients (MFCCs) approximate the frequency sensitivity of the human auditory system, they are frequently used in audio signal analysis. MFCCs offer a condensed depiction of the spectral envelope by converting linear frequencies into the perceptual Mel scale and using the discrete cosine transform (DCT) [26]. They have been widely used in speech processing due to their resilience in noisy environments, and they are also being used more and more in biomedical signals [27], such as respiratory sound analysis for conditions like COPD, asthma, and pneumonia.

Feature Optimization: To standardize the mean and variance of every feature, feature normalization was carried out in addition to PCA. This prevents learning, especially in SVM training, from being dominated by features with wider numerical ranges. In order to ensure a condensed and informative feature set, correlation analysis was also performed to eliminate highly correlated or redundant features.

### 2.5 Classifier Training

#### 2.5.1 CNN Classifier for Image-Based Features:

To learn hierarchical patterns from time-frequency images, a specially created 5-layer CNN was created. The network architecture comprised:

- Convolutional Layers: Low- to high-level feature patterns are captured by three convolutional layers with progressively larger filter sizes (32, 64, and 128).
- Batch Normalization and ReLU: ReLU adds non-linearity to model intricate relationships, while batch normalization stabilizes training by normalizing layer inputs.
- Pooling Layers: Translation-invariance is made possible by max-pooling, which minimizes spatial dimensions while maintaining important features.
- Fully Connected Layers and Dropout: Two dense layers combine features for classification and avoid overfitting by using dropout regularization.
- Softmax Output: Generates odds for every class of lung sounds.
- The Adam optimizer was used for training, with a batch size of 16 and an initial learning rate of 0.001, with early stopping determined by validation loss.

#### 2.5.2 SVM Classifier for MFCC Features:

A radial basis function (RBF) kernel was used to train the SVM classifier, which was then optimized using a grid search for the kernel width (γ) and penalty parameter (C). The input consisted of the PCA-reduced MFCC features. Ten-fold cross-validation reduced the chance of overfitting and guaranteed the accuracy of the performance metrics.

### 2.6 Performance Evaluation

The following standard metrics were calculated for each classifier:

- Accuracy (ACC): The ratio of correctly classified samples to the total samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

- F1-Score: The harmonic mean of precision and recall, providing a balance between them.

$$F - Score = \frac{2\times Precision\times Sensitivity}{Precision+Sensitivity} \quad (3)$$

Where:

- TP (True Positive): Correctly classified positive samples
- TN (True Negative): Correctly classified negative samples
- FP (False Positive): Negative samples incorrectly classified as positive
- FN (False Negative): Positive samples incorrectly classified as negative

Class-wise performance and misclassifications were also visualized through analysis of the confusion matrix.

# EXPERIMENTAL RESULTS

## 3.1 Classification results

Lung diseases were categorized into four groups using supervised learning algorithms: normal, COPD, pneumonia, and asthma. Spectrogram, Mel-Spectrogram, Log-Mel Spectrogram, Scalogram, Constant-Q Transform (CQT), and Gammatone were among the six time-frequency representations that were extracted. While Mel-Frequency Cepstral Coefficients (MFCC) were classified using a support vector machine (SVM), the image-based features were classified using a 5-layer convolutional neural network (CNN).

The overall classification performance for each feature type is shown in Table 1 along with the accuracy and F1-score. Following CQT (94% accuracy, 91% F1-score) and Mel-Spectrogram (93.6% accuracy, 91% F1-score), the Log-Mel Spectrogram had the highest accuracy of 95.4% and the highest F1-score of 94% among the CNN-based models. Gammatone features obtained 91.8% accuracy and 89% F1-score, while spectrogram and scalogram achieved moderate accuracy (89.1% and 89.8%, respectively). For lung sound classification, the SVM classifier with MFCC features performed the worst (77.96% accuracy, 77% F1-score), suggesting that image-based CNN techniques are more successful than classical feature representations.

The findings show that for all lung disease classes, CNN-based classification performs better than SVM on MFCC features. Higher classification accuracy and balanced F1-scores are the outcome of image-based representations, especially Log-Mel Spectrogram and CQT, which successfully capture discriminative patterns in lung sounds.

