Deep Learning Based Portable Respiratory Sound Classification System

Adithya Sunil Edakkadan and Abhishek Srivastava
Centre for VLSI and Embedded System Technologies (CVEST), IIIT Hyderabad, India
Email: adithya.edakkadan@research.iiit.ac.in and abhishek.srivastava@iiit.ac.in

2023 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS) | 979-8-3503-8119-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/APCCAS60141.2023.00039

***Abstract*—Respiratory diseases contribute to a majority of deaths worldwide every year. Diseases such as asthma, bronchitis and pneumonia also adversely impact a person's social and economic conditions. They can seriously threaten their health if left undiagnosed and untreated. Techniques such as auscultation are used in the diagnosis of most respiratory diseases. However, using such techniques requires an experienced physician and the diagnosis is subjective. To overcome these challenges, in this work, a portable handheld system has been proposed and a proof of concept implemented to detect and classify respiratory diseases automatically through the use of convolutional neural networks (CNN) running on mobile platforms. Mel spectrograms are generated from the audio signals and fed to the CNN for classification. This work makes use of the publicly available HF_Lung dataset. The classification accuracy achieved by the current implementation using a Raspberry Pi for processing is 80.55% with a sensitivity of 95.65% and specificity of 98.80% on the HF_Lung dataset.**

***Index Terms*—CNN, Deep learning, Respiratory sounds, Mel, Spectrogram**

I. Introduction

Respiratory diseases are among the leading causes of death with lung infections, lung cancer and chronic obstructive pulmonary disease (COPD) accounting for nearly one-sixth of the deaths worldwide. According to the Global Burden of Disease (GBD) study, infections in the lower respiratory tract are among the leading causes of death [1]. A delay in the diagnosis of such diseases can lead to serious irreversible health complications. These diseases also have an adverse impact on the social and economic conditions of a person. Access to diagnostic modalities such as lung ultrasounds is required in order to perform accurate diagnoses. However, under resource-limited situations where physicians do not have access to such infrastructure and equipment, auscultation which involves the analysis of respiratory sounds by a trained physician still serves as a prominent tool for preliminary diagnosis [2]. Respiratory sounds are generated by the movement of air within the respiratory system. These sounds vary depending on the state of the respiratory system and the health of the individual. Different respiratory disorders are characterized by a specific set of respiratory sounds which can serve as indicators of the particular disorder. Table I shows some of the common respiratory diseases and their respective respiratory sound indicators [3].

In this work, a portable system has been proposed and implemented to classify respiratory sounds through the use of audio processing and convolutional neural networks (CNN) in order to aid the auscultation process and allow for faster and more accurate diagnoses even in the absence of trained medical professionals and advanced diagnosis equipment.

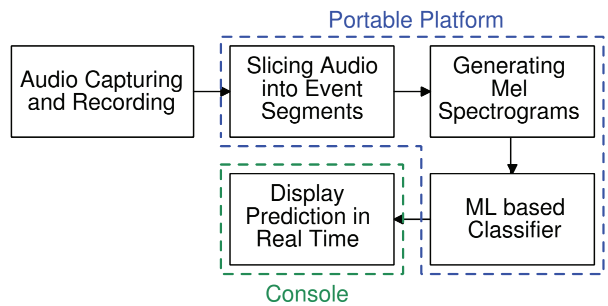

Fig. 1. Overview of proposed system

This work has made use of HF_Lung_V1 dataset which is a combination of a database used in a datathon in Taiwan Smart Emergency and Critical Care (TSECC), 2020, under the license of Creative Commons Attribution 4.0 (CC BY 4.0), provided by the Taiwan Society of Emergency and Critical Care Medicine (TSECCM) and sound recordings acquired from 18 residents of a respiratory care ward (RCW) or a respiratory care center (RCC) in Northern Taiwan [4]. This system is aimed at performing multi-class classification of normal, rhonchi, wheeze, stridor and crackles. This paper is organized as follows: Section II presents the background of this work and a review of related works. Section III elaborates the various signal processing techniques used to process the respiratory audio signals and extract features for classification and presents the deep learning model architecture used for classification. Section IV summarizes the results obtained and

TABLE I
Respiratory diseases and respective lung sounds

Disease	Rhonchi	Wheeze	Stridor	Crackling
Bronchitis	X	X		X
COPD	X	X		X
Pneumonia		X		X
Epiglottitis		X	X	
Laryngomalacia			X	
Cystic fibrosis				X
Heart disease		X		X

