IJIE

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 18 No. 4 (2026) 130-138
https://publisher.uthm.edu.my/ojs/index.php/ijie

Low-Cost COPD Screening from Respiratory Sounds Using MFCCs with k-NN, SVM and Decision Trees

Constantin Constantinescu[1]*, Remus Brad[1]

[1] *Department of Computer Science and Electrical Engineering, Faculty of Engineering, Lucian Blaga University of Sibiu, Emil Cioran 4, Sibiu, 550025, ROMANIA*

*Corresponding Author: ctin.constantinescu@ulbsibiu.ro
DOI: https://doi.org/10.30880/ijie.2026.18.04.009

Article Info

Received: 26 November 2025
Accepted: 17 March 2026
Available online: 24 June 2026

Keywords

COPD, respiratory sounds, MFCC, k-Nearest Neighbors

Abstract

The aim of our research is to develop a simple and inexpensive solution to detect Chronic Obstructive Pulmonary Disease (COPD) in places where sophisticated medical equipment or specialized medical personnel are not available. COPD is a long-term respiratory disorder that impacts a large portion of the global population. Early detection is crucial for effective management; however, conventional diagnostic methods are often complex, time-intensive, and expensive. This study seeks to simplify the process by introducing an accessible and non-invasive screening approach. We propose using an electronic stethoscope to listen and record the patient's respiratory sound and an artificial intelligence algorithm that automatically assists in detecting this condition. We combined two public lung-sound datasets (121 patients; 926 five-second segments) and extracted 20-dimensional MFCCs per frame. Using a patient-wise 70/30 split, k-NN achieved 81.65% segment-level accuracy, outperforming SVM and Decision Tree. Results support a low-cost screening aid for COPD in resource-limited settings. Machine learning models show promising results, indicating their ability to support COPD screening and assist clinical decision-making.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a major global health burden, ranking as the fourth leading cause of death worldwide with 3.5 million deaths in 2021 (approximately 5% of all global deaths). Chronic Obstructive Pulmonary Disease (COPD) continues to affect millions of individuals across the world, with nearly 90% of premature COPD-related deaths occurring in low- and middle-income countries [1]. It is often diagnosed only after it has reached an advanced stage, so early detection is essential for effective management, yet existing diagnostic procedures are expensive, time-consuming, and require specialized equipment and personnel. This study proposes an accessible and non-invasive screening approach that uses artificial intelligence (AI) to analyze recorded respiratory sounds and identify early signs of COPD.

Normally, diagnosing it requires detailed tests, scans, and specialist visits, which often requires significant time and expense. Our goal is to make this process simple and available to everyone. Using an electronic stethoscope, lung sounds can be recorded and automatically analyzed by AI to spot early signs of COPD. This allows doctors to provide quicker, more precise diagnoses while offering patients greater confidence.

The biggest advantage of this solution is early detection and screening of COPD. Taking steps on time can prevent hospitalizations and reduce morbidity. In addition, the solution is non-invasive and easy-to-use. Often, when we think about healthcare, our mind goes to fully equipped hospitals with specialized personnel for all

medical fields. However, this is only true for developed countries, whereas in underdeveloped countries people without access to modern healthcare still die from diseases that we cured decades ago. The solution we propose is very cheap and simple. This ensures that people in distant or low-resource areas can benefit from it, narrowing the healthcare gap and supporting fair access to good care.

However, the development of this solution also presents several challenges. A major one is training the AI model and validating it through rigorous tests and metrics. For this process, usually, a lot of diverse patient data is necessary. The more data, the better the machine learning algorithm will perform. But even if the algorithm is not perfect from the beginning, it can improve its robustness and accuracy by acquiring more data once it starts being used.

2. Related Work

The development of a solution that uses artificial intelligence (AI) to detect and diagnose Chronic Obstructive Pulmonary Disease, also known as COPD, is not without precedent. Most of the methods process medical images for the task. This study presents a method for automatically detecting emphysema regions in HRCT scans of COPD patients using multiple instances learning classifiers, showing improved correlation with lung function tests compared to traditional methods and manual radiologist annotations.

