



Automatic Lung Health Screening Using Respiratory Sounds

Himadri Mukherjee¹ · Priyanka Sreerama² · Ankita Dhar¹ · Sk. Md. Obaidullah³ · Kaushik Roy¹ · Mufti Mahmud⁴ · K.C. Santosh²

Received: 27 October 2020 / Accepted: 23 November 2020 / Published online: 11 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Significant changes have been made on audio-based technologies over years in several different fields. Healthcare is no exception. One of such avenues is health screening based on respiratory sounds. In this paper, we developed a tool to detect respiratory sounds that come from respiratory infection carrying patients. Linear Predictive Cepstral Coefficient (LPCC)-based features were used to characterize such audio clips. With Multilayer Perceptron (MLP)-based classifier, in our experiment, we achieved the highest possible accuracy of 99.22% that was tested on a publicly available respiratory sounds dataset (ICBHI17) (Rocha et al. *Physiol. Meas.* 40(3):035,001 2019) of size 6800+ clips. In addition to other popular machine learning classifiers, our results outperformed common works that exist in the literature.

Keywords Lung health · Respiratory infection · Respiratory sound · Healthcare

Introduction

Respiratory diseases are the third leading cause of death worldwide. As rapid growth of respiratory diseases is witnessed around the world, medical research field has gained interest in integrating potential audio signal analysis-based technique. Like in other application domains, audio signal analysis tools can potentially help in analyzing respiratory sounds to detect problems in the respiratory tract. Audio analysis aids in timely diagnosis of respiratory ailments more effortlessly in the early stages of a respiratory dysfunction. Respiratory conditions are diagnosed through spirometry and lung auscultation. Even though, spirometry is one of the most commonly available lung function tests, it is limited to patient's cooperation. As a result, it is error prone. Auscultation is a technique that involves listening to the internal human body sounds with the aid of a stethoscope. Over several years, it has been an effective tool to analyze lung disorders and/or abnormalities. Such

procedure is limited to trained physicians. Besides, for various reasons (e.g., faulty instrument), false positives can happen. Therefore, it opens a door to develop computerized respiratory sound analysis tools/techniques, where automation is integral.

Lung sounds are difficult to analyze and distinguish because they are non-stationary and non-linear signals. Automated analysis was made possible with the use of electronic stethoscope. In 2017, the largest publicly available respiratory sound database was compiled and encouraged the development of algorithms that can identify common abnormal breath sounds (wheezes and crackles) from clinical and nonclinical settings. Respiratory sounds are generally classified as normal or adventitious. Adventitious sounds are RS superimposed on normal respiratory sounds, which can be crackles or wheezes. Crackles are discontinuous sounds, explosive, and non-musical. that are typically less than 20 ms that occur frequently in cardiorespiratory diseases associated with lung fibrosis (fine crackles) or chronic airway obstruction (coarse crackles). Wheezes are high pitched sounds that last more than 100 ms. They are common in patients with obstructive airway diseases and indicate obstructive airway conditions, such as asthma and COPD. The dataset contains respiratory cycles that were recorded and annotated by professionals as wheezes, crackles, both, or no abnormal sounds.

Rao et al. [19] discussed acoustic techniques for pulmonary analysis. They studied acoustic aspects for different lung diseases. It includes different type of sounds in the

This article belongs to the Topical Collection: *Image & Signal Processing*

H Mukherjee and KC Santosh contributed equally.

✉ K.C. Santosh
santosh.kc@usd.edu

Extended author information available on the last page of the article.

thick of internal and external sounds. Aykanat et al. [3] presented a convolutional network plus mel frequency cepstral coefficient-support vector machine-based approach for lung sound classification. On a dataset of 17930 sounds from 1630 subjects, an accuracy of 86% (for healthy-pathological classification) was reported. Pramono et al. [18] classified normal respiratory sounds and wheezes on a dataset of 38 recordings. Of 425 events, 223 were wheezes and the rest were normal. They reported a AUC value of 0.8919 with MFCC-based features. Acharya et al. [1] presented a deep learning-based approach for lung sound classification. They reported an accuracy of 71.81% on the ICBHI17 dataset of size 6800+ clips. Dokur [10] used machine learning approaches to distinguish respiratory sounds. In their experiments, nine different categories from 36 patients were used. An accuracy of 92% was reported by using Multilayer Perceptron (MLP).

Melbye et al. [14] studied the classification of lung sounds by 12 observers. They worked with 1 clip each from 10 adults and children and obtained fleiss kappa values of 0.62 and 0.59 for crackles and wheezes, respectively. Among the 20 cases, they found that in 17 cases, the observers concluded presence of at least 1 adventitious sound. Bahoura and Pelletier [4] used cepstral features to distinguish normal and wheezing sounds. They worked with 12 instances from each class and reported the highest true positive value of 76.6% for wheezing sounds. They also reported 90.6% true positives for normal sounds with fourier transform-based features. Ma et al. [13] developed a system to distinguish lung sounds using a resnet-based approach. On ICBHI17 dataset, an accuracy of 52.26% was reported. Emmanouilidou et al. [11] proposed a robust approach to identify lung sounds in the presence of noise. In their experiments, with 1K+ volunteers (over 250 hours of data), an accuracy of 86.7% was reported.

