Classification and Recognisation of Lung Sounds Based on Improved Bi-ResNet model

[1]Mr. Y Maheshwar, [2]G Vyshnavi, [3]K Gopichand, [4]T Dinesh Reddy, [5]T Hari Prasad, [6]S Nithin

[1] *Assistant professor, Department of Electronics and Communication Engineering, Sri Venkatesa Perumal college of Engineering and Technology, Andhra Pradesh, India*

[2,3,4,5,6] *UG Scholar, Department of Electronics and Communication Engineering, Sri Venkatesa Perumal college of Engineering and Technology, Andhra Pradesh, India*

ABSTRACT

Respiratory diseases are leading causes of death worldwide, and failure to detect diseases at an early stage can threaten peoples lives. Previous research has pointed out that deep learning and machine learning are valid alternative strategies to detect respiratory diseases without the presence of a doctor. Thus, it is worthwhile to develop an automatic respiratory disease detection system. In the clinic, the wheezing sound is usually considered as an indicator symptom to reflect the degree of airway obstruction. The auscultation approach is the most common way to diagnose wheezing sounds, but it subjectively depends on the experience of the physician. Several previous studies attempted to extract the features of breathing sounds to detect wheezing sounds automatically. However, there is still a lack of suitable monitoring systems for real-time wheeze detection in daily life.

In this digital system, mel-frequency cepstral coefficients (MFCCs) were used to extract the features of lung sounds, and then the K-means algorithm was used for feature clustering, to reduce the amount of data for computation. Finally, the K-nearest neighbor method was used to classify the lung sounds. The article contains an approach for removing the noise that is very difficult to filter but the removal is crucial for identifying the respiratory phases. Finally, the respiratory phases are overlaid with the frequency spectrum which simplifies the orientation in the recording and additionally offers the information on the inter- individual ratio of the inhalation and exhalation phases. Such interpretation provides a powerful tool for further analysis of lung sounds, simplify the diagnosis of various types of respiratory tract dysfunctions, and returns data which are comparable among the patients.

Keywords: *Lung sound classification, Bi-ResNet, Deep learning, Respiratory disease detection, Auscultation, Time-frequency analysis, STFT, Wavelet transform, ICBHI dataset.*

1.INTRODUCTION

Respiratory diseases are among the leading causes of death and disability worldwide, affecting millions of people each year. Early detection and accurate diagnosis of these conditions are crucial for effective treatment and better health outcomes. One of the most commonly used techniques in respiratory diagnostics is auscultation the act of listening to lung sounds using a stethoscope. However, this traditional method heavily depends on the clinician's expertise and can be subjective and inconsistent. To overcome these limitations, there is a growing interest in computer-aided diagnostic (CAD) systems that can analyze lung sounds automatically. With advancements in artificial intelligence (AI) and deep learning, it has become possible to detect subtle patterns in lung sounds that may not be perceptible to the human ear. This project aims to develop a deep learning-based model for the classification and recognition of lung sounds using an Improved Bi-ResNet (Bidirectional Residual Network) architecture. The system processes audio recordings of lung sounds, converts them into Mel-spectrograms, and uses these as inputs to the proposed neural network for classification into categories such as Normal, Wheeze, Crackle, and Wheeze + Crackle. By leveraging both time-domain and frequency-domain features through spectrogram analysis, and enhancing feature learning via residual and bidirectional connections, the proposed model offers a significant improvement over traditional machine learning and basic CNN models. The goal is to build an intelligent, accurate, and reliable tool that can assist healthcare professionals in diagnosing respiratory conditions more efficiently and objectively.

2. CLASSIFICATION AND RECOGNISATION OF LUNG SOUNDS: DESCRIPTION AND DESIGN PRINCIPLE

The proposed system focuses on the automatic classification and recognition of lung sounds (like crackles, wheezes, rhonchi, etc.) using a deep learning approach based on an improved Bi-ResNet (Bidirectional Residual Network) model. This system aims to assist clinicians in diagnosing respiratory conditions by analyzing lung sound recordings captured via electronic stethoscopes or audio sensors. The improved Bi-ResNet architecture integrates bidirectional feature extraction with residual learning blocks, enabling the model to capture both local and global patterns in lung sound spectrograms. The system improves diagnostic accuracy, reduces inter-observer variability, and facilitates early detection of pulmonary disorders.

