WEARABLES FOR RESPIRATORY SOUND CLASSIFICATION

R. Shivapathy\textsuperscript{1}, Steny Saji\textsuperscript{2}, Dr. Nishi Shahnaj Haider, Assistant Professor\textsuperscript{3}

Department of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.

Email: shivapathyraguraman@gmail.com\textsuperscript{1}, sajisteny31@gmail.com\textsuperscript{2}, nishishahnaj@karunya.edu\textsuperscript{3}

Abstract: Respiratory disorders being one of the leading causes of deaths in the world, auscultation is one of the most popular methods used in early diagnosis and prevention, but this method faces drawbacks due to human errors. Hence the importance of an automated diagnosis method is being considered. This article investigates the classification of normal and adventitious respiratory sound analysis using the deep CNN RNN model. The classification models and strategies classify breathing sound anomalies such as wheeze and crackle for automated diagnosis of respiratory sounds. The data received through acquisition is denoised with Ensemble Empirical Mode Decomposition, a noise assisted version of the EMD algorithm. The features of respiratory sound are extracted and sent for training in the CNN-RNN model for classification. The proposed classification model scores an accuracy of 0.98, sensitivity of 0.96 and specificity of 1 for the four class prediction.

Keywords- Wheeze. Crackle. Chronic Obstructive Pulmonary Disease. Respiratory Sound. Ensemble Empirical Mode Decomposition. CNN RNN Model.

1. Introduction

According to the World Health Organization (WHO) [1], Respiratory Diseases are among the leading causes of deaths in the world. Respiratory sounds distinguished into healthy and unhealthy sound. Based on this respiratory sound of a healthy individual ranges from 20-2000Hz of frequency. Two most significant disturbances in lung sound are caused due to the presence of wheeze, crackle or both in an individual.

Wheeze is present in the frequency range 100Hz-2KHz and is considered as a high pitch sound that creates obstructions in the airway and is associated with many chronic obstructive pulmonary diseases (COPD). COPD is a non-curable progressive life threatening lung disease that restricts lung airflow and predisposes to exacerbations and serious illness, but early
treatment can relieve symptoms and reduce the risk of death. Statistics says that more than 3 million people die each year from chronic obstructive pulmonary diseases (COPDs), which is approximately 6% of all deaths worldwide.

Crackle is considered as an explosive and discontinuous sound seen during a breathing cycle for a short period. Lung disorders are mostly considered as incurable hence it is very important to control them in the initial stages.

The most effective method to manage chronic pulmonary disorder is prevention and access to medical treatment. Auscultation is considered as the most significant and traditional method for the early diagnosis of respiratory diseases. It is a non-invasive, cheap and easy procedure to assess the state of the patient's lungs by medical practitioners. However, this method comes with two drawbacks. Firstly, even experienced medical practitioners are subject to misinterpretation of the respiratory sound heard. Secondly, the disproportionate number of physicians compared to the overall population that hinders the speed of diagnosis and reduces the chance of early detection of respiratory diseases.

Hence the need to overcome these limitations several algorithms for feature extraction were designed [2] for the automated detection of the adventitious sounds produced by the lungs. Some popular feature extraction techniques used include spectrogram, Mel-Frequency Cepstral Coefficients (MFCC), wavelet coefficients, etc. Several machine learning (ML) algorithms have also been developed to detect breathing sound anomalies such as Dynamic Time Wrap (DTW), Gaussian mixture model (GMM) [6], Hidden Markov Model (HMM) [7], etc. But most of the strategies developed were for binary classification problem (wheeze or crackle) and therefore, not suitable for multi-class classification.

Deep Learning is a subset of machine learning and widely used methodology for many biomedical applications. Introduction to deep learning algorithms can overcome the drawbacks as network abstracts the useful features and data representations through a training model. The hybrid model of CNN-RNN is used to extract both spatial and temporal/sequential required for training the data. Hence this data can be used in many applications in the health industry for automated classification models.