To analyze lung sounds, Sen et al. [23] used Gaussian mixture model and support vector machine-based classifier. Using 20 healthy and non-healthy subjects, they reported an accuracy of 85%. Demir et al. [9] used a CNN-based approach. On ICBHI17 dataset, the highest accuracy of 83.2% was reported. Chen et al. [7] used a S-transform-based approach coupled with deep residual networks to classify lung sounds: crackle, wheeze, and normal. In their study, the reported accuracy was 98.79%. Kok et al. [12] employed multiple features, such as MFCC, DWT, and time domain metrics to distinguish healthy and non-healthy sounds. In their study, they reported accuracy, specificity, and sensitivity values of 87.1%, 93.6%, and 86.8%, respectively on the ICBHI17 dataset.

Chambers et al. [6] developed a tool to identify healthy/non-healthy patients using respiratory sounds. They used several spectral, rhythm, SFX, and tonal features coupled with decision tree-based classification. In their study, they

reported an accuracy of 85% on a dataset of 920 records. Altan et al. [2] developed a deep learning-based approach to detect chronic obstructive pulmonary disease. Their tool used Hilbert-Huang transform on multi-channel lung sounds. In their experiment, an accuracy of 93.67% was reported on a dataset of 600 sounds collected from 50 patients. Cohen and Landsberg [8] classified 7 different type of sounds using linear predictive coefficient-based technique. In their experiments, out of 105 instances, 100 were classified correctly.

Even though there exists a rich state-of-the-art literature for lung sound analysis, they do not guarantee optimal performance. Moreover, non-healthy cases are composed of several issues/criteria. Distinguishing healthy sounds from non-healthy sounds is not trivial. Handcrafted feature-based systems are preferred over deep learning-based systems, where computational resource is considered. Secondly, prior to deeper analysis of non-healthy sounds, it is essential to distinguish healthy and non-healthy sounds. A hierarchical approach can aid to reduce the workload of medical experts in resource-constrained regions. After ensuring that whether a person has lung infection, the true positive case can be taken for further treatment(s)/processing.

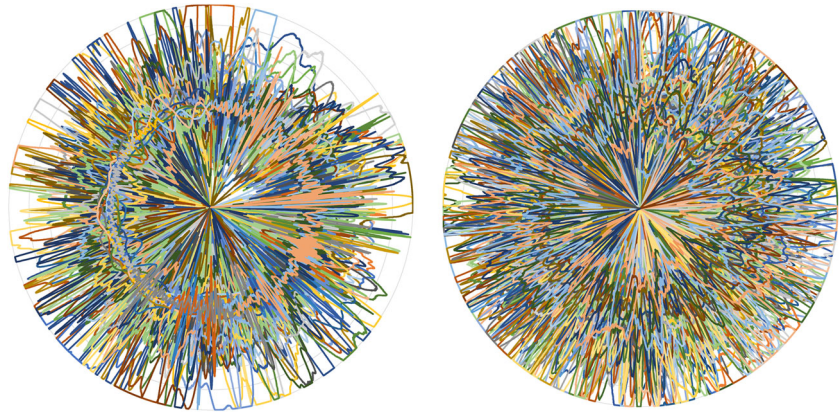
In this paper, we developed an automated tool, where LPCC-based features are employed. LPCC-based features were chosen due to its ability of modeling a variety of audio signals [15, 16]. In our experiments, on a dataset ICBHI17 (of size 6800+ clips), we achieved an accuracy of 99.22% using MLP.

The remainder of the paper is organized as follows. “[Dataset description](#)” discusses on dataset. In “[Proposed method: LPCC-based features and MLP](#)”, we describe the proposed tool. Experimental results are provided in “[Results and analysis](#)”. We conclude the paper in “[Conclusion](#)”.

Dataset description

To develop of a robust system, it is important to ensure that the dataset mimics real-world problems. Our system was trained on a publicly available respiratory sound database [20], which is associated with the International Conference on Biomedical and Health Informatics (ICBHI). Most of the database consists of audio samples recorded by the School of Health Sciences, University of Aveiro (ESSUA) research team at the Respiratory Research and Rehabilitation Laboratory (Lab3R), ESSUA and at Hospital Infante D. Pedro, Aveiro, Portugal. The second research team, from the Aristotle University of Thessaloniki (AUTH) and the University of Coimbra (UC), acquired respiratory sounds at the Papanikolaou General Hospital, Thessaloniki and at the General Hospital of Imathia (Health Unit of Naousa), Greece.

