Classification and Recognition of Lung Sounds Using Improved Bi- ResNet Model

Prof. Sangameshwar Kawdi[1], Avani Naik[2], Gangambika Deshmukh[3], Nitya Bagali[4], Poornima. C[5]
[1,2,3,4,5]*Department of Information Science and Engineering, Guru Nanak Dev Engineering College, Bidar*

***Abstract*—In order to help in the early detection of respiratory conditions like COPD, bronchitis, and asthma, this research offers a sophisticated deep learning framework for lung sound classification. The system combines an enhanced Bi-ResNet architecture that uses skip connections and residual blocks for deeper feature learning with the Short-Time Fourier Transform (STFT) and Wavelet Transform for feature extraction. The model outperforms conventional techniques in terms of accuracy by utilizing data augmentation and feature fusion. With an F1 score of 71.05%, experimental findings on the ICBHI 2017 dataset show a classification accuracy of 77.81%, which is 25.02% better than the baseline Bi ResNet. The potential of AI-driven diagnostic tools to lower subjectivity in medical practice and offer scalable respiratory healthcare solutions is highlighted in this paper.**

***Index Terms*—Lung sound classification, Deep learning, STFT and Wavelet features, Improved Bi-ResNet, Respiratory diagnostics**

I. INTRODUCTION

Millions of people worldwide suffer from respiratory conditions like COPD, and aberrant lung sounds like crackles and wheezes are important diagnostic markers. Conventional diagnosis is mostly dependent on the subjective knowledge of doctors. By automating the classification of lung sounds, deep learning provides a scalable approach. In order to improve feature extraction and classification accuracy, this study presents an enhanced Bi-Res Net model that combines CNN and Res Net modules. Reduced respiratory function brought on by a variety of chronic non-communicable disorders is the hallmark of Chronic Respiratory Disease (CRD), one of the four major chronic diseases in the world. The World Health Organization estimates that 400,000 individuals worldwide pass away from chronic respiratory illnesses each year, with Chronic Obstructive
Pulmonary Disease (COPD) accounting for the majority of these deaths. The respiratory system gradually deteriorates over time with COPD.

In general, new techniques and resources for the early detection and categorization of Chronic Obstructive Pulmonary Disease (COPD) are made possible by machine learning and deep learning technology. These technologies are anticipated to play a significant role in the management of COPD through additional research and development, enhancing the precision of diagnosis and the efficacy of treatment, and lessening the impact of COPD on people's health and the health of society. However, it is crucial to remember that in order to guarantee their safety and effectiveness, machine learning and deep learning technologies still require validation and improvement in clinical practice. In order to effectively forecast and diagnose lung disorders in the future, it is crucial to help medical professionals use deep learning to identify abnormal lung sounds. The model improves the precision of the model and the accuracy of lung sound categorization by successfully resolving the issue of leveraging feature information to extract richer features.

A. Problem Statement

Classifying lung sounds is an important yet difficult task in medical diagnosis. Despite its significance, the creation of dependable automated systems is hampered by a number of issues:

- Noise interference: Overlapping heartbeats, device-specific artifacts, and ambient noises frequently taint respiratory recordings. These distortions make it harder to accurately classify lung sounds and lessen their clarity.
- Limited and unbalanced datasets: There aren't many samples of some aberrant noises (like

wheezing) in publicly accessible datasets like ICBHI 2017. Biased models that perform well on majority classes but badly on minority ones are the result of this imbalance.
- Difficulty differentiating identical abnormal sounds: The frequency ranges and temporal patterns of crackles and wheezes frequently coincide. Even skilled doctors may find it difficult to recognize these small differences, leading to incorrect classification.
- Inadequate automated diagnostic tools: Few systems are incorporated into clinical practice, despite the existence of research prototypes. A lot of models don't work well in a variety of medical settings, recording devices, and patient demographics.
- Subjectivity in medical diagnosis: The clinician's experience and interpretation play a major role in traditional auscultation. Treatment delays and inconsistent diagnoses can result from practitioner variability.
- Variability in recording conditions: The quality of recordings produced by various hospitals and equipment varies. Inconsistencies brought about by this heterogeneity make model evaluation and training more difficult.
- Computational difficulties: Deep learning models need a lot of data and meticulous optimization. Accurate and efficient model creation is challenging due to limited datasets and high computational needs.
- Clinical adoption barriers: Despite models' excellent accuracy in research, they frequently lack real-time processing capabilities, user-friendly interfaces, and interaction with current healthcare operations.

