2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Classification of Respiratory Conditions using Auscultation Sound

Quan T. Do, Kirill Lipatov, Hsin-Yi Wang, Brian W. Pickering, and Vitaly Herasevich, *Mayo Clinic, Rochester, MN, USA.*

Abstract— **Management of respiratory conditions relies on timely diagnosis and institution of appropriate management. Computerized analysis and classification of breath sounds has a potential to enhance reliability and accuracy of diagnostic modality while making it suitable for remote monitoring, personalized uses, and self-management uses. In this paper, we describe and compare sound recognition models aimed at automatic diagnostic differentiation of healthy persons vs patients with COPD vs patients with pneumonia using deep learning approaches such as Multi-layer Perceptron Classifier (MLPClassifier) and Convolutional Neural Networks (CNN).**

Clinical Relevance— **Healthcare providers and researchers interested in the field of medical sound analysis, specifically automatic detection/classification of auscultation sound and early diagnosis of respiratory conditions may benefit from this paper.**

I. Introduction

Respiratory conditions such as Chronic Obstructive Pulmonary Disease (COPD) and pneumonia are among the leading causes of hospitalizations [1] and the most common causes of morbidity and mortality in the world [2,3]. Management of respiratory conditions relies on timely diagnosis and institution of appropriate management. Furthermore, COPD, asthma, pulmonary hypertension, and occupation lung diseases are among the most common respiratory conditions treated in ambulatory and hospital settings. Diagnostic evaluation ranges from bedside or clinical assessment involving history and physical exam to non-invasive radiographic and invasive testing. Despite years of research and characterization of lung pathology, many respiratory illnesses are diagnosed late and may no longer be responsive to therapy [4].

Thoracic auscultation remains the cornerstone of cardiopulmonary physical examination. This inexpensive, readily available, and noninvasive evaluation can pick up adventitious breath sounds, suggest abnormalities, and give clues to underlying conditions. The ability to rapidly note changes in breath sounds has been utilized in respiratory monitoring in a variety of ambulatory, acute care, and perioperative settings. However, the availability and simplicity of auscultation is associated with several potential drawbacks. First, the appreciation of breath sounds is operator-dependent with variable agreement. Second, qualitative assessment implies inherent subjectivity. Inconsistencies can arise from a multitude of patient and device-related factors. Third, respiratory sound evaluation relies on bedside examination and may be compromised in the settings of remote monitoring, self-management, and telemedicine. For these reasons, the standardization in assessment and digital analysis of breath sounds has a potential to enhance reliability and accuracy of this diagnostic modality while making it suitable for remote monitoring, personalized uses, and self-management uses.

Wheezing and crackles have been the focus of breath sound analysis. Wheezing refers to the sound created by the oscillation of opposing walls of the narrowed airways [5]. It has commonly been attributed to a wide frequency range (100Hz – 2Khz) and associated with obstructive airway diseases such as asthma or COPD [6]. Crackles, on the other hand, are typically a late inspiratory finding attributed to the re-expansion of collapsed alveolar spaces. Distinguishing between fine and coarse crackles can be accomplished by appreciating their respective frequencies of 650Hz and 350Hz and may help in diagnostic evaluation [7]. Over the years investigative efforts have concentrated on ways to automate the recognition of wheezing and crackles from breath sound recordings and use them as surrogates for the identification of cardiorespiratory pathologies.

It is of utmost importance to note that finding wheezing or crackles is nonspecific and could be attributed to a multitude of conditions. While wheezing has generally been attributed to airway disorders, it has also been described in the context of airspace diseases such as pneumonia and congestive heart failure [8]. The designation of airway disorders itself refers to a large spectrum of pathologically and clinically distinct conditions from upper airway obstruction to small airway inflammation. Similarly, crackles could represent chronic irreversible disorders such as interstitial fibrosis or more acute treatable conditions such as lower respiratory infection. Therefore, narrowing the classification to specific diagnoses rather than more sensitive auscultatory findings may augment the diagnostic yield of sound recognition models.

In this paper, we describe and compare breath sound recognition models aimed at automatic diagnostic differentiation of healthy persons vs patients with COPD vs patients with pneumonia using novel deep learning approaches such as MLPClassifier and CNN.

V. Herasevich is with the Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, N, 55902, USA. (phone: 507-255-9814; fax: 507-255-4267; e-mail: vitaly@mayo.edu).