Also using CT images, [3] presented in 2022 a new method based on a graph convolution neural network that achieved an accuracy of 77% and an AUC score of 0.81. In this study [4], a better AUC result of 0.86 is presented, which in addition to detecting COPD, also aims to classify COPD patients on the Gold scale. In 2023, [5] describes a solution based on attention-guided multiple instances learning models, performing with a prediction accuracy on an external dataset of 83%. Still using medical imaging, but in this case chest X-Rays, [6] has good results using CNN based on tensor flow, while [7] tries to differentiate between obstructive and non-obstructive pulmonary diseases based on spirometry reports.

A non-invasive solution based on questionnaires was presented a decade ago in [8]. For a group of 16 patients that were monitored for 6 months using questionnaires, a probabilistic neural network predicted COPD exacerbations with an 80% accuracy. Also, non-invasive and closer to our approach is this study [9] that describes an experiment where COPD is detected based on speech. In the experiment, a Support Vector Machine classifier is used to classify whether the person reading a story has COPD or not.

Several existing studies already showed the potential of the analysis of pulmonary sounds to diagnose respiratory diseases. [10] uses the lung sound recordings from the ICBHI dataset [11]. In the experiment, various sound features were extracted from the recordings and deep learning, namely CNN, was used for the classification. The best results were achieved with the MFCCs coefficients as feature and after performing data augmentation, with a Specificity and Sensitivity of 0.92. [12] [13] [14] propose various deep learning methods as well. The authors of this study conducted a very comprehensive analysis [15]. They present a comparative study of features and machine learning classifiers to detect COPD and pneumonia. Multiple combinations show very good results with sensitivity and specificity above 90%. Using their own data set and various machine learning classification techniques, this study [16] attempts to differentiate between COPD and Asthma.

If we move away from COPD for a moment, respiration has also been analyzed as signal. [17] proposes an evaluation method for respiratory signal extraction, while in [18] a respiratory symptoms evaluation is performed among people working in industrial environment. In [19], MFCCs are used for human breathing classification with electromyography signal.

In addition to scientific research, some software applications have also shown the effectiveness of sound analysis in the screening of respiratory diseases. ResApp [20], for example, has developed a pioneering approach that uses sound algorithms to diagnose respiratory diseases by coughing sounds. This emerging approach has produced encouraging results in the accurate identification of respiratory conditions, providing a new avenue for early detection and clinical support. Nevertheless, current solutions also underscore the need for continued research and technological advancement in sound analysis–based screening methods.

3. Datasets

To train a machine learning model, we needed to build a comprehensive dataset of respiratory sounds. We are using two public data sets that contain lung sounds recordings from patients with varying clinical backgrounds. One data set was published in 2021 [21] and another was the subject of a scientific challenge, namely the ICBHI Challenge in 2019 [22]. The two datasets were collected using different recording setups. The dataset from [20] was recorded using a Littmann electronic stethoscope under controlled conditions, while the ICBHI dataset includes recordings acquired using a variety of electronic stethoscopes and microphones in real-world clinical environments. Therefore, combining the datasets results in a variety of recording devices, acoustic conditions and noise levels.

Int. Journal of Integrated Engineering Vol. 18 No. 4 (2026) p. 130-138

Both data sets contain sounds from real patients, both healthy and with various lung conditions. The recordings were analyzed by physicians who labeled them accordingly. They also contain various noises to simulate normal conditions.

Since the data comes from two separate sources, there are many differences in the representation of the data that we had to standardize. For example, in the first set of data, the health status is encoded in the name of the file (N = Healthy), whereas in the second dataset, it is provided separately in an accompanying text file.

We combined both datasets, taking only what we needed from both and leaving the rest out, to build our own. Essentially, we compiled a dataset that includes all the recordings of healthy patients from both datasets, as well as all the recordings from patients with COPD. Patients with other conditions were not considered. Although the number of healthy patients is close to the number of patients with COPD, the number and length of the recordings are significantly higher for COPD patients. In some cases, multiple recordings were available for a single patient. We included only the first two. The goal is to have a data set that contains only the relevant data and is as balanced as possible.

All selected recordings were placed in one directory. We created a file called diagnosis.txt containing on each line the patient's ID and a binary indicator of their health status (0 for health and 1 for COPD). Each recording filename includes the patient ID to enable proper association with the diagnostic label. The resulting dataset contains recordings of 121 patients, 61 healthy and 60 diagnosed with COPD. The recordings are saved in .wav files.