B. Objectives

- Create a deep learning system that accurately classify lung sounds into three categories: wheeze, crackle, and normal.
- Promote early diagnosis of bronchitis, asthma, and COPD by lowering clinical evaluation subjectivity.
- To reduce noise and guarantee consistency, standardize lung sound recordings using resampling, filtering, and normalization.
- Combine the Wavelet Transform for multi-resolution analysis with the Short-Time Fourier Transform (STFT) for time-frequency analysis to improve feature extraction.
- To balance unusual sound samples and enhance model generalization, use non-linear mixed data augmentation.
- To capture multidimensional representations, use an improved Bi-ResNet architecture with residual connections and feature fusion.
- To optimize information consumption, integrate processed and original features concurrently.
- To verify efficacy, assess performance using accuracy, precision, recall, and F1 score.
- Create a system that is clinically relevant and helps medical professionals make decisions.
- Assure future scalability for integration with mobile applications, multimodal medical data platforms, and IoT-enabled stethoscopes.

II. LITERATURE SURVEY

[1] Training deeper neural networks is more challenging. In order to facilitate the training of networks that are significantly deeper than those previously employed, we introduce a residual learning approach. Rather than learning unreferenced functions, we explicitly reformulate the layers as learning residual functions with reference to the layer inputs. We present thorough empirical evidence demonstrating that these residual networks can improve accuracy with far greater depth and are simpler to optimize. We assess residual nets with up to 152 layers on the ImageNet dataset, which is eight times deeper than VGG nets but still less sophisticated. On the ImageNet test set, an ensemble of these residual nets achieves an error of 3.57%. In the ILSVRC 2015 classification task, this outcome took first place. Additionally, we offer a CIFAR-10 analysis with 100 and 1000 layers. For many visual recognition tasks, the depth of representations is crucial. We achieve a 28% relative improvement on the COCO object detection dataset only because of our incredibly deep representations. Our contributions to the ILSVRC & COCO 2015 competitions, where we also took first place in the ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation tasks, are based on deep residual nets.

[2] Most people agree that thousands of annotated training samples are necessary for deep networks to be successfully trained. In order to make better use of the

existing annotated samples, we offer a network and training technique in this research that heavily depends on data augmentation. The architecture comprises of a symmetric expanding path for accurate localization and a contracting path for context capture. We demonstrate that such a network outperforms the previous best approach (a sliding-window convolutional network) on the ISBI challenge for segmentation of neural structures in electron microscopic stacks and can be trained end-to-end using extremely few photos. We won the 2015 ISBI cell tracking competition in these categories by a significant margin using the same network that was trained on transmitted light microscopy images (phase contrast and DIC). Additionally, the network is quick. A 512x512 image's segmentation requires less than a second-rate GPU.

[3] The automatic analysis of respiratory sounds has attracted a lot of attention during the past few decades. Nevertheless, there aren't any sizable publicly accessible databases that may be used to assess and contrast new algorithms at this time. The establishment of such databases is necessary for future advancements in the sector. Method: The public respiratory sound database described in this study was created for the first scientific challenge of the IFMBE's International Conference on Biomedical and Health Informatics, an international competition. Two sets of annotations and 920 recordings obtained from 126 people are included in the database. There are 6898 annotated respiratory cycles in one set; some have crackles, wheezes, or both, while others have no accidental respiratory sounds. The locations of 10,775 crackles and wheezes were annotated in the other set. The best system that took part in the challenge received an average score of 91.2% with the event annotations and an average score of 52.5% with the respiratory cycle annotations. The scientific community will benefit from the construction and public publication of this database, which may draw attention to the respiratory sound classification issue.