## 3.2 Comparative Analysis

The table.1 shows the comparison of image-based features and MFCC-based feature results. Figure. 5 shows comparative analysis of feature extraction accuracy and F1- score. The results demonstrate that CNN-based image representations outperform MFCC features with classical SVM classifiers, underscoring the importance of time-frequency image features for automated lung disease analysis.

**Table 1.** Comparison of classification performance across feature types

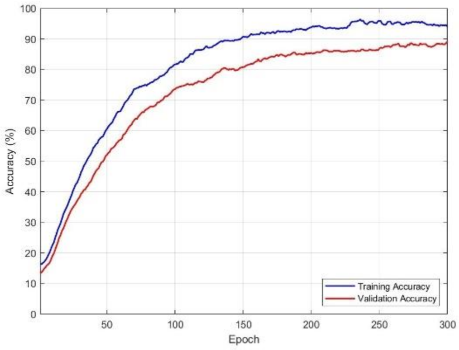
| Feature Type | Classifier | Accuracy | F1-Score |
|---|---|---|---|
| Spectrogram | CNN | 89.1 | 87 |
| Mel-Spectrogram | CNN | 93.6 | 91 |
| Log-Mel Spectrogram | CNN | 95.4 | 94 |
| Scalogram | CNN | 89.8 | 88 |
| CQT | CNN | 94 | 91 |
| Gammatone | CNN | 91.8 | 89 |
| MFCC | SVM | 77.96 | 77 |

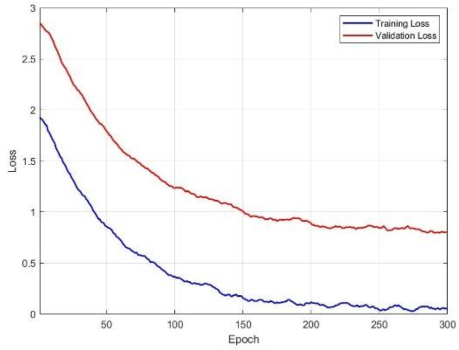
## 3.3 Visualization

Training vs. validation curves were plotted to show the learning progression. The accuracy curve shows each epoch's training and validation accuracy. Loss Curve: Shows the training and validation losses for every epoch. These plots are essential for demonstrating model convergence and preventing overfitting. The Figure. 3 shows the feature-based training and validation curves outcomes.