In addition to the audio recordings, partial demographic information is available for both datasets. In the first dataset, healthy participants span an age range of 18–81 years (11 female, 24 male), while COPD patients span 42–76 years (1 female, 8 male). For the ICBHI Challenge dataset, demographic metadata including age and sex are provided for most subjects, covering a wide age range from early childhood to over 90 years and including both male and female participants, although some entries contain missing values. Overall, the combined dataset used in this study includes adult subjects of both sexes and a broad age span. However, detailed demographic balance is not uniform and remains a limitation of the available public data. All data are publicly released; we adhered to the original licenses and used only de-identified recordings. No new human data was collected.

4. Preprocessing

4.1 Segmentation

To process the recordings, we used librosa [23], a renowned Python library that excels in audio signal processing. We loaded the .wav files, resulting in an array of numbers, representing the waveform of the recording as shown in Figure 1. To enable a more effective analysis of lung sounds, each patient's audio recording is segmented into shorter, manageable intervals. Specifically, the recordings are divided into uniform segments of 5 seconds each. We perform the segmentation on the one hand because the recordings differ in length from 5 to 90 seconds. On the other hand, having many small audio segments is better for the machine learning algorithm than having fewer larger ones. In addition, dividing the audio recordings into smaller, more manageable segments facilitates feature extraction and enables a more detailed examination of the underlying patterns and characteristics of lung sounds. After segmenting the recordings from the 121 patients in our dataset, we obtain 926 segments, each with a duration of 5 seconds.

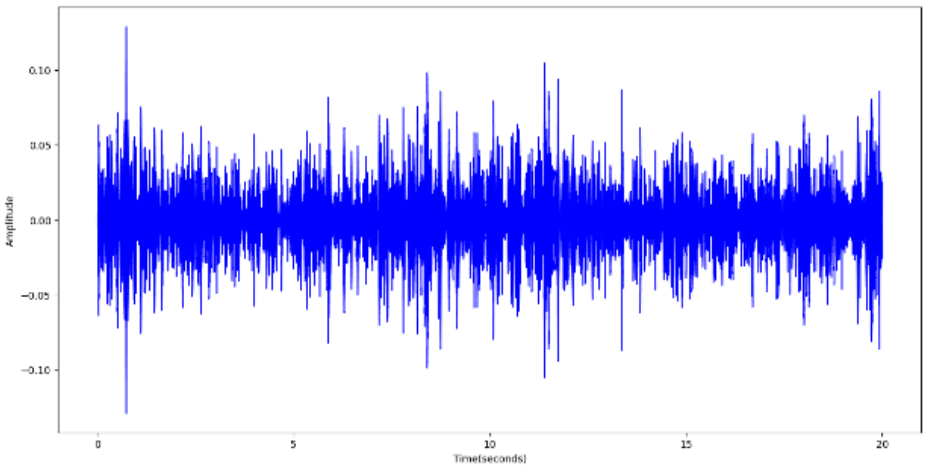

Fig. 1 *Waveform representation of a respiratory sound*

4.2 Feature Extraction

For every 5-second audio segment, a set of 20 Mel-Frequency Cepstral Coefficients (MFCCs) was extracted to represent the spectral characteristics of the signal. Mel-Frequency Cepstral Coefficients (MFCCs) constitute a widely used feature set in audio processing, particularly in the analysis of speech and music. The extraction process (Figure 2) begins by computing a time–frequency representation using the Short-Time Fourier Transform with a window length of 2048 samples and a hop size of 512 samples. The resulting power spectrogram is then projected onto 128 Mel-spaced frequency bands, followed by logarithmic compression and a Discrete Cosine Transform. The first 20 cepstral coefficients, continuous-valued real numbers, are retained, resulting in a concise and stable representation of the signal's spectral characteristics on the Mel scale. They describe the spectral envelope, which reflects how the signal's energy is distributed across various frequency bands.

In our case, for each 5 second segment, we obtain a two-dimensional array of MFCCs, as shown in figure 3. In this representation, the horizontal axis corresponds to time frames, the vertical axis corresponds to the MFCC coefficient index, and the color intensity represents the coefficient magnitude. This matrix-like structure is used as input for the machine learning classifiers. All arrays are of the same length and can be further used in machine learning classification. Finally, all feature vectors were normalized to unit L2 norm to reduce the effect of amplitude variations before classification.