In this paper we propose a classification strategy through a trained hybrid CNN RNN model. The dataset collected is preprocessed and denoised with an Ensemble Empirical Mode Decomposition. EEMD [11] is a noise assisted and non-stationary time series analysis method. Further the features are extracted from the denoised lung sounds and the data is send to the hybrid CNN RNN model for training and classification to the four class prediction of the respiratory sounds.
2. Related Study
Previously much analysis has been done for the analysis of respiratory sound for the same database. Jyotibdha Acharya, Aindham Basu et al. [2] used a deep CNN-RNN model that classifies respiratory sounds based on Mel spectrograms and implementation of a patient specific model tuning strategy. In this model the input data was converted into a 2D image through the Mel spectrogram. Further the features extracted from the image were sent to classifier for prediction class. The patient specific model was designed to improve the performance of the classification.

Jakovljevic et al. [7] used Hidden Markov Model with a Gaussian mixture model to classify respiratory cycles and spectral subtraction to pre-process the noisy parts and MFCC features are used for classification. The noisy parts were suppressed while the non-noisy parts were collected for further classifications.

Dokur et al. [4] used averaged power spectrum components for classification models of multi-layer perceptron (MLP), grow and learn (GAL) network and a novel incremental supervised neural network (ISNN) were examined for the classification of nine different RS classes into bronchial, broncho-vesicular, vesicular lung sounds, crackles, wheezes, stridor, grunting, squawks and friction rub.

A. Mondal et al. [3] used feature extraction technique based on statistical morphology of respiratory sounds. The input data had sounds that were classified into three class: wheeze, crackle and normal. The model then verified the features through ANN model.

L. Liu et al. [5] used algorithms to recognize sounds of children under noisy backgrounds. The features of Lung sounds were obtained in time frequency domain. The lungs sounds were then classified using MFCC and ANN. The model was proposed for monitoring children to reduce risks and health problems.

3. Materials and Method
The database used in this article is from International Conference on Biomedical and Health Informatics (ICBHI’17) scientific challenge respiratory sound database [2]. The database consists of 920 noisy respiratory lung sounds collected from 126 patients of different demographics and annotated by health professionals.

The database contains respiratory cycles out of which

| Respiratory cycles | 6898 |
|--------------------|------|
| Crackle            | 1864 |
| Wheeze             | 886  |
| Both               | 506  |
| Normal             | 3642 |

Table1. Respiratory Cycles in the dataset
The recordings collected are of varying lengths ranging from 10-90sec and different environments. The dataset contains samples recorded with different electronic stethoscopes. The data recorded from seven positions of the chest: the Trachea, Anterior left, Anterior right, Posterior left, Posterior right, Lateral left and Lateral right.

The next phase includes various steps that include Denoising, feature extraction, and development of a deep learning model. The model thus developed is then used for automated classification and diagnosis of chronic respiratory diseases.

4. Proposed Method

_Denoising and Feature extraction:_ The dataset collected are of different sampling frequency hence the data need to be sampled to avoid the loss of any relevant information.

The dataset is comparatively small for a training model hence the data must go through many data augmentation techniques that were applied to increase the size of the data. This method also helps the network learn useful data representations despite of varying conditions.

The dataset used is collected from noisy environments, different recording conditions and different equipment’s denoising of the wave signals is to be considered. Ensemble Empirical Mode Decomposition denoising method is been used for the same. Denoising is done mainly to avoid artifacts in the original lung sound and for better performance.

Denoising demonstrated that EEMD can extract the features more effectively than common measurements. EEMD is a self-adaptive algorithm. EMD can be directly performed for decomposition on a time-domain signal [14].

To evaluate EEMD denoising algorithm, an attenuation signal is established, and numerical verification of noise signal is implemented. The original signal time is 10 s, and the sampling frequency is 200 Hz.

---

Figure 1 Block Diagram of the proposed system
The calculation formula is shown in the equation.

\[ X(t) = \exp(0.2t) \cdot \cos(6\pi t + 0.5 \sin(6\pi t)) + 0.5 \sin(20\pi t) \]

Further details about the denoising method can be found in [15].