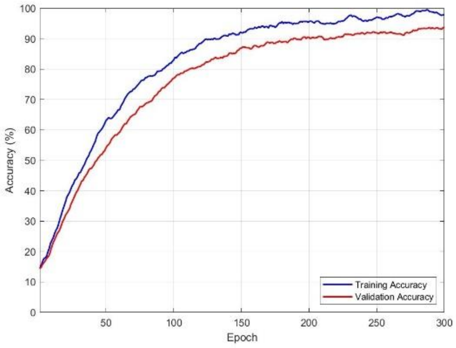
Spectrogram

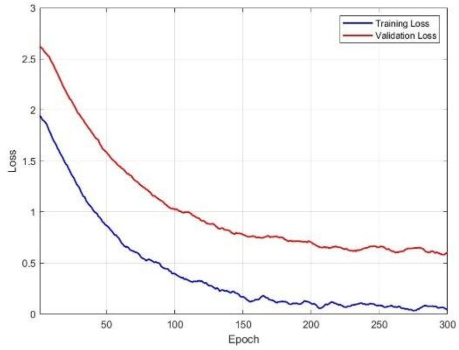



Accuracy | Loss

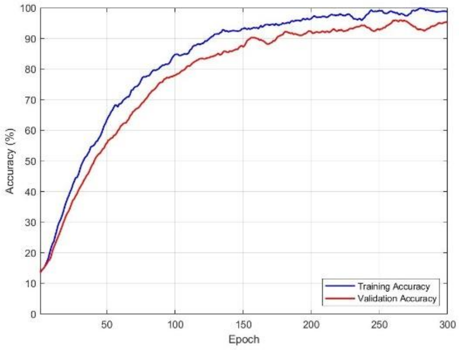
Mel-Spectrogram

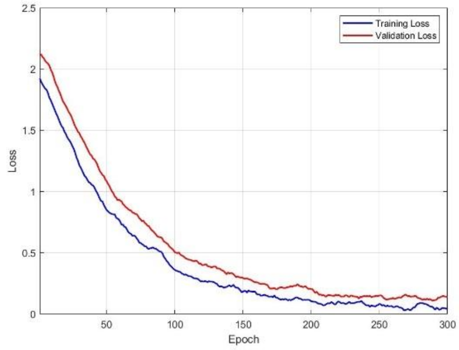



Accuracy | Loss

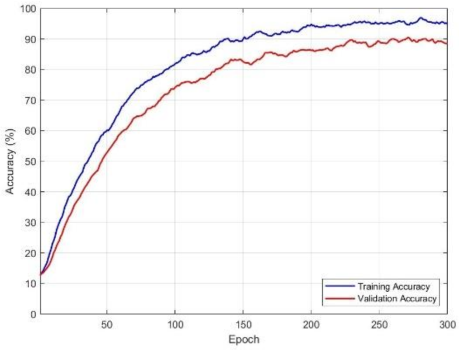
Log-Mel Spectrogram

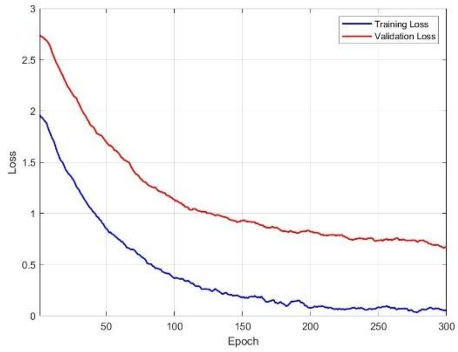



Accuracy | Loss

Scalogram

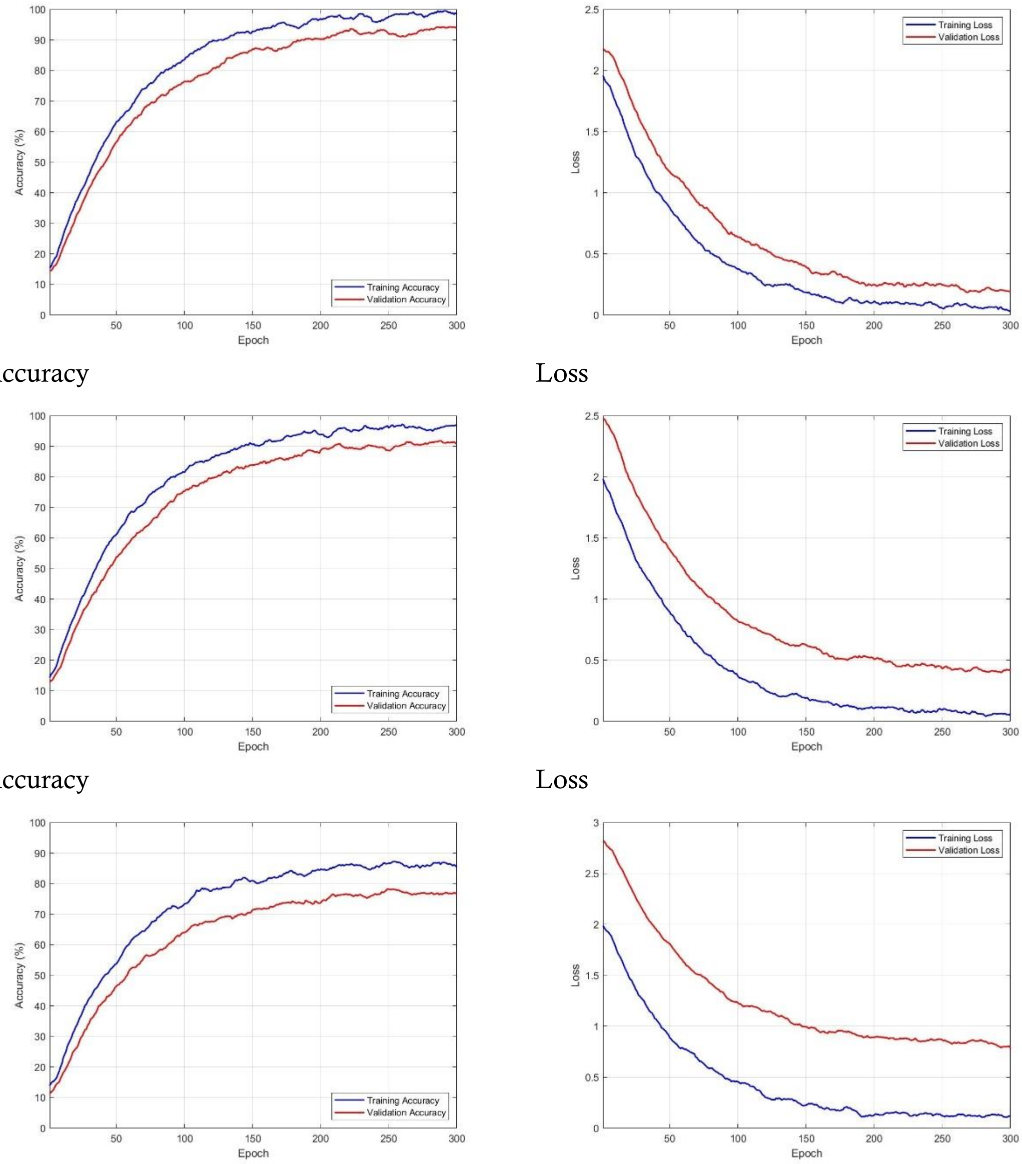



Accuracy | Loss

CQT




Accuracy Loss

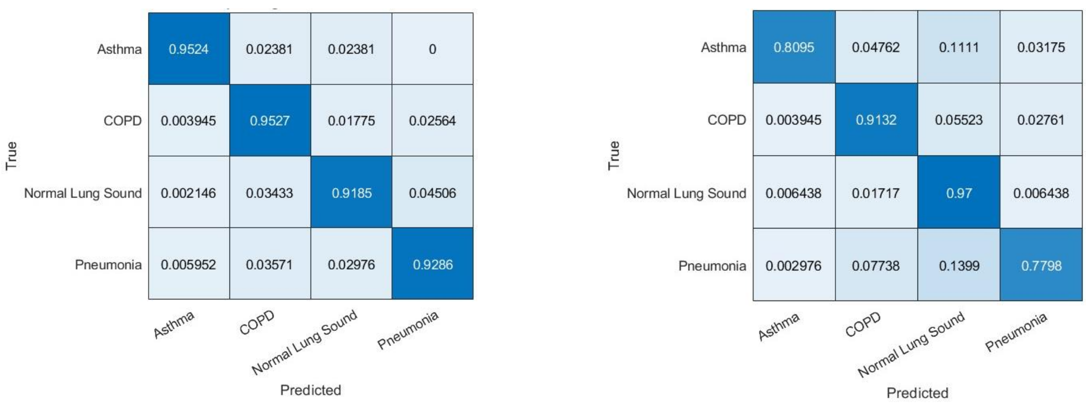
Gammatone

Accuracy Loss

MFCC

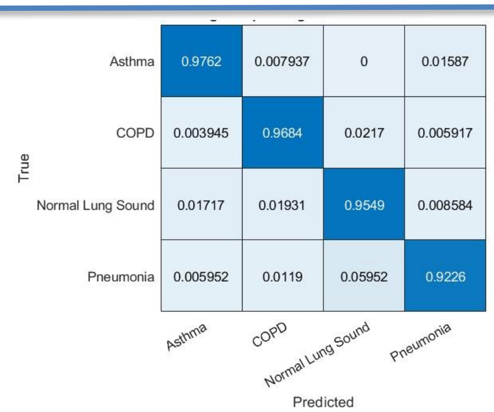
Accuracy Loss

**Figure 3.** Training and validation accuracy/loss curves for each representation.

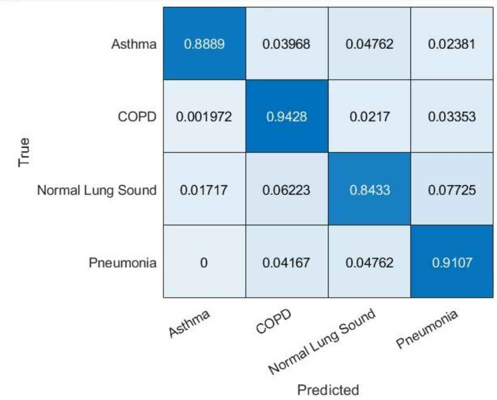



Spectrogram Mel-Spectrogram

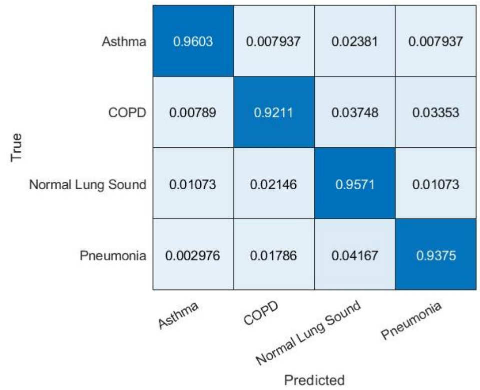
 



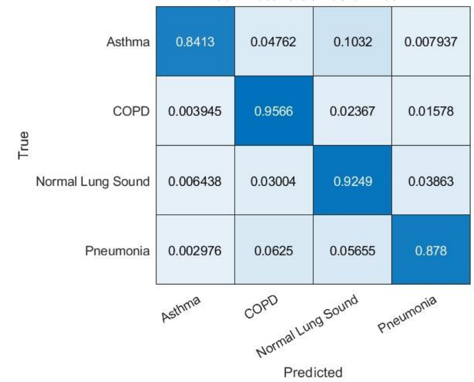

Log- Mel Spectrogram