Fig. 1 200 audio clips (original): healthy class (left) and non-healthy class (right)



To collect data, disparate stethoscopes and microphones were used. The audios were recorded from the trachea and 6 other chest locations: left and right posterior, anterior, and lateral. The audios were collected in both clinical and non-clinical settings from adult participants of disparate ages. Participants encompassed patients with lower and upper respiratory tract infections, pneumonia, bronchiolitis, COPD, asthma, bronchiectasis, and cystic fibrosis.

The ICBHI database consists of 920 audio samples from 126 subjects. These are annotated by respiratory experts, and used as a benchmark in the field. Each respiratory cycle in the dataset is annotated amidst 4 classes. The annotations basically cover 2 broad groups: healthy and non-healthy. The non-healthy category is further divided into wheeze and crackle with some cycles having both issues. Among 6898 cycles totaling to 5.5 hours, 1864 cycles have crackles while 886 have wheezes. There are 506 cycles, which have both wheezes and crackles.

While recording, the participants were seated. The acquisition of respiratory sounds was performed on adult and elderly patients. Many patients had COPD with comorbidities (e.g., heart failure, diabetes, and hypertension). Further, noise exists, such as rubbing sound of the stethoscope with the patient’s dress, and background talking. Such varieties in the data made it challenging to identify problems in the respiratory sounds. One of the most challenging aspects of the audio clips was the presence of heartbeat sound along with the breath sounds. No preprocessing was performed to remove the heartbeat sounds.

For better understanding, visual representations of 200 audio clips from the healthy and non-healthy sounds are shown in Fig. 1. In Table 1, a complete dataset is provided.

Table 1 Respiratory sound database [20]

| Clip type | Number of clips |
|-------------|-----------------|
| Healthy | 3642 |
| Non-healthy | 3256 |

Proposed method: LPCC-based features and MLP

Respiratory sound representation: LPCC-based features

As audio clip contains high deviations across its entire length, its analysis is not trivial. Therefore, each audio clip is broken down into smaller segments called frames to facilitate analysis. In our study, we divided each clip into frames consisting of 256 sample points with a 100-point overlap in between them. The parameters were empirically designed. The same 200 audio clips (as in Fig. 1) are shown in Fig. 2 after framing. The number of S_z sized overlapping frames O_f with O overlapping points for a signal having S points is presented below:

$$O_f = \left\lceil \frac{S - S_z}{O} + 1 \right\rceil. \tag{1}$$

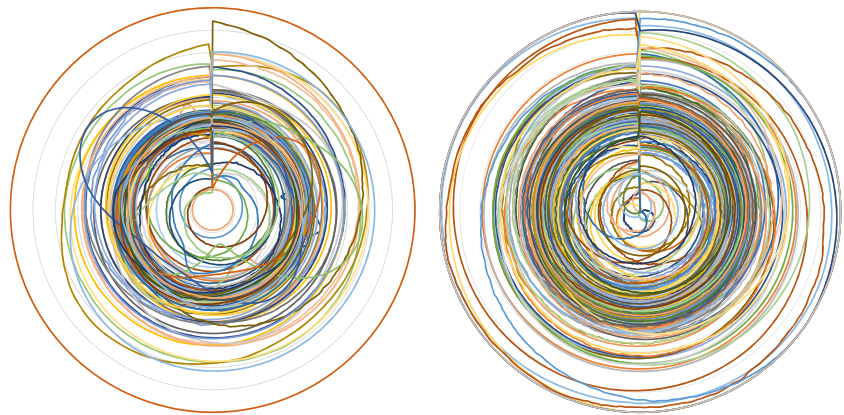
After framing audio clips (into shorter segments), it was observed that in various instances the starting and ending points were not aligned in a frame. These discontinuities/jitters lead to smearing of power across the frequency spectrum. This posed a problem in the form of spectral leakage during frequency domain analysis which produced additional frequency components. To tackle this, the frames were subjected to a window function. Hamming window was selected for this purpose due to its efficacy as reported in [16]. The same frames (Fig. 2) are presented in Fig. 3 after windowing. The hamming window is mathematically illustrated as

$$A(z) = 0.54 - 0.46 \cos \left(\frac{2\pi z}{S_z - 1} \right), \tag{2}$$

where $A(z)$ is the hamming window function and z is a point within a frame.