In this process, the critical portion of the amplitude of added noise determines the restraining level. This parameter mainly follows the statistical regularity as follow [16]

\[ \varepsilon_c = \frac{\varepsilon}{\sqrt{P}} \]

Where, \( \varepsilon_c \) is the deviation between the original signal and reconstructed signal, \( \varepsilon \) represents the amplitude of added noise, and \( P \) is the ensemble number.

The formula concludes that the signal decomposition is proportional to the amplitude of added noise and inversely proportional to the ensemble number [16].

Various denoising parameters were calculated through algorithms to improve the performance of the classification model. Parameters like PSNR (75.28956dB), MSE (7.4301e+19dB) were found for the input audio signals.

Features were extracted from the denoised lung sound. This is mainly done to increase the accuracy and to improve the system performance in the four class prediction.
Hybrid CNN-RNN model: The extracted lung sound is taken to the CNN RNN hybrid model for further classification.

The model mainly consists of three divisions: the first division is a deep CNN-RNN model which is mainly used to extract the abstract features from the input data from input signal. The second division consists of a bidirectional long short term memory layer (Bi-LSTM) that learns temporal relations of the data and finally in the third division we have a fully connected and softmax layers that convert the output of previous layers to prediction classes.

The first division consists of batch normalization, max pool layers and convolution. The batch normalization layer scales the input data over each batch to stabilize the training. Each convolution layer is followed by Rectified Linear activation functions (ReLU). The max-pool layer selects the maximum values from a pixel neighborhood which reduces the overall network parameters and results in shift-invariance.

LSTM consists of gated recurrent cells that allow or block the data to pass in a sequence or time series. The following is done by learning the perceived importance of data points. The second division consists of a Bidirectional LSTM that is constructed based on two interconnected LSTM layers, one of which operates in the same direction as the data sequence while the other operates in the reverse direction. So, the current output of the Bi-LSTM layer is a function of data in various timelines.

Finally, the third division comprises of fully connected softmax layers where the output of the Bi-LSTM layer is taken and converted to class predictions. Finally, the classification model is trained for the four-class classification problem of respiratory problems [2].

5. Discussions

It is observed that trained models for image classification perform better than a trained speech or audio recognition. A possible explanation for this could be that while image-trained models are trained on a much larger dataset and therefore, has better performance compared to models trained on relatively smaller audio datasets.

Another possibility could be due to presence of artifacts, since the audio signals are acquired in environments that are not in ideal conditions hence a lot of irrelevant information can be found in the input signal. Hence it becomes difficult to train a classification model for audio signals.

In comparison to the existing studies discussed in the paper, Jyotibdha Acharya, Aindham Basu et al.[2] model provided a significant and reliable results by achieving a score of 71.81% for the four class prediction by extracting features using a Mel Spectrogram.
Jakovljevic et al. [7] model extracted features from the lung sound using MFCC. Their proposed model could obtain a performance score of 39.56% by GMM and HMM methods for a 10 fold cross-training data.

A. Mondal et al.[3] used EMD for feature extraction and the features were with WT ,MFCC and SSA by and ANN and SVM classifiers. The proposed system had a performance accuracy score of 94.16%.

L. Liu et al. [5] classified the data using ANN and MFCC acquired classification rate of 83.3% for diagnosis and classification of lung sound for children.

The proposed strategy in this literature was able to comparatively classify the respiratory sounds into four classes and perform well with a score of 98% accuracy and 96% sensitivity.

6. Conclusion and Future Scopes
In this paper, the developed a hybrid CNN-RNN model classifies the ICBHI’17 respiratory audio dataset. The dataset could be verified by running various algorithms to classify them into the four class predictions.

Since the data used was collected from different demographics denoising the artifacts using EEMD improved the performance of the proposed model. In the proposed model various algorithms were applied to estimate the denoising parameters such as PSNR (75.28956dB), MSE (7.4301e+19dB). This proposed model could provide more significant and reliable results in comparison to the original training set. The classification model scores an accuracy of 98%, sensitivity of 96% and specificity of 1.