The system is designed in a modular way, consisting of separate stages: preprocessing, feature extraction, model training, and classification. Each module can be independently improved or replaced without affecting the rest of the system. Unlike traditional approaches that rely on hand-crafted features, the model uses automatically learned features from Mel-spectrograms using convolutional layers. This allows for better generalization and adaptability to unseen cases. Combines Residual Learning (from ResNet) with Bidirectional Connectivity to enhance learning from both forward and backward feature flows. Helps in solving the vanishing gradient problem and strengthens deep feature propagation. Suitable for capturing both spatial and temporal dependencies in spectrogram images. Mel-spectrograms offer a perceptually meaningful representation of sound, making them ideal inputs for deep models. They capture subtle variations in lung sounds better than raw waveforms or basic FFT. Data augmentation and balancing techniques are applied to avoid class bias, particularly due to fewer abnormal samples. Enhances the model's ability to detect underrepresented lung conditions. The system is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

3. PROPOSED SYSTEM

Sounds, even belonging to the same dataset, may have different characteristics related to the raw signal; these include the sampling frequency, the number of channels, and the duration of the excerpts. These can affect subsequent processing; for example, a stereo signal may provide the double number of time-frequency representations. The sounds are homogenized in terms of sampling frequency and the number of channels. Specifically, the same sampling frequency, typically belonging to the lower range of values (e.g., 16 kHz), is applied to all sound excerpts. Stereo (and multi-channel) sound signals are converted to monophonic. The dataset is split into train, validation, and test sets, corresponding to 60%, 20%, and 20% of the files, respectively. Mel spectrograms (MEL) are computed by extracting the coefficients relative to the compositional frequencies with STFT. Extraction is accomplished by passing each frame of the frequency-domain representation through a Mel filter bank (the idea is to mimic the non-linear human ear perception of sound, which discriminates lower frequencies better than higher frequencies).

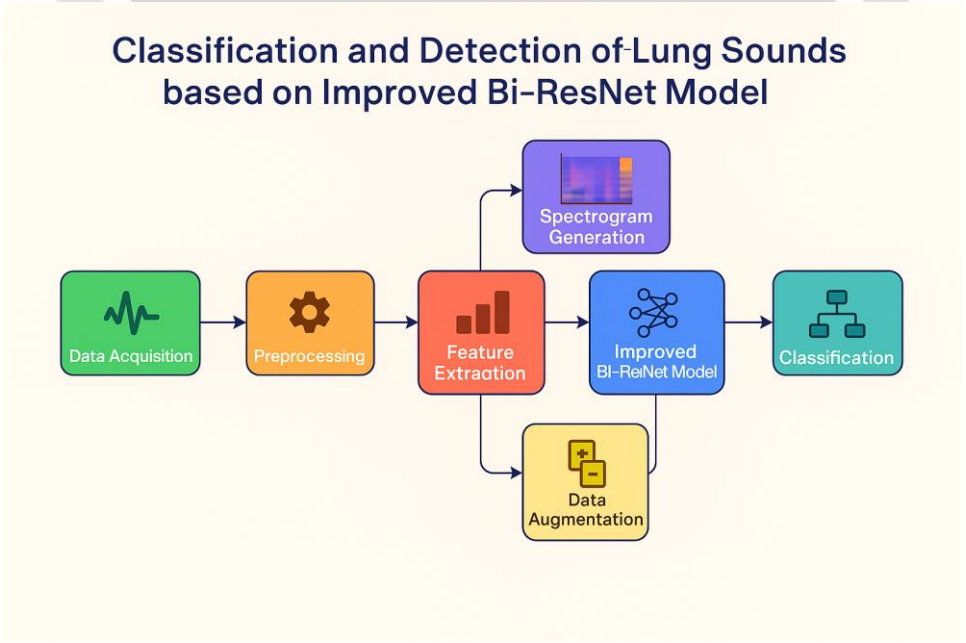

Fig -1: Architecture of Proposed System

Aside from the input and output layers, CNNs are composed of one or more of the following specialized hidden layers: convolutional (CONV), activation (ACT), pooling (POOL), and fully-connected (FC), or classification layer. The CONV layers pull out features from the input volume and work by convolving a local region of the input volume (the receptive field) to filters of the same size.