Thereafter, we performed Linear Predictive Coefficient (LPC) based analysis [15] on each of them. The previous

Fig. 2 200 audio clips (as in Fig. 1) after framing: healthy class (left) and non-healthy class (right)



P samples are used to present the r^{th} sample in a signal $s()$ as

$$s(r) \approx p_1s(r - 1) + p_2s(r - 2) + p_3s(r - 3) + \dots + p_Ps(r - P), \tag{3}$$

where p_1, p_2, \dots, p_P are the LPCs or predictors. The error of this prediction $E(r)$ bounded by the actual and predicted samples: ($s(r)$ and $\hat{s}(r)$) can be explained as

$$E(r) = s(r) - \hat{s}(r) = s(r) - \sum_{k=1}^P p_k s(r - k). \tag{4}$$

The error of sum of squared differences (as shown below) is minimized to generate the unique predictors for a x sized frame, which can be expressed as

$$E_r = \sum_x \left[s_r(x) - \sum_{k=1}^P p_k s_r(x - k) \right]^2. \tag{5}$$

Thereafter, a recursive technique is used to compute the Cepstral coefficients (C), which is expressed as

$$\left. \begin{aligned} C_0 &= \log_e P \\ C_r &= p_r + \sum_{q=1}^{r-1} \frac{q}{r} C_q p_{r-q}, \text{ for } 1 < r \leq P \text{ and} \\ C_{r0} &= \sum_{q=r-P}^{r-1} \frac{q}{r} C_q p_{r-q}, \text{ for } r > P \end{aligned} \right\}. \tag{6}$$

Since clips in the dataset were of unequal lengths and number of frames obtained varied. When features were extracted in frame level, it produced different dimensions. To handle this, we performed two operations: a) grading and b) standard deviation measurement.

1. Firstly, the sum of LPCC coefficients in each of the frequency ranges (bands) across all the frames was computed. Based on the sum of these energy values, bands were graded in an ascending order. This sequence

of band numbers was used as features that helped in identifying dominance of different bands for the clips from various categories.

2. Secondly, standard deviation was computed for every band. These two metrics were stacked to form the feature, which is independent of the clip length. 10, 20, 30, 40 and 50 dimensional features were extracted for the 2 classes. The trend of the 30 dimensional feature values (best result) for the 2 classes is shown in Fig. 4.

Classification: MLP

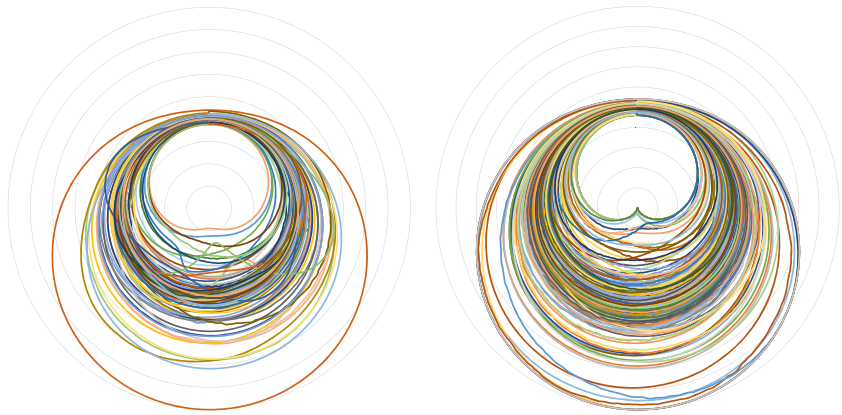
We employed MLP classifier – feed-forward artificial neural network – for classification purpose [17]. Feedforward neural networks are made up of the input layer, output layer and hidden layer. It is a supervised learning algorithm trained on a dataset using a function $f() : Z_n \rightarrow Z_o$, where n and o represent the dimensions for input and output. For a given set of features $P = p_1, p_2, \dots, p_n$ and aim x , a non-linear function is learned for classification. The difference between MLP and logistic regression lies in the existence of one or more non-linear layers (hidden layers) between the input and the output layer. MLP consists of three or more layers (input layer, output layer and one or more hidden layers) of non-linear activating neurons. The number of hidden layers can be increased according to the requirement of developing a model to accomplish certain task.

The initial layer is the input layer which comprises of a set of neurons $\{p_i \mid p_1, p_2, \dots, p_n\}$ denoting the features. Each neuron of the hidden layer modifies the values from the previous layer using sum of weights as $w_1 p_1 + w_2 p_2 + \dots + w_n p_n$.

The activation function that represents the relationship between input and output layer in of non-linear nature. It makes the model flexible in defining unpredictable relationships. The activation function can be expressed as

$$y_i = \tanh(w_i) \text{ and } y_i = (1 + e^{w_i})^{-1}, \tag{7}$$

Fig. 3 Representation of 200 audio clips (as in Fig. 1) after windowing: healthy class (left) and non-healthy class (right)



where y_i and w_i denotes the outcome of the i^{th} neuron and weighted sum of the input features. The values from the ultimate hidden layer are provided to the output layer as output values. Each layer of MLP contains several fully connected layers as each neuron in a layer is attached to all the neurons of the previous layer. The parameters of each neuron are independent of the remaining neurons of the layer ensuring possession of unique set of weights. The initial momentum and learning rate were set to 0.2 and 0.3 respectively.