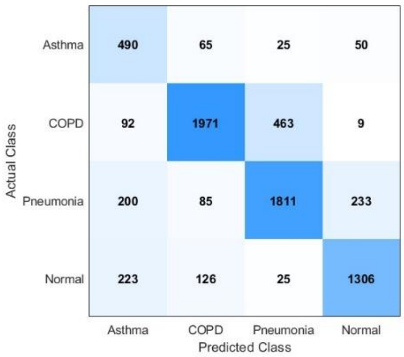

Scalogram




CQT




Gammatone




MFCC

**Figure 4.** Confusion Matrix

# DISCUSSION

A few important findings can be focused on during the comparison of the different lung disease classification. The use of CNN to classify lung diseases has yielded significantly higher performing results than those of the SVM models based on MFCC features. The CNNs achieved their best results when using the Log-Mel Spectrogram and Constant-Q Transform (CQT) features; the best accuracies from the CNN models were found to be 95.4% for Log-Mel Spectrogram and 94% for Constant-Q Transform features. The outstanding performance of the CNN models is likely due to the Log-Mel Spectrogram and CQT features being convolutional in nature, as both are Time-Frequency representations of lung sound signals and they both hold the temporal and spectral characteristic of non-stationary lung sound signals. The use of logarithmic scales for Log-Mel features provides a means of separating subtle differences in respiratory sounds to assist in identifying normal versus abnormal classes. From the classifier performance comparison, we observed that the CNNs produced a higher accuracy score and F1 score than the SVM classifiers trained on MFCC features. This difference is believed to be due to the ability of CNNs to learn and automatically extract hierarchical and discriminative representations directly from lung sound images. The large differences in accuracy and F1-scores between SVM and CNN-based models can be attributed to this distinction. Although the results obtained from the current study are encouraging, there are several limitations to consider.