[4] In order to improve early disease identification, this study presents a deep learning-based system for lung sound analysis that blends Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architectures. The algorithm is capable of differentiating between a variety of respiratory conditions, such as pneumonia, asthma, and Chronic Obstructive Pulmonary Disease (COPD), using extensive datasets from Coswara and ICBHI. Mel-spectrograms are essential input features in the model's segmentation analysis of lung sound recordings and high pass filtering. In addition to include two Long Short-Term Memory (LSTM) layers in the RNN component, the entire fusion model design combines three CNN layers, three max-pooling layers, and two fully linked layers to produce a feature map that emphasizes the presence of observed features. The cross-entropy loss function and the Adam optimizer are the focus of the training procedure. To address class disparities and improve the generalizability of the model, data augmentation approaches were used. Across a range of respiratory conditions, the experimental results show excellent accuracy, sensitivity, specificity, and F1-score. The model's remarkable accuracy is demonstrated by the performance metrics on the ICBHI dataset: 93.3% for healthy individuals, 93.8% for pneumonia patients, 91.7% for asthma patients, and 94.0% for COPD patients. In terms of precision, recall, F1 score, and accuracy across the ICBHI and Coswara datasets, the model performs better than alternative techniques including decision trees, support vector machines, and random forests.

[5] One of the main causes of death globally, respiratory illnesses have a direct impact on a patient's quality of life. The effective treatment of respiratory disorders depends on early diagnosis and patient monitoring, which typically involve lung auscultation. The method of manually interpreting lung sounds is laborious, subjective, and demands a high level of medical knowledge. Robust lung sound categorization models could be created by utilizing deep learning's capabilities. In order to address training data imbalance, we provide a novel hybrid neural model in this study that uses the focal loss (FL) function. A long short-term memory (LSTM) network receives features that were first extracted from short-time Fourier transform (STFT) spectrograms using a convolutional neural network (CNN). The LSTM network learns the temporal dependencies between the data and classifies four different types of lung sounds: normal, crackles, wheezes, and both crackles and wheezes. The ICBHI 2017 Respiratory Sound Database was used to train and test the model, which produced state-of-the-art results using three different data splitting strategies: sensitivity 47.37%, specificity 82.46%, score 64.92%, and accuracy 73.69% for the official 60/40 split;

sensitivity 52.78%, specificity 84.26%, score 68.52%, and accuracy 76.39% using interpatient 10-fold cross validation; and sensitivity 60.29% and accuracy 74.57% using leave-one-out cross validation.

[6] Deep learning algorithms are transforming medical diagnostics by providing previously unheard-of levels of efficiency and accuracy in the identification and categorization of diseases. The most recent developments in deep learning techniques applied to several diagnostic modalities, such as genetic analysis, medical imaging, and electronic health record data, are examined in this paper. We provide a thorough analysis of the performance of cutting-edge deep learning architectures, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), across a range of clinical contexts. We also look at the benefits and problems of integrating deep learning into diagnostic workflows, such as clinical validation, model interpretability, and data accessibility. Our results show that deep learning has a great deal of promise for improving patient outcomes, early disease identification, and diagnostic accuracy. This study adds to the increasing amount of data that supports the use of deep learning in everyday clinical practice.

[7] Early detection of respiratory illnesses has become more important in the wake of the COVID-19 pandemic. Even though they are accurate, traditional diagnostic techniques like computed tomography (CT) and magnetic resonance imaging (MRI) can have accessibility issues. A less complicated option is lung auscultation, which is heavily dependent on the clinician's skill and is subjective. These difficulties have been made worse by the pandemic, which has limited in-person consultations. By creating an automated respiratory sound classification system utilizing deep learning, this project seeks to get over these restrictions and enable precise and remote diagnosis. Methods: Using spectrographic representations of respiratory sounds in an image classification framework, we created a deep convolutional neural network (CNN) model. To improve classification efficacy, our model incorporates attention feature fusion of low-to-high-level information based on a knowledge propagation method.