2768-3516/23/$31.00 ©2023 IEEE
DOI 10.1109/APCCAS60141.2023.00039

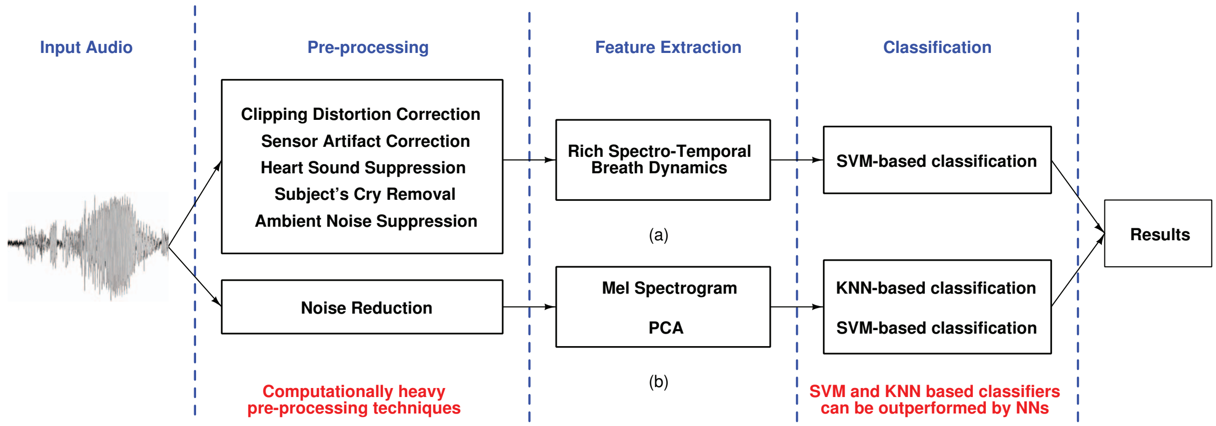

Fig. 2. Respiratory sound classification methods in related works: (a) [5] and (b) [6]

Section V provides the conclusion.

II. Background and Prior Works

Various efforts have been made towards automated lung auscultation for diagnostic applications. [5] proposes a computer-aided approach to classify respiratory sounds even in noisy environments using noise suppression schemes, spectrotemporal features and support vector machines. The framework is shown to be effective in noise suppression suitable for auscultation and explores the use of rich bio-mimetic feature mappings yielding notable improvement in classifying adventitious respiratory events. Fig. 2(a) shows the pipeline with a variety of pre-processing techniques and feature extraction and classification methods. The fallback is that complex and computationally heavy algorithms have been used which necessitate the use of computing systems capable of handling the load and poses a challenge to making the system function in real-time. [6] proposes an ML-based approach to classify respiratory sounds based on Mel Spectrograms and develops a comparison between performance obtained using different algorithms. PCA is applied to the Mel spectrograms to extract audio signature features which are fed to the classifier models. Fig. 2(b) shows the flow of data through the noise reduction stage followed by Mel spectrogram generation, PCA and the classifier. [7] explores the possibility of using smartphones to record respiratory audio for the development of portable solutions. It is concluded that lung auscultation with smartphone built-in microphones is feasible in a clinical context but with heavy limitations on the accuracy of diagnosis. These works have solely focused on the design and implementation of computerized lung auscultation software and are heavily dependent on complex processing algorithms or handicapped due to the lack of sufficiently robust algorithms. There have been no efforts towards making this system completely portable to function outside the clinical environment and in real-time. In this work, we focus on developing a system capable of performing real-time lung auscultation on a portable platform.

III. Respiratory Sound Processing and Classification

A. Respiratory Sound Processing

The HF_Lung dataset contains 9765 15-second audio files with 8,457 wheeze labels, 686 stridor labels, 4,740 rhonchi labels and 15,606 discontinuous adventitious sound labels. Various respiratory events are characterised by sounds of distinctive properties. Ronchi is a low-pitched snoring-like sound with a frequency range of less than 200 Hz while wheeze is a high pitched whistling like sound with a frequency range greater than 400 Hz. Stridor is a high-pitched musical sound with a frequency range greater than 500 Hz while coarse and fine crackles are explosive sounds with frequencies ranging near 350 Hz and 650 Hz respectively [8].