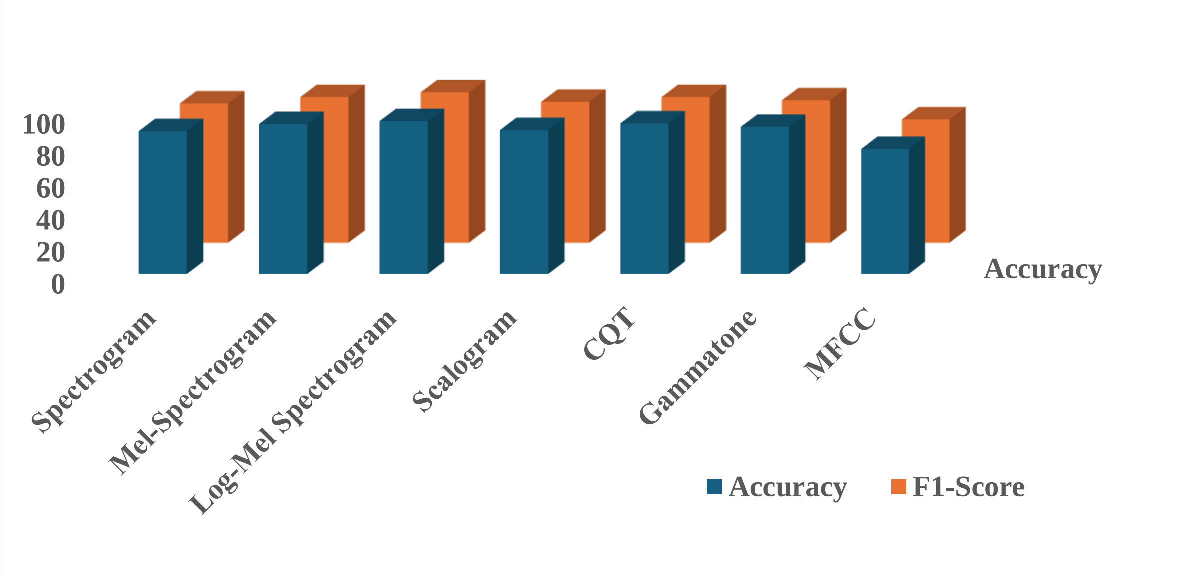



**Figure 5.** Comparative analysis of Feature Extraction methods

The dataset used for this study was a multi-class lung sound dataset that included a comparably larger number of samples than previous studies; nevertheless, it is possible that the dataset does not adequately represent the differences in patient characteristics or variables that may impact generalizability from recording settings. The recordings needed for classification were conducted under a controlled laboratory environment; thus, it is likely that the performance in clinical settings will differ from the laboratory setting where data collection was performed. The results of this study suggest that CNN models may provide an opportunity to develop an automated approach to analyzing lung sounds through the use of log-mel spectrograms and CQT representations. This type of automated system may enable consistent, unbiased, and objective preliminary assessment of possible respiratory problems in patients to help clinicians make better informed treatment decisions. Further studies should validate these models using much larger, more diverse datasets and evaluate their practical application as real-time diagnostic aids.

## CONCLUSION

Automated lung disease classification can greatly reduce the human effort required for accurate and timely diagnosis of respiratory conditions by utilizing supervised machine learning and deep learning techniques. For prompt treatment and patient care, lung sound abnormalities must be identified early and accurately. Six time-frequency representations were taken from lung sound recordings for this study: Spectrogram, Mel-Spectrogram, Log-Mel Spectrogram, Scalogram, Constant-Q Transform (CQT), and Gammatone. Mel-Frequency Cepstral Coefficients (MFCC) were employed with a support vector machine (SVM) as a baseline, and a 5-layer convolutional neural network (CNN) was used to classify image-based features. The best classification performance was obtained by the Log-Mel Spectrogram and CQT features, which had balanced F1-scores and accuracies of 95.4% and 94%, respectively. With an accuracy of 77.96%, MFCC + SVM was used as a traditional benchmark, demonstrating CNNs' superior ability to extract discriminative and hierarchical features from time-frequency images.

The findings show that CNN-based classification using CQT and Log-Mel Spectrogram features offers a reliable and effective method for automatically identifying abnormalities in lung sounds. This method can help physicians screen for respiratory diseases more consistently and with less diagnostic burden.

### Conflict of Interest

The authors declare no conflict of interest.

### Funding

This research received no external funding.

## REFERENCES

1. Amy M. Kwon, et al., "A temporal dependency feature in lower dimension for lung sound signal classification", Scientific Reports | (2022), https://doi.org/10.1038/s41598-022-11726-3.
2. Irin bandyopadhyava, et al., "Automatic lung sound cycle extraction from single and multichannel acoustic" Biomedical Signal Processing and Control 64 (2021), https://doi.org/10.1016/j.bspc.2020.102332.