III. PROPOSED SYSTEM

Even within the same dataset, sounds may differ in terms of their basic signal parameters, such as sampling frequency, channel count, and excerpt duration. These may have an impact on further processing; for instance, a stereo signal might produce twice as many time-frequency representations. The number of channels and sample frequency of the sounds are homogenized. In particular, every sound clip is subjected to the same sampling frequency, which is usually in the lower range of values (e.g., 16 kHz). Monophonic sound signals are created from stereo (and multi-channel) sound signals. Train, validation, and test sets make up 60%, 20%, and 20% of the files in the dataset, respectively. Mel spectrograms (MEL) are calculated by using STFT to derive the coefficients in relation to the compositional frequencies. Each frame of the frequency-domain representation is sent through a Mel filter bank in order to complete extraction.

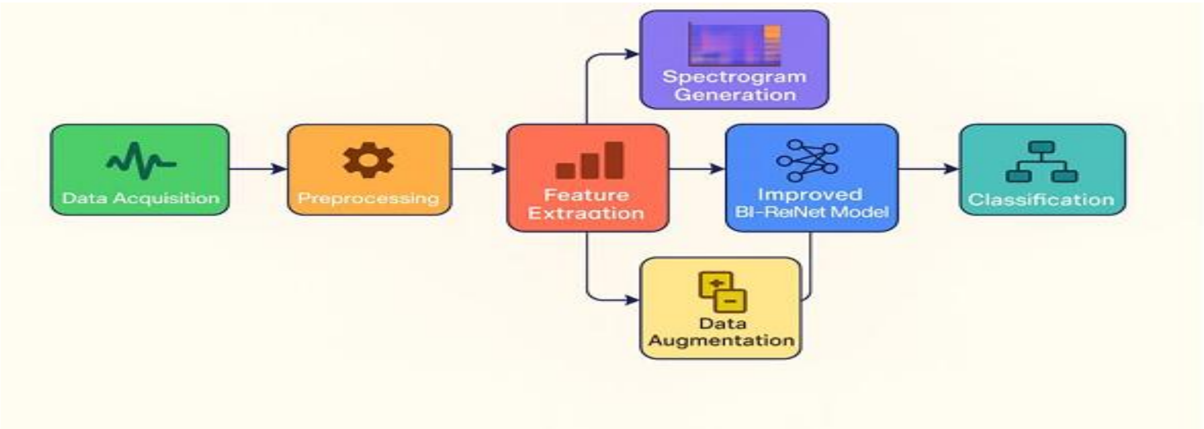


Figure 3.1: Classification and Detection of Lung Sounds based on Improved Bi-ResNet Model

A. Key Features

- STFT Feature Extraction: Captures precise temporal frequency information by converting lung sound waves into spectrograms.
- Wavelet Transform: Offers multi-resolution analysis, making it possible to identify minute changes in unusual sounds.
- Data Augmentation: By increasing aberrant sound samples, nonlinear mixed augmentation enhances the generalization of the model.
- Enhanced Bi-ResNet Model: This model incorporates skip connections between CNN and ResNet modules for deeper learning and fewer vanishing gradient problems.
- Feature Fusion: Richer representation is ensured by processing original and altered features in parallel.
- Multi Class Classification: Recognizes the normal, crackle, wheeze, and crackle+wheeze categories with accuracy.
- Noise Robustness: Preprocessing filters lessen noise from devices and the surroundings.

SYSTEM ARCHITECTURE AND DESIGN

The figure 4.1 illustrates a dual-path deep learning architecture designed for respiratory sound classification, combining STFT (Short-Time Fourier Transform) and Wavelet Transform features.

A. Parallel Feature Pipelines

- STFT Path and Wavelet Path run in parallel.
- Each begins with a 2D convolution layer (64 filters, 3×3 kernel) to extract local patterns.
- Both pipelines then pass through multiple ResNet blocks:
- ResNet(a): A deeper block with multiple convolution layers, ReLU activations, and dropout for regularization.
- ResNet(b): A lighter residual block with fewer layers, still maintaining skip connections.
- After several stacked residual blocks (8 repetitions of ResNet(b)), each path ends with average pooling to reduce dimensionality.