Q. Do is with the Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, 55902, USA. (email: do.quan@mayo.edu).

K. Lipatov is with the Department of Medicine, Mayo Clinic, Rochester, MN, 55902, USA. (e-mail: Lipatov.kirill@mayo.edu).

B. Pickering is with the Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, 55902, USA (e-mail: pickering.brian@mayo.edu).

Hsin Yi Wang is with the Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, 55902, USA. Department of Anesthesiology, Taipei Veterans General Hospital and National Yang Ming Chiao Tung University, Taiwan (e-mail: vicky8101@gmail.com) .

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

II. Related work

Sound is traveling vibration, where a wave moves through air. Basically, a wave has two main properties: amplitude (loudness) and frequency (the number of a wave's vibrations or samples over a time period).

In order to recognize patterns, a machine learning (ML) algorithm first must obtain/extract an informative set of features (strong predictors) regarding the desired properties of the raw data before feeding these features into a ML model for training. For audio feature extraction, the audio signal is split into short-term windows (frames) with a set of short-term audio features for each windowed frame. The common extracted features are Mel Frequency Cepstral Coefficients (MFCCs) which signify the short-term power spectrum of a frame, Wavelet (waveform of limited duration), and a Short-Time Fourier Transform (STFT) which is a sequence of Fourier transforms of a frame. Recent efforts on computerized breath sound recognition used features such as MFCCs [9, 10, 11], wavelets [12, 13], and STFT [12, 13, 14], and optimized S-transform [7]. The MFCC process includes STFT together with a Mel-frequency scaled filter bank and a Discrete Cosine Transform [15]. MFCC has been acknowledged as the most popular audio feature extraction method [15]. The extracted features are then fed into a traditional ML model (HMM, GMM, BDT, SVM) or a deep learning model such as MLP, CNN, Recurrent Neural Networks (RNN) and Residual Neural Networks (ResNet) to train a sound recognition model.

For this study, we obtained the publicly available Respiratory Sound database [17] that was provided for the scientific challenge organized at Int. Conf. on Biomedical Health Informatics (ICBHI'17) to train models for respiratory conditions classification. The summary of previous work on sound recognition using this dataset is described below:

TABLE I. Summary Performance of Previous Work

Author	Classification	Feature	Mode	Sen %	Spe %	Acc %
Jakovljevic[16]	Anomaly-driven[1]	MFCC	HMM,GMM	N/A	N/A	39.6
Chen [7]	Anomaly-driven[1] (3)	OST	ResNet	N/A	N/A	98.8
Chambers [9]	Anomaly-driven[1] (4)	MFCC	BDT	22	78	49.6
Serbes [12]	Anomaly-driven[1] (4)	STFT+ wavelet	SVM	55	83	57.9
Perna [11]	Pathology-driven[2]	MFCC	CNN	N/A	N/A	82.0
Ma [13]	Anomaly-driven[1] (4)	STFT+ wavelet	Bi-ResNet	31	69	52.8
Kochetov [10]	Anomaly-driven[1] (3)	MFCC	NMRNN	58	73	65.7
Demir [14]	Anomaly-driven[1] (4)	STFT	CNN	N/A	N/A	65.5
Acharya [6]	Anomaly-driven[1] (4)	MFCC	CNN-RNN	49	84	66.3

HMM - Hidden Markov models; OST - Optimized S-transform; BDT - Boosted decision tree; GMM - Gaussian mixture model; NMRNN - noisemaking RNN; MFCC+ - MFCC and low-level features; STFT - short-time Fourier transform.

1-Anomaly-driven classification of breath sounds differentiates between wheezes, crackles, and the combination of or absence of adventitious sounds, with the number in parenthesis representing the number of classifiers used.

2-Pathology-driven classification refers to the identification of healthy vs unhealthy individuals based on breath sounds and may further include subclassification of chronic vs acute findings.