The proposed system represents a very promising framework also in the field of telemedicine and can support a remote diagnostic system that would allow users to examine their respiratory conditions, identify possible problematic situations that require urgent intervention in real-time. The classification model embedded in a wearable can help in the real-time acquisition of early diagnosis of respiratory disorders and further can be used for many applications.

7. Reference
[1] World Health Organization. Global Surveillance, Prevention and Control of Chronic Respiratory Diseases: A Comprehensive Approach. World Health Organization, 2007.

[2] Jyotibdha Acharya, Arindam Basu“Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model” IEEE publication 18 March 2020 Page(s): 535 – 544
[3] A. Mondal and H. Tang, "Respiratory sounds classification using statistical biomarker," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 2952-2955.

[4] Dokur Z. Respiratory sound classification by using an incremental supervised neural network. Pattern Anal Appl 2009; 12: 309-319.

[5] L. Liu, W. Li and C. Ji, "Breath Sounds Recognition and Classification for Respiration System Diseases," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6

[6] I. Sen, M. Saraclar, and Y. P. Kahya, “A comparison of svm and gmmbased classifier configurations for diagnostic classification of pulmonary sounds,” IEEE Transactions on Biomedical Engineering, vol. 62, no. 7, pp. 1768–1776, 2015.

[7] N. Jakovljević and T. Loncar-Turukalo, “Hidden markov model based respiratory sound classification,” in Precision Medicine Powered by pHealth and Connected Health. Springer, 2018, pp. 39–43.

[8] Lozano M, Fiz JA, Jané R.Automatic Differentiation of Normal and Continuous Adventitious Respiratory Sounds Using Ensemble Empirical Mode Decomposition and Instantaneous Frequency. IEEE J Biomed Health Informa 2016; 20:486-497.

[9] Lozano M, Fiz JA, Jané R. Performance evaluation of the Hilbert–Huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization. Signal Process 2016; 120: 99-116.37.

[10] Shih-Hong Li 1,5, Bor-Shing Lin 2, Chen-Han Tsai 3, Cheng-Ta Yang 4,5 and Bor-Shyh Lin “Design of Wearable Breathing Sound Monitoring System for Real-Time Wheeze Detection” MDPI published on January 2017.

[11] Xiaochuan He, Rafik A. Goubran, Xiaoping P. Liu, Senior Member ‘Ensemble Empirical Mode Decomposition and Adaptive Filtering for ECG Signal Enhancement’. IEEE published 2012, 978-1-4673-0882-3

[12] N. Gavioli, Y. Palti, G. Alroy, and J. B. Grotberg, “Measurement and theory of wheezing breath sound” Journal of Applied Physiology, vol. 57, no. 2, pp. 481–492, 1984.

[13] P. Piirila and A. Sovijärvi, “Crackles: recording, analysis and clinical significance,” European Respiratory Journal, vol. 8, no. 12, pp. 2139–2148, 1995

[14] Hsuan Ren, Yung-Ling Wang, Min-Yu Huang, Yang-Lang Chang *, and Hung-Ming Kao “Ensemble Empirical Mode Decomposition Parameters Optimization for Spectral Distance
Measurement in Hyperspectral Remote Sensing Data” ISSN 2072-4292 Remote Sens. 2014, 6, 2069-2083, Published: 7 March 2014

[15] Bo Li, Lixin Zhang, Qiling Zhang, Shengmei Yang, "An EEMD-Based Denoising Method for Seismic Signal of High Arch Dam Combining Wavelet with Singular Spectrum Analysis", Shock and Vibration, vol. 2019, Article ID 4937595, 9 pages, 2019.

[16] Ma Z, Wen G, Jiang C. EEMD independent extraction for mixing features of rotating machinery reconstructed in phase space. Sensors (Basel). 2015;15(4):8550-8569. Published 2015 Apr 13.