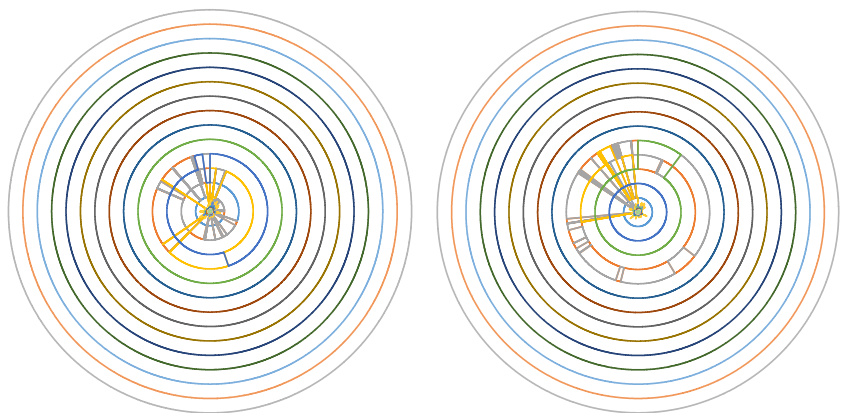
Results and analysis

Evaluation metrics and protocol

Accuracy is not enough to measure the performance of any system. It is also much important to analyze the disparate misclassifications. Hence, to evaluate our tool, the following performance metrics are used: Precision, Accuracy, Sensitivity (Recall), Specificity, and Area under ROC curve (AUC). They are computed as

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N},$$

Fig. 4 Representation of 30 dimensional features for the audio clips: healthy class (left); non-healthy class (right)



$$Precision = \frac{T_P}{T_P + F_P},$$

$$Sensitivity (Recall) = \frac{T_P}{T_P + F_N},$$

$$Specificity = \frac{T_N}{T_N + F_P}, \text{ and}$$

$$F1 \text{ score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

where T_P , T_N , F_P , and F_N refer to true positive, true negative, false positive, and false negative, respectively.

To avoid possible bias in evaluation, 5-fold cross validation was used.

Our results

The performance of the different features are provided in Table 2. It is observed that the best result was obtained with 30 dimensional features and it's corresponding confusion matrix is provided in Table 3.

Next, the momentum was varied from 0.1 to 0.5 with a step of 0.1, and results are provided in Table 4. The best result was obtained for a momentum of 0.1 whose inter-class confusions are provided in Table 5. As compared to the default scenario, there were 4 more misclassifications in

Table 2 Performance of different feature dimensions using MLP

| Feature dim. | Accuracy(%) |
|--------------|-------------|
| 10 | 93.91 |
| 20 | 90.01 |
| 30 | 99.07 |
| 40 | 89.19 |
| 50 | 98.78 |

the case of the healthy cases (and 9 less misclassifications for the non-healthy cases).

Finally, the momentum was varied from 0.1-0.6 with a step of 0.1 whose results are provided in Table 6. In our experiment, the highest performance was obtained when a learning rate of 0.5 was selected. We presented a confusion matrix for this setup in Table 7. It is observed that the number of misclassifications for both classes was reduced as compared to the initial setup. The misclassified instances were analyzed, and it was found that many of them had heartbeat sounds. Along with this, other unwanted artefacts, such as talking and movement of the probe helped in misclassifying.

It is observed that the misclassified instances was reduced by almost 15.63% as compared to the original setup using default settings. As compared to best result, after momentum tuned, a decrease of nearly 8.47% occurred for the misclassified instances.

A deeper analysis of the misclassifications revealed that approximately 0.74% of the healthy cases were misclassified as opposed to non-healthy. In the case of non-healthy instances, approximately 0.83% of the clips were misclassified as healthy, which we call false negative.

The different performance metrics were computed for the default setup, best momentum, and best learning rate (overall highest). Such results are provided in Table 8. The ROC curves for these scenarios are shown in Fig. 5.

Comparative study

The performance of several other classifiers was compared in order to establish the efficacy of MLP. For comparison, the 30 dimensional feature set (best performance) was chosen. We experimented with BayesNet, SVM, RNN,

Table 3 Inter-class confusions for the 30 dimensional features (Best result) using MLP

| | Healthy | Non-healthy |
|-------------|---------|-------------|
| Healthy | 3611 | 31 |
| Non-healthy | 33 | 3223 |

Table 4 Performance for different momentum values on 30 dimensional features with learning rate of 0.3

| Momentum | Accuracy(%) |
|----------|-------------|
| 0.1 | 99.14 |
| 0.2 | 99.07 |
| 0.3 | 99.04 |
| 0.4 | 99.07 |
| 0.5 | 99.12 |

Table 5 Inter-class confusions for momentum value of 0.1 on 30 dimensional features

| | Healthy | Non-healthy |
|-------------|---------|-------------|
| Healthy | 3607 | 35 |
| Non-healthy | 24 | 3232 |