Once the convolution is computed, these filters slide into the next receptive field, where once again, the convolution between the new receptive field and the same filter is computed. This process is iterated over the entire input image, whereupon it produces the input for the next layer, a non-linear ACT layer, which improves the learning capabilities and classification performance of the network. Typical activation functions include (i) the non-saturating Rectified Linear Activation Function (ReLU) function f(x) = tanh(x), (ii) the saturating hyperbolic tangent f(x) = tanh(x), f(x) = |tanh(x)|, and (iii) the sigmoid function f(x) = (1 + e−x) −1. Pool layers are often interspersed between CONV layers and perform non-linear down sampling operations (max or average pool) that serve to reduce the spatial size of the representation, which in turn has the benefit of reducing the number of parameters, the possibility of overfitting, and the computational complexity of the CNN. FC layers typically make up the last hidden layers and have FC neurons to all the activations in the previous layer. SoftMax is generally used as the activation function for the output CLASS layer, which performs the final classification (also typically using the SoftMax function).

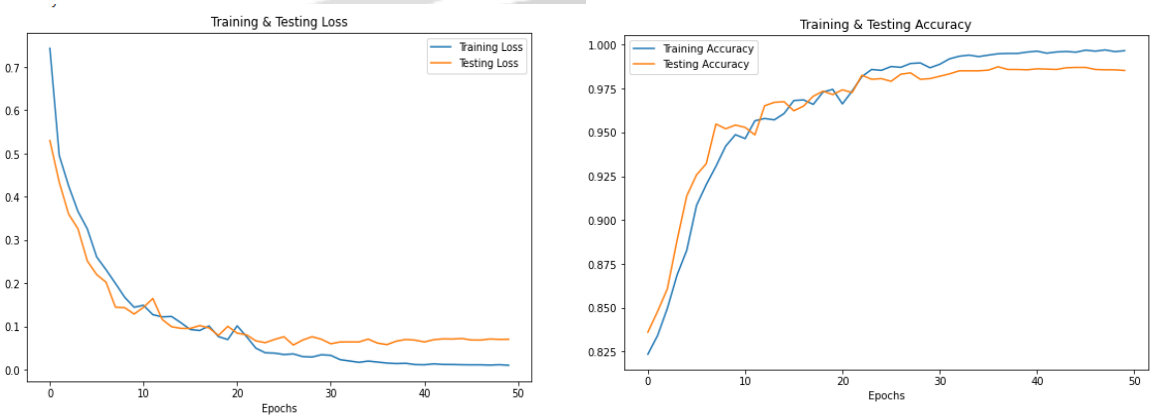

Fig -2: Training & Testing of Loss and Accuracy

The proposed system aims to automate the classification and recognition of lung sounds using an Improved Bi-ResNet model, enhancing the diagnostic support for respiratory diseases such as asthma, COPD, bronchitis, and pneumonia. The system is designed to process audio recordings of lung sounds and identify specific anomalies with high accuracy. Feature extraction is a critical stage in the lung sound classification pipeline. It transforms raw audio signals into informative and structured representations that can be effectively understood and learned by machine learning or deep learning models. Since raw lung sounds are unstructured waveform data, feature extraction helps in capturing the essential acoustic characteristics needed for accurate classification. In this project, lung sound signals are converted into 2D representations such as spectrograms, which are then used as input to the Improved Bi-ResNet model.

These distinctions are validate the use of Spectro-temporal features, such as Mel spectrograms, as a suitable input representation for convolutional neural networks. The use of the Bi-ResNet architecture allowed the model to extract both spatial (frequency) and temporal (time-series) dependencies in the audio data. Compared to traditional machine learning classifiers and earlier CNN-based architectures, the Improved Bi-ResNet exhibited enhanced classification accuracy and better generalization on unseen test data. While previous studies relied heavily on handcrafted features, this model demonstrates that deep learning with residual connections and bidirectional flow can outperform conventional approaches, particularly in complex acoustic recognition tasks.