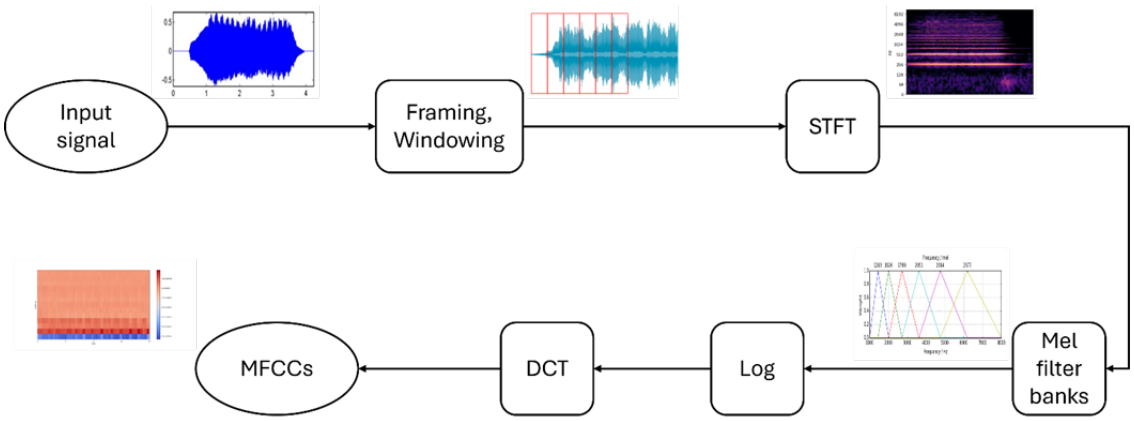

Fig. 2 *Schematic of the MFCC extraction pipeline*

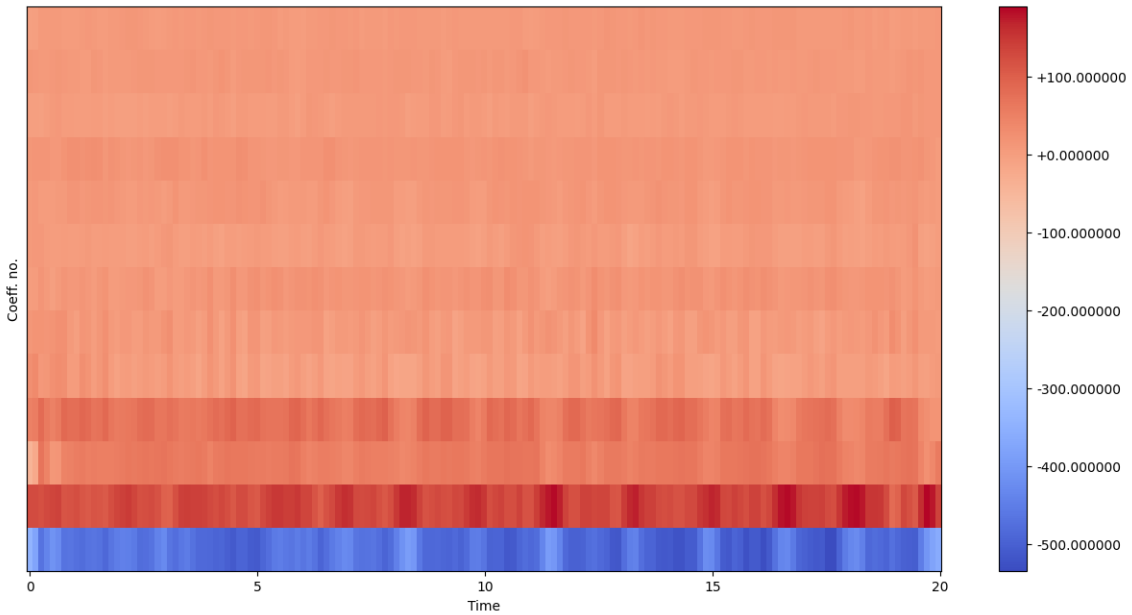

Fig. 3 *MFCCs representation of a respiratory sound*

Int. Journal of Integrated Engineering Vol. 18 No. 4 (2026) p. 130-138

5. Experimental Results

5.2 Experiment Description

We developed and tested three distinct machine learning models to try and automatically categorize these sounds: k-Nearest Neighbors (k-NN), Support Vector Machine (SVM) and Decision Tree. The k-Nearest Neighbors (k-NN) algorithm assigns a class to each audio segment by determining the majority class among its k nearest neighbors in the feature space [24]. The Support Vector Machine (SVM) algorithm constructs a hyperplane in high-dimensional space to separate different classes [25], offering robust performance against overfitting. Finally, the Decision Tree algorithm classifies data by dividing it into subsets based on feature values, forming a tree-like model of decisions that is both intuitive and easy to interpret [26]. Through performance evaluation of these models, we sought to determine the most accurate approach for detecting COPD using sound analysis.

Our previously processed data, which now contains the MFCCs arrays extracted from the recordings of the 121 patients, formed the basis for training three machine learning models. First, the dataset was divided into a training set and a testing set, with a 70/30 ratio, using a strict patient-wise split, so that all segments belonging to a given patient were added either to the training or the testing set, to prevent data leakage. This way we made sure that the models are evaluated only on previously unseen patients, rather than on different segments of the same subject. The 70/30 ratio was picked because due to the limited number of available patients, this split allowed a balanced distribution, both at segment level and at patient level.

Using the prepared training data, we trained each of the three models and saved their outputs in separate files for later analysis and evaluation. The models were then assessed on the same dataset, enabling a comparative examination of their individual strengths and limitations.

5.3 Results