Table 6 Performance for different learning rates with momentum of 0.2

| Learning rate | Accuracy(%) |
|---------------|-------------|
| 0.1 | 99.03 |
| 0.2 | 99.13 |
| 0.3 | 99.07 |
| 0.4 | 99.06 |
| 0.5 | 99.22 |
| 0.6 | 99.13 |

Table 7 Interclass confusions for learning rate of 0.5 and momentum of 0.2 on 30 dimensional features

| | Healthy | non-healthy |
|-------------|---------|-------------|
| Healthy | 3615 | 27 |
| non-healthy | 27 | 3229 |

Table 8 Performance metrics for default scenario, best results after tuning momentum value and best result after tuning learning rate

| Metrics | Default | Best momentum | Best learning rate |
|---------------------|---------|---------------|--------------------|
| Sensitivity | 0.9915 | 0.9904 | 0.9917 |
| Specificity | 0.9899 | 0.9926 | 0.9926 |
| Precision | 0.9909 | 0.9834 | 0.9917 |
| False positive rate | 0.0101 | 0.0074 | 0.0074 |
| False negative rate | 0.0085 | 0.0096 | 0.0083 |
| Accuracy(%) | 99.07 | 99.14 | 99.22 |
| F1 score | 0.9912 | 0.9919 | 0.9917 |
| AUC | 0.9994 | 0.9995 | 0.9993 |

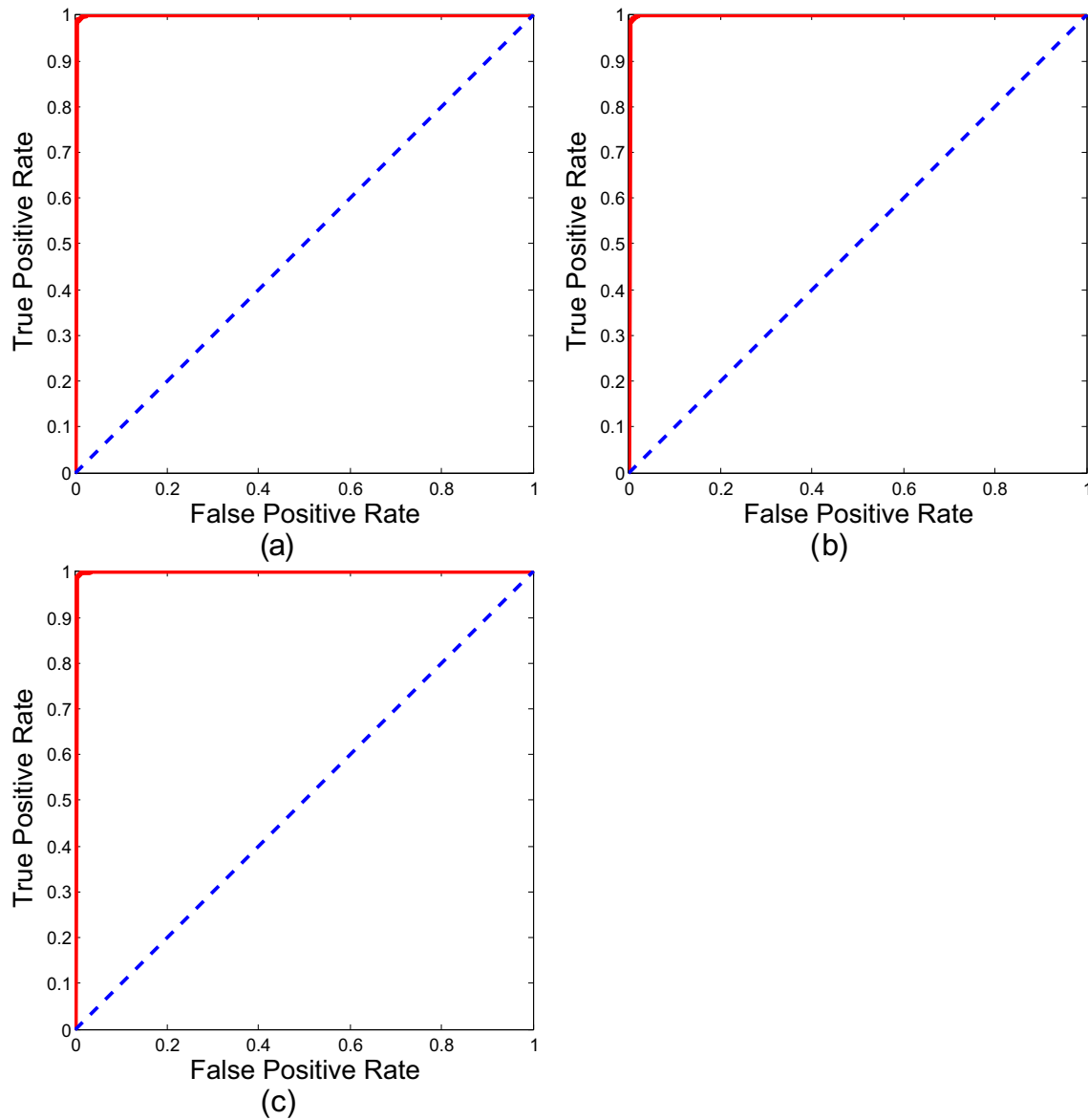


Fig. 5 ROC curves: **a** default settings, **b** best momentum value (0.1), and **c** best learning rate (0.5, overall highest result)