3. Murat Aykanat, et al., "Classification of lung sounds using convolutional neural networks", EURASIP Journal on Image and Video Processing (2017), DOI 10.1186/s13640-017-0213-2.
4. Bruno M Rocha, et al., "An open access database for the evaluation of respiratory sound classification algorithms", Physiological Measurement (2019), https://doi.org/10.1088/1361-6579/ab03ea.
5. Yun Chu, et al., "CycleGuardian: a framework for automatic respiratory sound classification based on improved deep clustering and contrastive learning", Complex &Intelligent Systems (2025), https://doi.org/10.1007/s40747-025-01800-4.
6. Ximing Liao, et al., "Automated detection of abnormal respiratory sound from electronic stethoscope and mobile phone using MobileNetV2", BiocyberneticsandBiomedicalEngineering43(2023)763–775, https://doi.org/10.1016/j.bbe.2023.11.001.
7. FATIH DEMIR, et al., "Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure", IEEE access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.3000111.
8. L.D. Mang, et al., "Cochleogram-based adventitious sounds classification using convolutional neural networks", Biomedical Signal Processing and Control 82 (2023) 104555, https://doi.org/10.1016/j.bspc.2022.104555.
9. Dat Ngo, et al., "Deep Learning Framework Applied For Predicting Anomaly of Respiratory Sounds", 26 Dec 2020.
10. Xiaoping Xie, et al., "A Voice Disease Detection Method Based on MFCCs and Shallow CNN", https://doi.org/10.1016/j.jvoice.2023.09.024.
11. Fan Wang, et al., "Classification of Abnormal Lung Sounds Using Deep Learning", https://www.researchgate.net/publication/374561628.
12. Ali Mohammad Alqudah, et al., "Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds", Soft Computing (2022) 26:13405–13429, https://doi.org/10.1007/s00500-022-07499-6.
13. Youngjin Choi, et al., "Interpretation of lung disease classification with light attention connected module", Biomedical Signal Processing and Control 84 (2023) 104695, https://doi.org/10.1016/j.bspc.2023.104695.
14. YOUNGJIN CHOI, et al., "Lightweight Skip Connections With Efficient Feature Stacking for Respiratory Sound Classification", IEEE Access, Digital Object Identifier 10.1109/ACCESS.2022.3174678.
15. Cristina Jácome, et al., "Convolutional Neural Network for Breathing Phase Detection in Lung Sounds", Sensors 2019, 19, 1798, doi:10.3390/s19081798.
16. Shing-Yun Jung, et al., "Efficiently Classifying Lung Sounds through Depthwise Separable CNN Models with Fused STFT and MFCC Features", Diagnostics 2021, 11, 732, https://doi.org/10.3390/diagnostics11040732.
17. LUKUI SHI, et al., "Lung Sound Recognition Algorithm Based on VGGish-BiGRU", IEEE Access, Digital Object Identifier 10.1109/ACCESS.2019.2943492.
18. Jaenal Arifin, et al., "Enhancing Lung Sound Classification Using Transfer Learning with ResNet50 and Mel Spectrogram Pre-processing", International Journal of Intelligent Engineering and Systems, Vol.18, No.6, 2025, DOI: 10.22266/ijies2025.0731.39.
19. Antonio Castejón-Barrio, et al., "Leveraging unlabeled data for lung sound classification through self-supervised contrastive learning", Biomedical Signal Processing and Control 112 (2026) 108477, https://doi.org/10.1016/j.bspc.2025.108477.
20. Mohamed Loey, et al., "COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models", Computers in Biology and Medicine 139 (2021) 105020, https://doi.org/10.1016/j.compbiomed.2021.105020.
21. Mohammed Saddek Mekahila, et la., "A comparative analysis of Constant-Q Transform, gammatonegram, and Mel-spectrogram techniques for AI-aided cardiac diagnostic", Medical engineering and physics 137 (2025) 104302, http://doi.org/10.1016/j.medengphy.2025.104302.
22. Yun Chu, et al., "EZhouNet: A framework based on graph neural network and anchor interval for the respiratory sound event detection", Biomedical Signal Processing and Control 112 (2026) 108491, https://doi.org/10.1016/j.bspc.2025.108491.
23. Ahmed I. Taloba, et al., "Detecting respiratory diseases using machine learning-based pattern recognition on spirometry data", Alexandria Engineering Journal 113 (2025) 44–59, https://doi.org/10.1016/j.aej.2024.11.009.
24. Arnab Kumar Das, et al., "A deep learning model for depression detection based on MFCC and CNN generated spectrogram features", Biomedical Signal Processing and Control 90 (2024) 105898, https://doi.org/10.1016/j.bspc.2023.105898.
25. Jyotibdha Acharya, et al., "Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning", DOI: 10.1109/TBCAS.2020.2981172.

26. Sangeetha Balasubramanian, et al., "Machine Learning-Based Classification of Pulmonary Diseases through Real-Time Lung Sounds", International Journal of Engineering and Technology Innovation, vol. 14, no. 1, 2024, pp. 85-102, DOI: https://doi.org/10.46604/ijeti.2023.12294.
27. Rishabh, et al., "Respiratory sound classification utilizing human auditory-based feature extraction", Physica Scripta 100(2025)046003, https://doi.org/10.1088/1402-4896/adb80f.