Discerning different respiratory sounds is very difficult from an audio signal due to the presence of noise and lack of visible frequency information. The feeble nature of respiratory audio makes the task of classification more challenging. This brings us to the requirement of more robust representations of the audio signals such as spectrograms in order to aid the task of classification. Spectrograms are an ideal choice for representing respiratory audio signals as they are much more suited for the classification task as compared to audio signals as they contain frequency spectrum information variation with time [9], [10]. With the advances in computer vision oriented machine learning models, the feature extraction and classification of image data such as spectrogram images are more effective and efficient as compared to audio signals.

B. Mel Spectrogram

Spectrograms visually represent the signal strength of a signal at various frequencies over time. Spectrograms can be plotted over the raw magnitude or over a logarithmic scale. However, linear or logarithmic spectrograms may prove to be insufficient to obtain sufficiently differentiable spectrogram representations for respiratory signals as the frequency bands formed in spectrograms will not be discrete enough to discern.

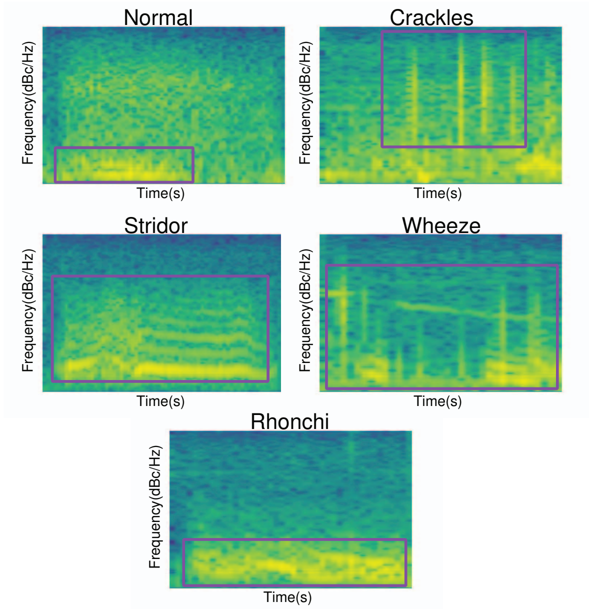

Fig. 3. Mel spectrograms for different respiratory sounds in HF_Lung

The Mel scale allows for better differentiation between lower frequencies as compared to higher frequencies which are ideal for respiratory sounds as they fall in the 50 Hz to 2500 Hz range and is used in such instances for improved detection and classification [8]. Eq. 1 gives the Mel scale value, m, for a frequency of f in Hz.

$$m = 2595 \cdot log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

The spectrogram parameters need to be adjusted in order to ensure that the spectrograms generated are best representative of the data and have sufficient resolution. The size of the FFT, N, defines the frequency resolution of the spectrogram. $\frac{SR}{N}$ gives the frequency resolution, res, of the generated spectrogram with respect to the sampling rate, SR, and N. There is an increase in frequency resolution with the increase in N. Hence, the N value selected must be sufficiently high in order to ensure sufficient frequency resolution in the generated spectrogram. Another parameter to be aware of is the window size since a smaller window size implies not enough information is present in each window to give sensible spectral information while a larger window size implies there will be excess leakage of information leading to a lack of resolution. The type of window is also a determining factor as rectangular windows are very poor choices as they have very minimal side lobe attenuation i.e., their side lobes in the frequency domain have significant magnitude with respect to the main lobe. Hamming and Hanning windows are a far better choice of windows as their side lobes are of negligibly lower magnitude as compared to the main lobe. The magnitude of the first side lobe of the Hamming window is lower than the first side lobe of the Hanning which would make it a better choice for eliminating the contribution of those side lobe frequencies but the distant side lobes of the Hanning are far more attenuated than the Hamming making Hanning a better choice for signals with larger bandwidths i.e., energy spread across a wider frequency range.

Fig. 3 shows the Mel spectrograms generated for various respiratory sounds with 1024 point FFT with Hanning window of length of 128 and hop length of 64. The distinction between different sounds is clearly visible in their respective spectrograms.

C. Deep Learning Classifier

Deep learning is used across various machine learning and artificial intelligence (AI) applications to obtain best possible results by imitating human learning abilities. CNNs are a class of artificial neural networks widely used for image classification applications due to their ability to abstract image data to feature maps [11] - [13]. The Mel spectrograms generated can be processed as images. Mobilenet_V2 is the chosen CNN architecture due to its drastically lower parameter count and memory and processing requirements allowing for use on mobile platforms [14]. In this work, we make use of a transfer learning-based approach. The available MobileNet_V2 model is trained on the ImageNet dataset which implies that the convolution layers are heavily trained for multi-class feature extraction. These pre-trained convolutional layers are used in conjunction with classifier layers trained on the application-specific dataset to allow for state-of-the-art feature extraction and classification.