B. Residual Learning (ResNet Blocks)

- ResNet(a): Input passes through a 1×1 convolution for dimensional alignment.
- Parallel branch applies ReLU → Conv → Dropout → ReLU → two Conv layers.
- Outputs are combined via addition, enabling residual learning.
- ResNet(b): Simpler structure: ReLU → Conv → Dropout → ReLU → Conv.
- Output is added back to the input, preserving information and easing optimization.

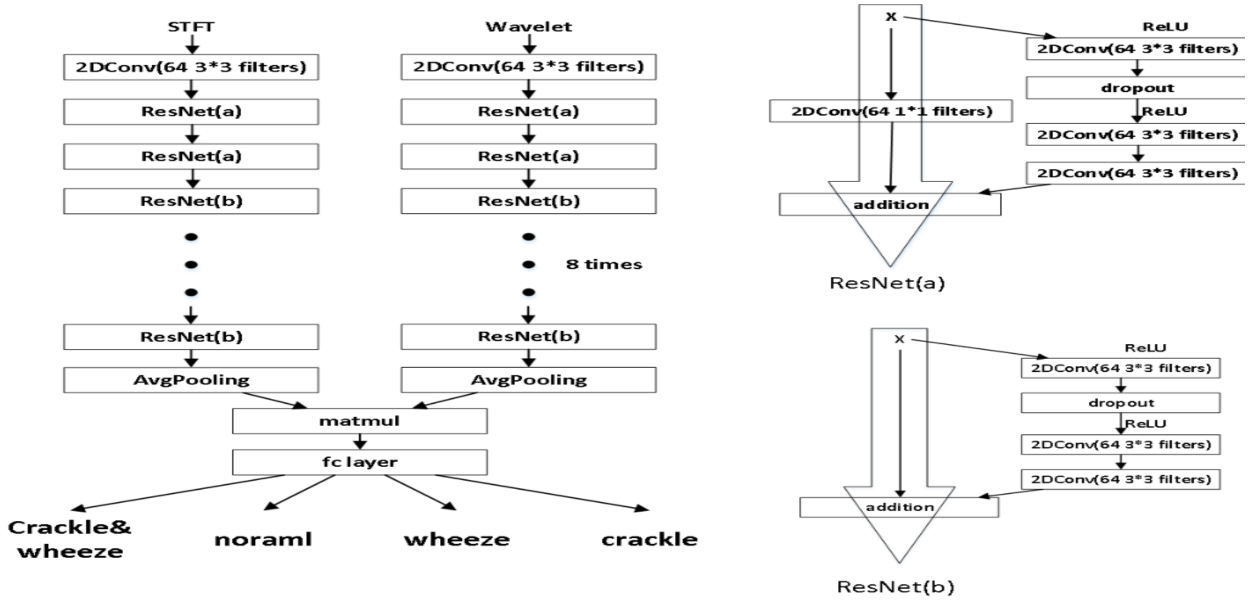


Figure 4.1: Dual-path architecture design

C. Feature Fusion

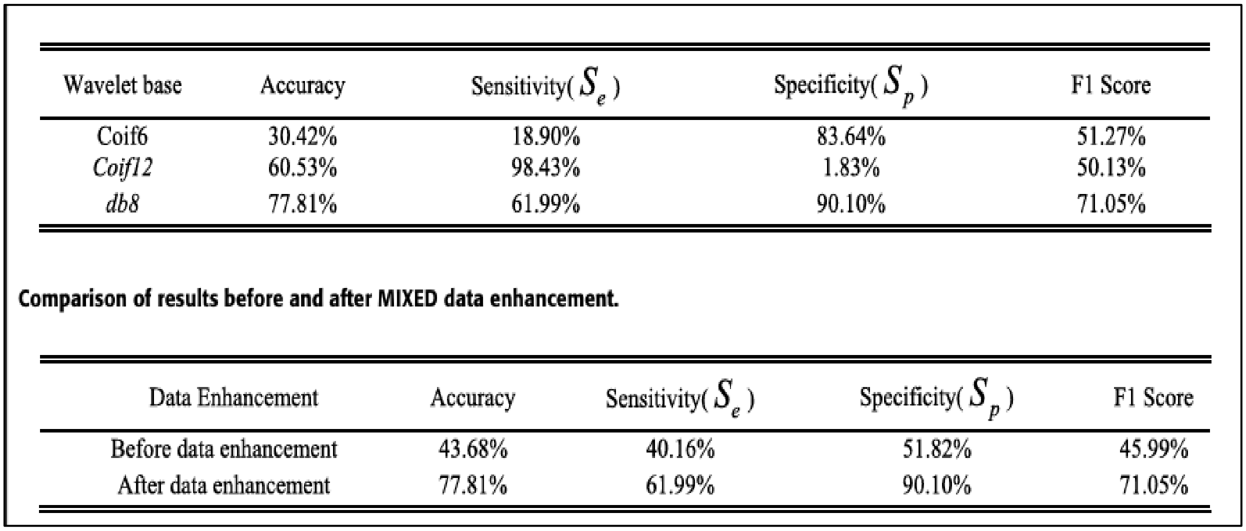
- Outputs from STFT and Wavelet pipelines are combined using matrix multiplication (matmul).
- This fusion integrates time-frequency features (STFT) with multi-resolution features (Wavelet), enriching representation.
- The fused features are passed to a fully connected (fc) layer.