Prior work on sound classification using the ICBHI database revolved around the detection and differentiation of adventitious lung sounds. Data was classified into wheezes, crackles, and the combination or absence of both anomalies in most studies (Table 1). MFCC or STFT features (alone or in combination with wavelets) were employed in all but one study. A wide variety of deep learning and traditional ML models have been created for sound recognition. Several subsequent studies employed CNN models resulting in accuracy ranging from 65.5% to 82% [14, 11]. Notably, one of the author's later works using RNN achieved even better performance with accuracy over 90% [18]. A different variant of RNN that included both noise and respiratory classifiers was successfully developed in an anomaly-driven model reaching accuracy of 66% [10]. Finally, a hybrid CNN-RNN model has been invented to tackle the temporal and frequency variance commonly seen in adventitious lung sounds [8]. In this design CNN obtained abstract features while temporal relationships were established by the Long Short-Term Memory (LTSM) layer. In addition, softmax classifier was implemented to output results. The hybrid model achieved 66.3% accuracy. Besides, ResNet models have also been successfully developed to differentiate spectrogram data of respiratory anomalies [7, 13]. One study used STFT and the wavelet feature in combination with two ResNet blocks to recognize adventitious sounds from ICBHI database with the accuracy of 50.16% [13]. In a different study an impressive 98.8% accuracy was accomplished when optimized S-transform feature was classified with ResNet [7].

Pathology-driven work focused on distinguishing healthy vs unhealthy respiratory sounds. For this binary classification, a CNN model together with MFCC feature reached 83% accuracy [11]. A categorical classification was also proposed distinguishing from chronic vs acute respiratory pathologies, resulting in similar performance. Further work of these authors incorporating RNN led to superior results [18]. While acuity of pathologic findings has clinical implications, diagnostic classification of respiratory sounds could further clinical usefulness and dramatically change management. No prior study to our knowledge evaluated diagnostic performance of sound classification.

III. Method

The ICBHI'17 Respiratory Sound Database [17] contains audio samples of recordings obtained independently by the researchers at the University of Aveiro, the Aristotle University of Thessaloniki, and the University of Coimbra. The samples were recorded at Hospital Infante D. Pedro (Aveiro, Portugal), the Papanikolaou General Hospital, (Thessaloniki, Greece) and the General Hospital of Imathia (Naousa, Greece) in real life conditions with high noise levels. The cycles were confirmed by respiratory experts as "including crackles, wheezes, a combination of them, or no adventitious respiratory sounds" [17]. A total of 920 annotated audio samples from 126 patients were recorded through 5.5 hours to obtain 6898 cycles with duration range from 10s to 90s. Among these cycles, 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes. The diagnosis/illness of the subjects are also provided in a text file with 2 columns: PatientNumber and his/her respiratory condition. None of the subjects had comorbidity conditions. The recorded audio files were named by the combination of 5 elements, separated with underscores. These elements were PatientNumber, RecordingIndex, ChestLocation, AcquisitionMode, and EquipmentType. When checking each filename for the PatientNumber and matching it with the diagnosis file, we can know the respiratory condition of the recorded subject. Among 126 subjects, there are 8 respiratory conditions: 1 asthma sample, 16 Bronchiectasis samples, 13 Bronchiolitis samples, 793 COPD samples, 35 Healthy samples, 2 LRTI samples, 37 Pneumonia samples, and 23 URTI samples.

A. Data Preprocessing

a. Data Selection

In this paper, we report only the results of the classification of Healthy-COPD-Pneumonia conditions using breath sound. The data included 793 COPD samples, 35 Healthy samples, and 37 Pneumonia samples.

b. Data Exploratory and Data Engineering

The main properties of the audio files were accessed to ensure property consistency of data (data exploratory). These attributes included sample rate, number of audio channels, and bit-depth. The "sampling rate" reflects how frequent it will take samples. All files have the sampling rate of 44.1khz, which indicates the samples are taken 44,100 times per second. Bit depth describes how detailed it will take samples. A 16-bit depth indicates that any sample can take a value from range 65,536 values corresponding to its amplitude. Samples taken with 8 bit will be 256 times less detailed than that of 16 bit. All the audio files are monophonic (single audio channel). Since the bit-depth was not consistent, all files were converted (data engineering) to 16 bit audio.

c. Data Visualization

The visualization of waveform and spectrogram of selected samples were conducted. Below are the samples of the 3 groups. X-axis represents time (distance) and Y-axis represents amplitute of the sound:

Figure 1. Waveform visualization of COPD Sample

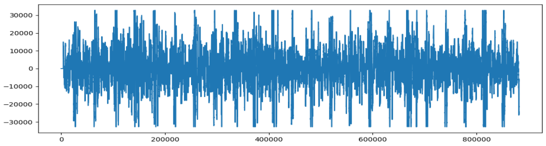


Figure 2. Waveform visualization of Pneumonia Sample

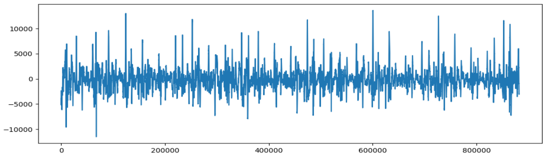
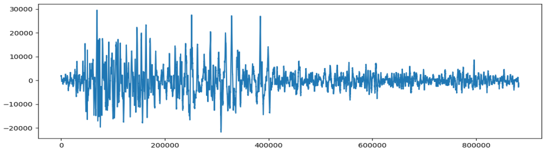


Figure 3. Waveform visualization of Healthy Sample



c. Data Splitting

For all the classification models that are reported in this paper, 80% of the audio samples of each diagnosis were selected randomly for training models, and 20% of the audio samples of each type were used for testing.

B. Features Extraction

There are 4 features that are extracted from the breath sound samples. These features include (1) MFCC short-term power spectrum, (2) mel-scaled spectrogram frequency, (3) chromagram (the 12 different pitch classes), and (4) tonnetz (computed tonal centroid features). A total of 166 attributes were obtained as a result of feature extraction.

C. Multi-layer Perceptron Classifier (MLPClassifier)

The initial classification model was developed with MLPClassifier. MLP is a feedforward Artificial Neural Networks model with at least one input layer, one hidden layer, and one output layer. MLP adopts a backpropagation algorithm to calculate a gradient of the error function with respect to the weights by comparing the desired output with the expected output then adjusting the weight in order to minimize the difference between the actual output and the oncoming output. Each node in the hidden and output layers uses a nonlinear activation function to solve nonlinear pattern classification problems. MLP works best with data that is not linearly separable, therefore, it has been employed for biological analysis, image and speech recognition. In order to classify Healthy-COPD-Pneumonia breath sound, we trained a MLPClassifier model with the activation function = 'relu' by default, the batch_size (the size of minibatches for stochastic optimizers) of 250, the hidden_layer_size (the number of neutrons in a hidden layer) of 300, the maximum number of iterations of 500, and a constant initial learning rate as long as the learning loss (error) keeps on reducing (learning_rate='adaptive'). Variables obtained by the above feature extraction step were used to train the model.

D. Convolution Neural Network (CNN) Model

As MLPClassifier has both strengths and limitations, another CNN model was trained to classify Healthy-COPD-Pneumonia breath sound. The CNN model includes two main components which are the feature extractors and a classifier. Each layer in the feature extractor receives its immediately preceding layer's output as input, then its output is transferred as an input to the succeeding layers. For the classifier (also called the dense layer), the output of the feature extractor is transformed into a 1D feature vector. Since we had conducted the features extraction manually by extracting MFCCs, mel, chroma, and tonnetz of the audio files, we could use them to feed directly into the Dense layer. This way also made sure that the data that was used to train the MLP and CNN models were the same. Finally, we used the Softmax activation function for outputting the results. With Softmax as an activation function, the output would be the prediction probability of each class so the sum of all Softmax units would be 1 in the case of categorical classification model. The architecture of the CNN included 4 Dense layers. The first 3 Dense layers had 166, 256, and 128 units respectively with 'relu' activation function. The output Dense layer had the output of 3 units since we had 3 classes to classify. Two Dropout layers were added to the model to reduce overfitting and, therefore, increase generalizability.

IV. Result

The accuracy rate of MLP model was 94.12% while the accuracy rate for the CNN model was 99.02% for classifying Healthy-COPD-Pneumonia diseases. For the MLP model, the sensitivities for COPD, Healthy, and Pneumonia were 96.7%, 96.7%, 80% respectively while the specificities were 90.3%, 96.8%, 100% respectively. For the CNN model, there were only 2 Pneumonia samples misclassified as COPD. As MLPClassifier is computationally costly when the number of weights is high, it becomes inefficient due to the reduced generalization ability when dealing with data that has a spatial or temporal relationship. Although the strength of MLP is being able to learn non-linear data, the flexibility of MLP also indicates that MLP has high potential of fitting to noise and systemic variation in the data. CNN has been known for eliminating this problem.