IV. Results and Discussions

Fig. 4 shows the complete pipeline used in this work. The classifier is trained on the dataset on a multi-GPU system and testing is performed on a Raspberry Pi 4B with 8 GB of RAM in order to verify complete functionality on a portable platform. Raspberry Pi offers sufficient resources for the use of CNN-based classifiers while maintaining a small form factor making it an ideal choice for a portable system. Fig. 5 shows the Raspberry Pi-based portable respiratory sound classification system. Since the Raspberry Pi only has a CPU and no GPUs, the model is reconfigured to run on CPU alone. The system implementation is able to perform the tasks of multiclass classification of normal, rhonchi, wheeze, stridor and crackle sound events from respiratory audio data obtained in real-time. The HF_Lung dataset is split in an 80:20 ratio into training and testing data. Fig. 6 shows the confusion matrix obtained on the test dataset. It is observed that the model is able to classify each of the adventitious sounds with high accuracy. The model achieves an accuracy of 80.55% with a sensitivity of 95.65% and a specificity of 98.80% on the testing dataset. Table II shows the performance summary and compares it with that of related works. It is clearly visible that this work ensures improved sensitivity and specificity while maintaining high accuracy. The accuracy obtained in this work

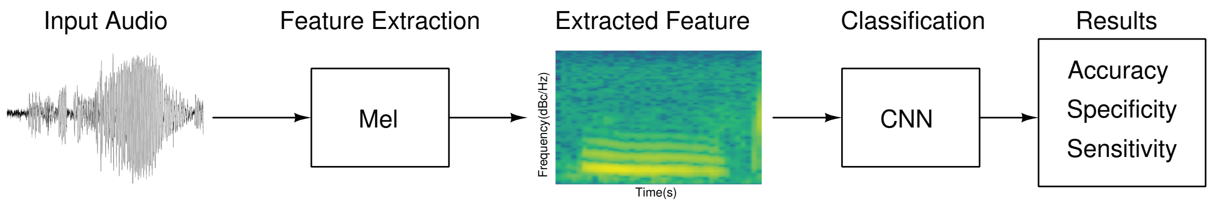

Fig. 4. Respiratory sound classification pipeline

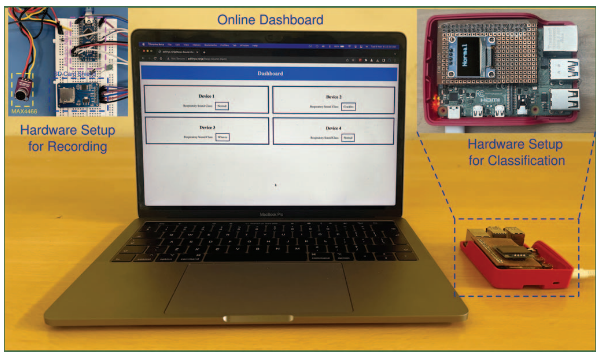

Fig. 5. Portable Respiratory sound classification system with recording system

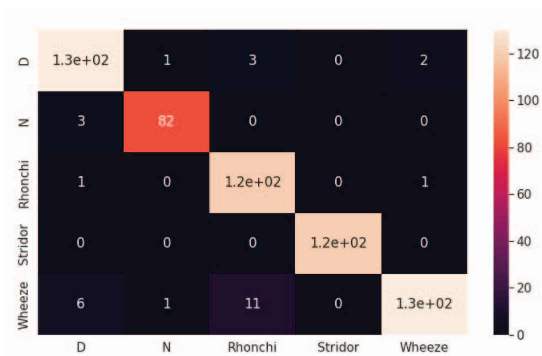

Fig. 6. Confusion matrix obtained for test split of HF_Lung dataset

is also comparable to [5] which makes use of complex and computationally heavy pre-processing algorithms.

Fig. 5 also shows the recording system consisting of Arduino Nano and MAX4466 which can capture audio in real time for use with the classification system. The results of classification from the system are pushed to an IoT cloud service like ThingSpeak through MQTT. This allows for the data to be accessed remotely through a web dashboard which has been set up to centrally monitor all the portable devices. Fig. 7 shows the developed online dashboard.