Table 9 Performance of different classifiers on the 30 dimensional features

| Classifier | Accuracy(%) |
|-----------------------|-------------|
| BayesNet | 98.26 |
| Naïve Bayes | 88.98 |
| SVM | 98.59 |
| RBF Network | 95.82 |
| LibLINEAR | 98.59 |
| Simple Logistic | 98.70 |
| Decision Table | 98.62 |
| RNN | 93.82 |
| Multilayer Perceptron | 99.22 |

Naive Bayes, RBF network, Decision Table, LibLINEAR, and Simple logistic. The results are provided in Table 9.

We also compared the performance of our system with reported works by Kok et al. [12] and Chambers et al. [6]. The average accuracies for both the systems along with the proposed system are provided in Table 10.

Table 10 Comparison with reported works

| Work | Accuracy(%) |
|---------------------|-------------|
| Kok et al. [12] | 87.10 |
| Chambers et al. [6] | 85.00 |
| Proposed technique | 99.22 |

Conclusion

In this paper, we developed a tool to detect respiratory sounds that come from respiratory infection carrying patients. We have employed Linear Predictive Cepstral Coefficient (LPCC)-based features to characterize respiratory sounds. With Multilayer Perceptron (MLP)-based classifier, in our experiment, we have achieved the highest possible accuracy of 99.22% (AUC = 0.9993) on a publicly available dataset of size 6800+ clips. In addition to other popular machine learning classifiers, our results outperformed common works that exist in the literature.

Not limiting to binary classification (health/non-healthy), our immediate plan is to classify disease types from non-healthy category. This will help identify the nature and severity of infection. As we observed that COVID-19 could possibly be screened by analyzing respiratory sound [5], we are now extending our experiments on COVID-19 [21, 22].

Compliance with Ethical Standards

Conflict of interests Authors declared no conflicts of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.