The confusion matrix offers a comprehensive breakdown of the model's performance, indicating true positives (instances correctly classified with COPD), false positives (instances incorrectly classified with COPD), true negatives (instances correctly classified as healthy), and false negatives (instances incorrectly classified as healthy), enabling a thorough evaluation of the model's performance and dependability. Additionally, to better understand the results, we evaluate the model's accuracy, precision, true positive rate or sensitivity, true negative rate or specificity, false positive rate, and false negative rate. Table 1 shows the metrics values.

Table 1 *Model evaluation results at segment-level*

Model	Acc	Precision	Sensitivity (TPR)	Specificity (TNR)	FPR	FNR
KNN	**0.82**	0.84	0.80	0.83	0.17	0.20
SVM	0.75	**0.87**	0.61	**0.90**	**0.10**	0.39
DT	0.77	0.82	**0.82**	0.70	0.30	**0.18**

5.3.1 KNN Results

The KNN model accurately classified 118 true positives and 109 true negatives, resulting in a total accuracy of around 81.65% for the classification of all 278 samples in the test set (Figure 4). The precision of the model is 0.84. The sensitivity of the model is 0.8, meaning that it correctly identifies 80\% of all actual positive cases. The false positive rate of the model is 0.16. TNR was as high as 0.83, showing the model is quite able to differentiate between healthy segments and segments with COPD. The model was trained with many values for the k constant, from 3 to 13. The best values of these metrics were achieved with k equal to 3 and Euclidean distance.

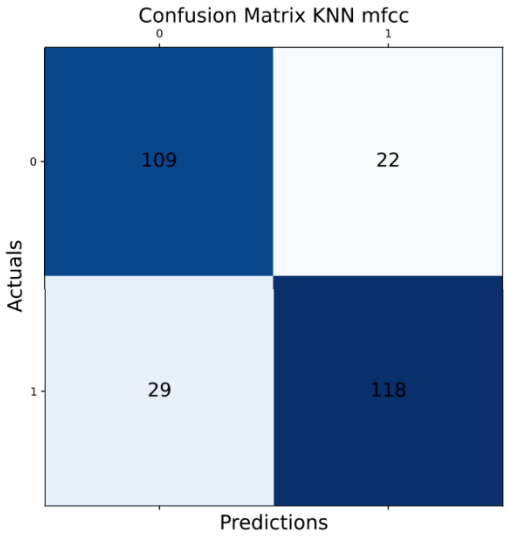

Fig. 4 *Confusion matrix of the results from the KNN algorithm*

5.3.2 SVM Results

The SVM model's performance is lower than KNN. Its confusion matrix indicates 90 true positives and 118 true negatives, achieving an overall accuracy of about 74.82 % for the classification of all samples in the test set (Figure 5). Although it has better precision than the KNN, this model's main issue is its 0.61 true positive rate, which is very low for this metric, making its ability to distinguish between healthy patients and those with COPD questionable. The model favors healthy subjects, as seen in the high TNR and FNR, which also negatively impacts its accuracy. In the case of SVM, the winning hyperparameter configuration was the RBF Kernel with C=1.