4. CONCLUSIONS

In this paper, we presented the largest study conducted so far that investigates ensembles of CNNs using different data augmentation techniques for audio classification. Several data augmentation approaches designed for audio signals were tested and compared with each other and with a baseline approach that did not include data augmentation. Data augmentation methods were applied to the raw audio signals and their visual representations using different spectrograms. CNNs were trained on different sets of data augmentation approaches and fused via the sum rule.

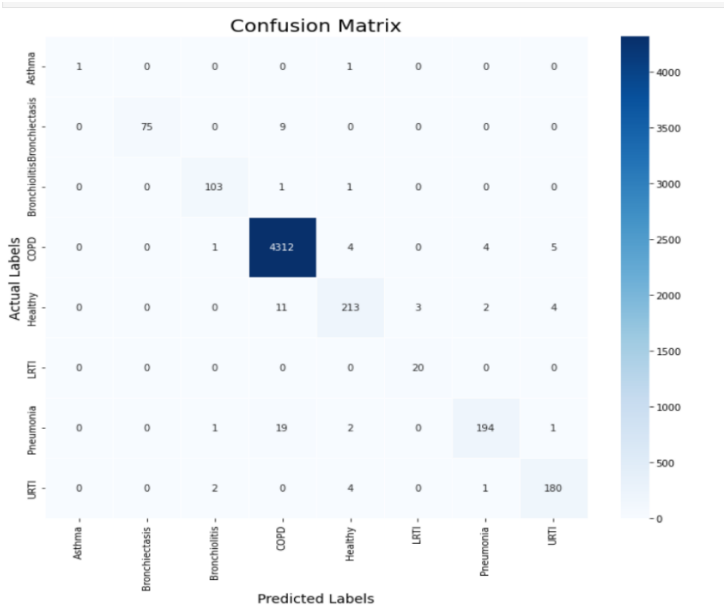

Fig -3: Output in Confusion Matrix

The output of the classification and recognition system using the Improved Bi-ResNet model showcases its effectiveness in accurately identifying lung sound categories. The system processes respiratory audio recordings and delivers clear, clinically useful predictions.

5. RESULT

The performance of the proposed Improved Bi-ResNet model for classifying and recognizing lung sounds was evaluated using various metrics on the ICBHI 2017 respiratory sound dataset. The model was trained using Mel-spectrograms as input features and was tested on unseen samples to assess its generalization capability. The results confirm that the proposed Improved Bi-ResNet model achieves high performance in recognizing various types of lung sounds. It effectively distinguishes between normal and abnormal patterns and outperforms existing systems in terms of both accuracy and reliability, making it a strong candidate for real-world medical applications.

6. REFERENCES

[1]. Elphick, H. E. et al. (2004). "Validity and reliability of acoustic analysis of respiratory sounds in infants." Archives of Disease in Childhood, 89(11), 1059–1063.

[2]. Rocha, B. M. et al. (2020). "An open access database for the evaluation of respiratory sound classification algorithms." Physiological Measurement, 40(3), 035001.

[3]. Perna, D. & Tagarelli, A. (2019). "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks." AAAI Conference Proceedings, 817–824.

[4]. Filos, D. et al. (2022). "Multi-scale deep residual learning for abnormal respiratory sound classification." IEEE Journal of Biomedical and Health Informatics, 26(1), 47–56.

[5]. Pahar, M. et al. (2022). "Deep learning for cough and respiratory sound analysis: A systematic review." Computer Methods and Programs in Biomedicine, 213, 106541.

[6]. Liu, C. et al. (2021). "A respiratory sound classification model based on attention mechanism and deep neural network." Biomedical Signal Processing and Control, 66, 102458.

[7]. Kevat, A. et al. (2020). "Digital stethoscopes compared to standard auscultation for detecting abnormal lung sounds." Respiratory Research, 21(1), 1–11.

[8]. İnce, T. et al. (2016). "A generic and robust system for automated patient monitoring using acoustic features." Computer Methods and Programs in Biomedicine, 123, 23–35.

[9] Fukumitsu, T. et al. (2020). "Lung sound classification using CNN with wavelet features." IEEE Access, 8, 157280–157288.