References

- Acharya, J., and Basu, A., Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE Trans. Biomed. Circ. Syst.* 14(3):535–544. <https://doi.org/10.1109/TBCAS.2020.2981172>, 2020. <https://ieeexplore.ieee.org/document/9040275>.
- Altan, G., Kutlu, Y., and Allahverdi, N., Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE J. Biomed. Health Inform.* 24(5):1344–1350. <https://doi.org/10.1109/JBHI.2019.2931395>, 2019. <https://ieeexplore.ieee.org/document/8777195>.
- Aykanat, M., Kilić, Ö., Kurt, B., and Saryal, S., Classification of lung sounds using convolutional neural networks. *EURASIP J. Image Video Process.* 2017(1):65. <https://doi.org/10.1186/s13640-017-0213-2>, 2017.
- Bahoura, M., and Pelletier, C., New parameters for respiratory sound classification, 2003. <https://doi.org/10.1109/CCECE.2003.1226178>. <https://ieeexplore.ieee.org/document/1226178>.
- Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P., and Mascolo, C., Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3474–3484. New York: Association for Computing Machinery, 2020. <https://doi.org/10.1145/3394486.3412865>.
- Chambres, G., Hanna, P., and Desainte-Catherine, M., Automatic detection of patient with respiratory diseases using lung sound analysis. In: *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6: IEEE, 2018. <https://ieeexplore.ieee.org/document/8516489>. <https://doi.org/10.1109/CBMI.2018.8516489>.
- Chen, H., Yuan, X., Pei, Z., Li, M., and Li, J., Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access* 7:32,845–32,852. <https://doi.org/10.1109/ACCESS.2019.2903859>, 2019. <https://ieeexplore.ieee.org/document/8663379>.
- Cohen, A., and Landsberg, D., Analysis and automatic classification of breath sounds. *IEEE Trans. Biomed. Eng.* 9:585–590. <https://doi.org/10.1109/TBME.1984.325301>, 1984. <https://ieeexplore.ieee.org/document/4121905>.
- Demir, F., Ismael, A. M., and Sengur, A., Classification of lung sounds with cnn model using parallel pooling structure. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.3000111>. <https://ieeexplore.ieee.org/document/9108294>, 2020.
- Dokur, Z., Respiratory sound classification by using an incremental supervised neural network. *Pattern. Anal. Applic.* 12(4):309. <https://doi.org/10.1007/s10044-008-0125-y>, 2009.
- Emmanouilidou, D., McCollum, E. D., Park, D. E., and Elhilali, M., Computerized lung sound screening for pediatric auscultation in noisy field environments. *IEEE Trans. Biomed. Eng.* 65(7):1564–1574. <https://doi.org/10.1109/TBME.2017.2717280>, 2017. <https://ieeexplore.ieee.org/document/7953509>.
- Kok, X. H., Imtiaz, S. A., and Rodriguez-Villegas, E., A novel method for automatic identification of respiratory disease from acoustic recordings. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2589–2592: IEEE, 2019. <https://ieeexplore.ieee.org/document/8857154>. <https://doi.org/10.1109/EMBC.2019.8857154>.
- Ma, Y., Xu, X., and Li, Y., Lungm+ nl: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation. *Proc. Interspeech* 2020:2902–2906. <https://doi.org/10.21437/Interspeech.2020-2487>, 2020.
- Melbye, H., Garcia-Marcos, L., Brand, P., Everard, M., Priftis, K., and Pasterkamp, H., Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings, Vol. 3. <https://doi.org/10.1136/bmjresp-2016-000136>. <https://bmjopenrespres.bmj.com/content/3/1/e000136>, 2016.
- Mukherjee, H., Dhar, A., Obaidullah, S. M., Santosh, K., Phadikar, S., and Roy, K., Linear predictive coefficients-based feature to identify top-seven spoken languages. *Int. J. Pattern Recognit. Artif. Intell.* 34(06):2058,006. <https://doi.org/10.1142/S0218001420580069>, 2020.
- Mukherjee, H., Obaidullah, S. M., Phadikar, S., and Roy, K., Misna-a musical instrument segregation system from noisy audio with lpcc-s features and extreme learning. *Multimed. Tools Appl.* 77(21):27,997–28,022. <https://doi.org/10.1007/s11042-018-5993-6>, 2018.
- Pal, S. K., and Mitra, S., Multilayer perceptron, fuzzy sets classification. <https://doi.org/10.1109/72.159058>. <https://ieeexplore.ieee.org/document/159058>, 1992.
- Pramono, R. X. A., Imtiaz, S. A., and Rodriguez-Villegas, E., Evaluation of features for classification of wheezes and normal respiratory sounds. *PloS One* 14(3):e0213,659. <https://doi.org/10.1371/journal.pone.0213659>, 2019.
- Rao, A., Huynh, E., Royston, T. J., Kornblith, A., and Roy, S., Acoustic methods for pulmonary diagnosis. *IEEE reviews in biomedical engineering* 12:221–239. <https://doi.org/10.1109/RBME.2018.2874353>, 2018. <https://ieeexplore.ieee.org/document/8514011>.

20. Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., Jakovljevic, N., Turukalo, T. L., Vogiatzis, I. M., Perantoni, E. et al., An open access database for the evaluation of respiratory sound classification algorithms. *Physiol. Meas.* 40(3):035,001. <https://doi.org/10.1088/1361-6579/ab03ea>, 2019.
21. Santosh, K.C., Ai-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multitudinal/multimodal data. *J. Medical Syst.* 44(5):93. <https://doi.org/10.1007/s10916-020-01562-1>, 2020.
22. Santosh, K.C., COVID-19 prediction models and unexploited data. *J. Medical Syst.* 44(9):170. <https://doi.org/10.1007/s10916-020-01645-z>, 2020.
23. Sen, I., Saraclar, M., and Kahya, Y. P., A comparison of svm and gmm-based classifier configurations for diagnostic classification of pulmonary sounds. *IEEE Trans. Biomed. Eng.* 62(7):1768–1776. <https://doi.org/10.1109/TBME.2015.2403616>, 2015. <https://ieeexplore.ieee.org/document/7041226>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Himadri Mukherjee¹ · Priyanka Sreerama² · Ankita Dhar¹ · Sk. Md. Obaidullah³ · Kaushik Roy¹ · Mufti Mahmud⁴ · K.C. Santosh² 

Himadri Mukherjee
himadrim027@gmail.com

Priyanka Sreerama
priyanka.sreerama@coyotes.usd.edu

Ankita Dhar
ankita.ankie@gmail.com

Sk. Md. Obaidullah
sk.obaidullah@gmail.com

Kaushik Roy
kaushik.mrg@gmail.com

Mufti Mahmud
mufti.mahmud@ntu.ac.uk

- ¹ Department of Computer Science, West Bengal State University, Kolkata, India
- ² Department of Computer Science, University of South Dakota, Vermillion, SD, 57069, USA
- ³ Department of Computer Science and Engineering, Aliah University, Kolkata, India
- ⁴ School of Science and Technology, Nottingham Trent University, Nottingham, UK