5.3.3 Decision Tree Results

The Decision Tree model, as opposed to the SVM, performed better at detecting COPD but also frequently misclassified healthy patients as having COPD. Its confusion matrix indicates 121 true positives and 92 true negatives, producing an overall accuracy of roughly 76.62% for the classification of all 278 samples (Figure 6). With a better TPR of 0.82, the model has demonstrated its ability to detect COPD effectively. However, the false positive rate of 0.29 is somewhat high, suggesting potential for further improvement in this domain. The best results were achieved with the Gini criterion which suggests that some MFCC features are more informative than others, and the decision tree can exploit this by prioritizing splits on the most useful features.

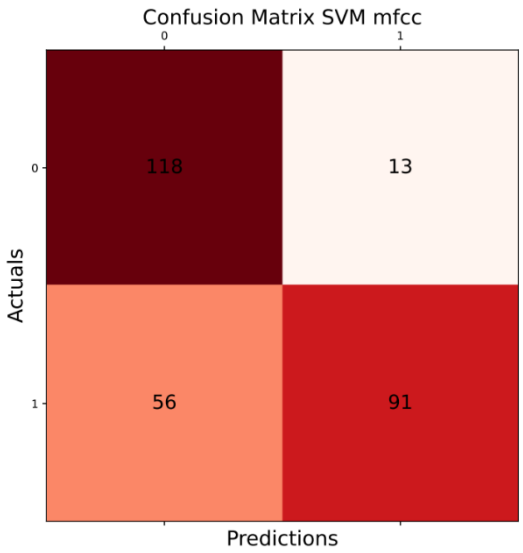

Fig. 5 *Confusion matrix of the results from the SVM algorithm*

Int. Journal of Integrated Engineering Vol. 18 No. 4 (2026) p. 130-138

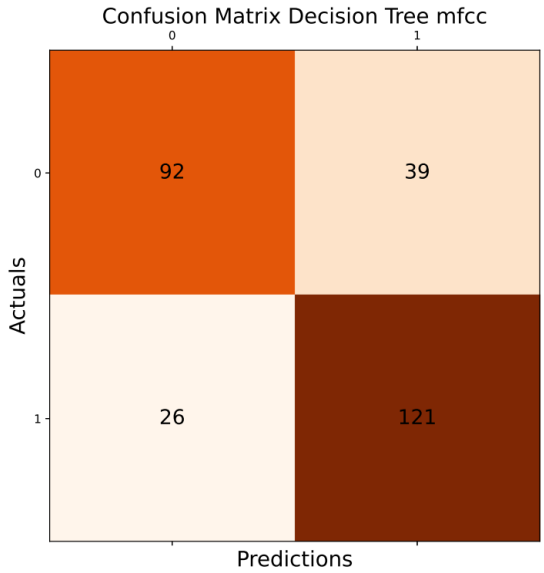

Fig. 6 *Confusion matrix of the results from the decision tree algorithm*

5.3.4 Patient-Level Evaluation

Since COPD screening is a patient-level decision task, we additionally evaluated the proposed system at patient level. We aggregated segment-level predictions by computing, for each patient, the proportion of segments classified as COPD. This value lies between 0 and 1 and represents the fraction of segments predicted as positive. The final patient label was obtained by comparing this proportion to a decision threshold. For example, if 7 out of 10 segments of a patient were classified as COPD, the aggregated score for that patient is 0.7. Furthermore, to account for potential class imbalance and to optimize the clinical trade-off between sensitivity and specificity, we performed a threshold optimization. We varied the decision threshold from 0 to 1 with a step of 0.05. For each value, we computed the balanced accuracy, defined as (TPR+TNR)/2. The threshold with the highest balanced accuracy was selected for each model. As can be observed in Table 2, the k-NN classifier maintains the most balanced performance at patient level, achieving 80% accuracy with comparable sensitivity and specificity. The Decision Tree favors sensitivity, detecting most COPD patients but at the cost of lower specificity, while the SVM remains conservative, achieving high specificity but missing a significant proportion of COPD cases.