TABLE II. Performance Comparision of Models

Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
MLP	91.1	95.7	94.1
CNN	99.3	100	99.2

V. Discussion

Prior attempts at sound classification targeted identification and differentiation of wheezing and crackles. While this information is clinically useful, these findings are nonspecific and definitive diagnostic conclusions can rarely be drawn from them in isolation. Furthermore, narrowing the spectral waveform analysis of breath sounds to only patterns appreciable by the human ear can greatly limit predictive potential. Relieving the expectations of identifying only wheezing and crackles could uncover previously unknown feature associations and help create generalizable prediction of respiratory pathologies [19]. Repeatedly new highly predictive associations have been discovered in addition to and sometimes contradicting the human perception-derived features previously thought to carry unequivocal predictive potential [20].

We are working to apply developed recognition models on the database of auscultation sounds from patients at Mayo Clinic. The collected recordings will be used to further investigate the model performance in classification of multiple respiratory diseases in real-time. We are also open to explore more possibilities to improve the classification accuracy rate by adding the LSTM classifier into the CNN model.

VI. Conclusion

This paper reported methods to classify Healthy-COPD-Pneumonia conditions from auscultation sound using MLPClassifier and CNN methods. The CNN model's prediction results were 99% accurate. Our paper is among the first papers reporting the diagnostic performance of breath sound classification, which confirmed the possibility of using deep learning to recognize respiratory diseases by auscultation sound.

References

[1] Lash, T. L., et al. (2011). Hospitalization rates and survival associated with COPD: a nationwide Danish cohort study. Lung, 189(1), 27–35.

[2] File T. M. (2000). The epidemiology of respiratory tract infections. Seminars in respiratory infections, 15(3), 184–194.

[3] Hurd S. (2000). The impact of COPD on lung health worldwide: epidemiology and incidence. Chest, 117(2 Suppl), 1S–4S.

[4] Mooney, J., et al. (2019). Potential Delays in Diagnosis of Idiopathic Pulmonary Fibrosis in Medicare Beneficiaries. Annals of the American Thoracic Society, 16(3), 393–396.

[5] Loudon, R, and R L Murphy Jr. "Lung sounds." The American review of respiratory disease vol. 130,4 (1984): 663-73.

[6] Acharya, J., & Basu, A. (2020). Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. IEEE transactions on biomedical circuits and systems, 14(3), 535-544.

[7] Chen, H., Yuan, X., Pei, Z., Li, M., & Li, J. (2019). Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. IEEE Access, 7, 32845-32852.

[8] Pratter, M R., et al. "Diagnosis of bronchial asthma by clinical evaluation. An unreliable method." Chest vol. 84,1 (1983): 42-7. doi:10.1378/chest.84.1.42.

[9] Chambres, G.,et all. (2018). Automatic detection of patient with respiratory diseases using lung sound analysis. The 2018 International Conference on Content-Based Multimedia Indexing (CBMI) (pp. 1-6).

[10] Kochetov, K., et al. (2018). Noise masking recurrent neural network for respiratory sound classification. In International Conference on Artificial Neural Networks (pp. 208-217). Springer, Cham...

[11] Perna, D. (2018). Convolutional neural networks learning from respiratory data. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2109-2113). IEEE.

[12] Serbes, G., et al. (2017). An automated lung sound preprocessing, and classification system based on spectral analysis methods. In International Conference on Biomedical and Health Informatics (pp. 45-49). Springer, Singapore.

[13] Ma, Y., et al. (2019). LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm. In 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1-4). IEEE.

[14] Demir, F., Sengur, A., & Bajaj, V. (2020). Convolutional neural networks based efficient approach for classification of lung diseases. Health information science and systems, 8(1), 1-8.

[15] Toledano, D. T., et al. (2018). Multi-resolution speech analysis for automatic speech recognition using deep neural networks: Experiments on TIMIT. PloS one, 13(10), e0205355.

[16] Jakovljević, N., & Lončar-Turukalo, T. (2017). Hidden markov model based respiratory sound classification. In International Conference on Biomedical and Health Informatics (pp. 39-43). Springer, Singapore.

[17] Rocha, B. M., et al. (2019). An open access database for the evaluation of respiratory sound classification algorithms. Physiological measurement, 40(3), 035001.

[18] Perna, D., & Tagarelli, A. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 50-55). IEEE.

[19] Leisman, D. E., et al. (2020). Development and Reporting of Prediction Models: Guidance for Authors. Respiratory, Sleep, and Critical Care Journals. Critical care medicine, 48(5), 623–633.

[20] Galloway, C. D., et al. (2019). Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. JAMA cardiology, 4(5), 428-436.