V. Conclusion

Respiratory diseases pose a major threat to our well-being. Methods of diagnosis of such diseases are often complex and require extensive training and experience to perform. Delay in diagnosis can be fatal and such methods are also seen to be less effective in minors. In this work, an automated system has been proposed and a proof of concept implemented to classify respiratory sounds to help recognise anomalies which may be the sign/symptom of respiratory diseases. The respiratory audio recordings obtained from the recording system are converted into spectrograms and fed to CNNs for classification. The system designed and implemented in this work achieves an accuracy of 80.55% with a sensitivity of 95.65% and a specificity of 98.80% on the HF_Lung dataset on the tasks of multiclass classification of normal, rhonchi, wheeze, stridor and crackle respiratory sound events.

TABLE II
Performance Summary and Comparison with Similar Works

Ref.	[5]	[6]	[8]*	This work
Accuracy (%)	86.67	64.45	64.90	80.55
Sensitivity (%)	86.82	-	51.93	95.65
Specificity (%)	86.55	-	77.88	98.80

* Performance on Task 1-2 in the mentioned work which is most comparable to this work

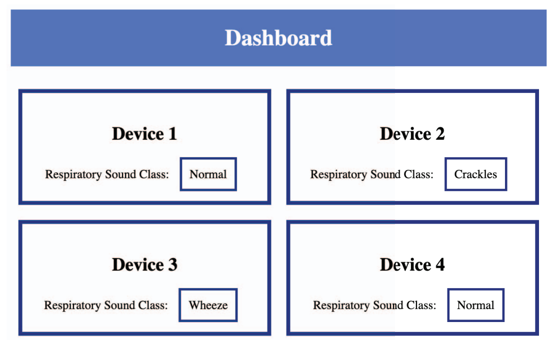

Fig. 7. Online dashboard layout

Acknowledgements

The authors would like to acknowledge Kohli Center on Intelligent Systems (KCIS), IIIT Hyderabad, India for supporting this research.

References

[1] A. Lopez, C. Murray, ”The global burden of disease, 1990–2020,” in *Nat Med 4*, 1241–1243, 1998.

[2] L. Arts, E.H.T. Lim, P.M. van de Ven et al, ”The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: a meta-analysis”, Sci Rep 10, 7347, 2020.

[3] Y. Kim, Y. Hyon, S. S. Jung et al, ”Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning”, Sci Rep 11, 17186, 2021.

[4] F. S. Hsu, S. R. Huang, C.W. Huang, C.J. Huang, Y.R. Cheng, et al., ”Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1”, PLOS ONE 16(7): e0254134, 2021.

[5] D. Emmanouilidou, E. D. McCollum, D. E. Park and M. Elhilali, ”Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments,” in IEEE Transactions on Biomedical Engineering, vol. 65, no. 7, pp. 1564-1574, July 2018.

[6] J. H. L. Serato and R. Reyes, ”Automated lung auscultation identification for mobile health systems using machine learning,” 2018 IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 287-290.

[7] H. Ferreira-Cardoso, C. Jácome, S. Silva, A. Amorim, M. T. Redondo, J. Fontoura-Matias, M. Vicente-Ferreira, P. Vieira-Marques, J. Valente, R. Almeida, J. A. Fonseca, I. Azevedo, Lung Auscultation Using the Smartphone-Feasibility Study in Real-World Clinical Practice, Sensors (Basel), 2021 Jul 20;21(14):4931.

[8] Q. Zhang et al., ”SPRSound: Open-Source SJTU Paediatric Respiratory Sound Database,” in IEEE Transactions on Biomedical Circuits and Systems, 2022.

[9] M. H. Tanveer, H. Zhu, W. Ahmed, A. Thomas, B. M. Imran and M. Salman, ”Mel-spectrogram and Deep CNN Based Representation Learning from Bio-Sonar Implementation on UAVs,” 2021 International Conference on Computer, Control and Robotics (ICCCR), 2021, pp. 220-224.

[10] R. Palaniappan and K. Sundaraj, ”Respiratory sound classification using cepstral features and support vector machine,” 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2013, pp. 132-136.

[11] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano and S. Kido, ”Automatic Classification of Large-Scale Respiratory Sound Dataset Based on Convolutional Neural Network,” *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, 2019, pp. 804-807.

[12] K. Simonyan and A. Zisserman, ”Very Deep Convolutional Networks for Large-Scale Image Recognition”, *2015 International Conference on Learning Representations*, 2015.

[13] K. He, X. Zhang, S. Ren and J. Sun, ”Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, MobileNetV2: Inverted Residuals and Linear Bottlenecks, 4510-4520, 10.1109/CVPR.2018.00474.