Table 2 *Model evaluation results at patient-level*

Model	Acc	Precision	Sensitivity (TPR)	Specificity (TNR)	Threshold
KNN	**0.80**	**0.80**	0.81	0.78	0.40
SVM	0.70	0.76	0.59	**0.81**	0.30
DT	0.76	0.72	**0.87**	0.64	0.50

5.4 Discussion

All three models show interesting results. The SVM and Decision Tree have similar accuracy values but perform oppositely in other metrics. One misclassifies healthy subjects as having COPD, while the other misclassifies COPD patients as healthy. KNN, on the other hand, is more balanced, providing a reliable tool for detecting COPD in patients who have the condition. Although an accuracy slightly above 80% might not seem impressive, considering the limited amount of training data and the low complexity of the model, we believe this is a very good result. The low sensitivity of the SVM model suggests that the decision boundary learned by the RBF kernel favors the majority structure of the healthy class, potentially due to overlap in the MFCC feature space and the limited number of training patients. In contrast, k-NN benefits from local neighborhood structure, which appears better suited for this dataset.

When evaluated at patient level using optimized decision thresholds, the performance trends remain consistent with the segment-level analysis. The k-NN classifier continues to offer the best balance between sensitivity and specificity, making it the most suitable candidate for screening applications. The Decision Tree achieves the highest sensitivity (0.87), which is desirable for screening, but produces more false positives. In contrast, SVM exhibits high specificity but insufficient sensitivity, which limits its suitability for a screening scenario where missing true cases is critical.

In this study, we have shown that aggregating predictions at patient levels using thresholder proportion of positive segments and threshold optimization leads to more clinically meaningful performance estimates. Also, as suggested in other studies, another way to improve the model could be through data augmentation. In addition, the accuracy of the screening results can improve over time as the dataset naturally grows. This means that if our solution is used, it is expected to become more effective over time, offering users a more dependable and accurate tool for COPD detection.

6. Conclusion

The solution we propose in this paper has been specifically designed for the detection of COPD, providing enhanced insights and screening precision compared to conventional respiratory assessment tools. By addressing the distinct features of COPD, it delivers a more extensive and precise approach for both patients and healthcare practitioners. By integrating machine learning with lung sound analysis, we developed a system that is simple, affordable, and highly accurate. In addition, patient-level evaluation using fraction of COPD-positive segments and balanced-accuracy-based threshold optimization confirms that the proposed k-NN approach provides a robust and clinically meaningful screening performance. Both medical practitioners and patients can use the system effectively. Although not perfect, this technology is helping broaden access to advanced respiratory screenings in rural and underserved regions, while simultaneously lowering healthcare expenses for both patients and providers.

Acknowledgement

The authors declare that they have nobody or no-company to acknowledge.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Constantin Constantinescu, Remus Brad;* ***data collection:*** *Constantin Constantinescu;* ***analysis and interpretation of results:*** *Constantin Constantinescu;* ***draft manuscript preparation:*** *Constantin Constantinescu, Remus Brad. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] World Health Organization. *Chronic obstructive pulmonary disease (COPD) – fact sheet.* Available at: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) (Accessed: 09.01.2026).

[2] I. Pino Peña et al., "Automatic emphysema detection using weakly labeled HRCT lung images," PLoS One, vol. 13, no. 10, p. e0205397, 2018.

[3] Z. Li, K. Huang, L. Liu, and Z. Zhang, "Early detection of COPD based on graph convolutional network and small and weakly labeled data," Med Biol Eng Comput, vol. 60, no. 8, pp. 2321–2333, 2022.

[4] J. Sun et al., "Detection and staging of chronic obstructive pulmonary disease using a computed tomography–based weakly supervised deep learning approach," Eur Radiol, vol. 32, no. 8, pp. 5319–5329, 2022.

[5] Y. Wu et al., "Attention-guided multiple instance learning for COPD identification: To combine the intensity and morphology," Biocybern Biomed Eng, vol. 43, no. 3, pp. 568–585, 2023, doi: https://doi.org/10.1016/j.bbe.2023.06.004.

[6] K. K. Agrawal et al., "A Novel Approach for Chronic Obstructive Pulmonary Disease Diagnosis with TensorFlow-Based Image Analysis," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 20s, pp. 713–719, Mar. 2024, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/5234

[7] S. Bhattacharjee, B. Saha, P. Bhattacharyya, and S. Saha, "Classification of obstructive and non-obstructive pulmonary diseases on the basis of spirometry using machine learning techniques," J Comput Sci, vol. 63, p. 101768, 2022, doi: https://doi.org/10.1016/j.jocs.2022.101768.

[8] M. A. Fernández-Granero, D. Sánchez-Morillo, A. León-Jiménez, and L. F. Crespo, "Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms," Biomed Mater Eng, vol. 24, no. 6, pp. 3825–3832, 2014.

[9] L. van Bemmel, W. Harmsen, C. Cucchiarini, and H. Strik, "Automatic Selection of the Most Characterizing Features for Detecting COPD in Speech," in Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings 23, 2021, pp. 737–748.

[10] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease," PeerJ Comput Sci, vol. 7, p. e369, 2021.

[11] B. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. Kahya, et al., "An open access database for the evaluation of respiratory sound classification algorithms," Physiol Meas, vol. 40, Mar. 2019, doi: 10.1088/1361-6579/ab03ea.

[12] G. Altan, Y. Kutlu, and N. Allahverdi, "Deep Learning on Computerized Analysis of Chronic Obstructive Pulmonary Disease," IEEE J Biomed Health Inform, vol. 24, no. 5, pp. 1344–1350, 2020, doi: 10.1109/JBHI.2019.2931395.

[13] G. Altan, Y. Kutlu, A. Ö. Pekmezci, and S. Nural, "Deep learning with 3D-second order difference plot on respiratory sounds," Biomed Signal Process Control, vol. 45, pp. 58–69, 2018, doi: https://doi.org/10.1016/j.bspc.2018.05.014.

[14] G. Altan, Y. Kutlu, and A. Gökçen, "Chronic obstructive pulmonary disease severity analysis using deep learning onmulti-channel lung sounds," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 28, no. 5, pp. 2979–2996, 2020.

[15] 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

[16] N. S. Haider and A. K. Behera, "Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD)," Biocybern Biomed Eng, vol. 42, no. 1, pp. 42–59, 2022, doi: https://doi.org/10.1016/j.bbe.2021.12.004.

[17] Mohd Amin, A. T. ., Mokri, S. S. ., Ahmad, R. ., Ismail, F. ., & Abd Rahni, A. A. . (2021). Evaluation Methodology for Respiratory Signal Extraction from Clinical Cone-Beam CT (CBCT) using Data-Driven Methods. International Journal of Integrated Engineering, 13(5), 1-8.

[18] Zainal Bakri, S. F., Hariri, A., Ismail, M., Abdullah, S., & Kassim, N. I. (2018). Evaluation of Respiratory Symptoms, Spirometric Lung Patterns and Metal Fume Concentrations among Welders in Indoor Air-Conditioned Building at Malaysia. International Journal of Integrated Engineering, 10(5). https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/3038

[19] Norali, A. N. (2017). Human Breathing Classification Using Electromyography Signal with Features Based on Mel-Frequency Cepstral Coefficients. International Journal of Integrated Engineering, 9(4). https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/2019

[20] K. M. Le, "RESAPP-A RESUME MOBILE APPLCIATION," 2022, [Online]. Available: https://scholarworks.calstate.edu/downloads/r494vr592

[21] 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: https://doi.org/10.1016/j.dib.2021.106913.

[22] "Icbhi 2017 challenge," https://bhichallenge.med.auth.gr/ICBHI%202017%20Challenge, accessed: 2025-05-11.

[23] B. McFee, M. McVicar, D. Faronbi, I. Roman, M. Gover et al., "librosa/librosa: 0.10.2.post1," May 2024, Zenodo. doi: 10.5281/zenodo.11192913.

[24] E. Fix and J. L. Hodges, "Discriminatory analysis, nonparametric discrimination," 1951.

[25] V. Vapnik, The nature of statistical learning theory. Springer science & business media, 2013.

[26] J. R. Quinlan, "Induction of decision trees," Mach Learn, vol. 1, pp. 81–106, 1986.