D. Classification Layer

- Final output categories:
- Normal
- Crackle
- Wheeze
- Crackle & Wheeze

This ensures the system can distinguish between healthy and abnormal respiratory sounds.

IV. RESULTS AND ANALYSIS



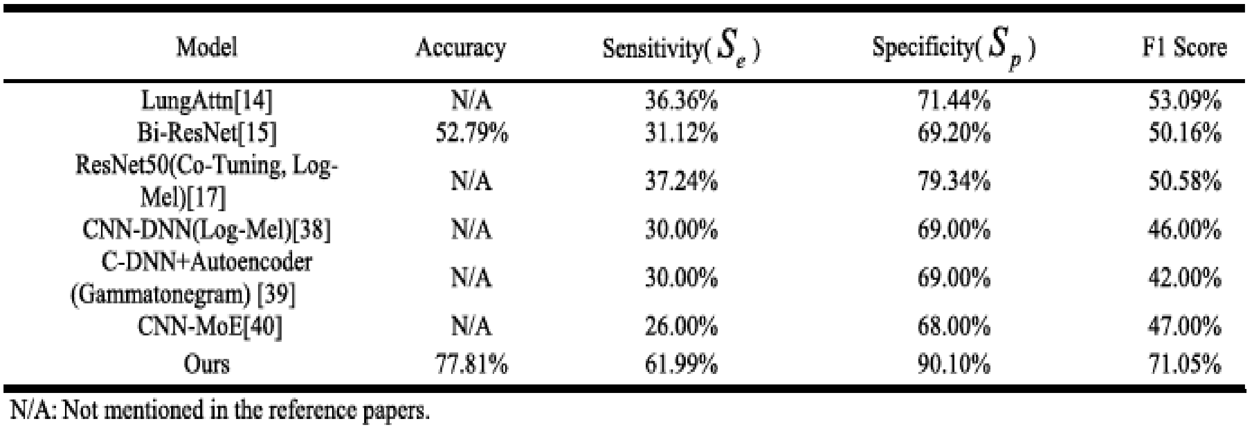
Wavelet base	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Coif6	30.42%	18.90%	83.64%	51.27%
Coif12	60.53%	98.43%	1.83%	50.13%
db8	77.81%	61.99%	90.10%	71.05%

Comparison of results before and after MIXED data enhancement.

Data Enhancement	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Before data enhancement	43.68%	40.16%	51.82%	45.99%
After data enhancement	77.81%	61.99%	90.10%	71.05%

Figure 5.1: Comparison of Wavelet Base Selection Results

- db8 delivers the best overall performance, with the highest accuracy (77.81%) and F1 score (71.05%), indicating a strong balance between sensitivity and specificity.
- Coif12 has extremely high sensitivity (98.43%) but very poor specificity (1.83%), meaning it detects almost all positives but misclassifies many negatives.
- Coif6 has high specificity (83.64%) but very low sensitivity (18.90%), missing most positive cases.

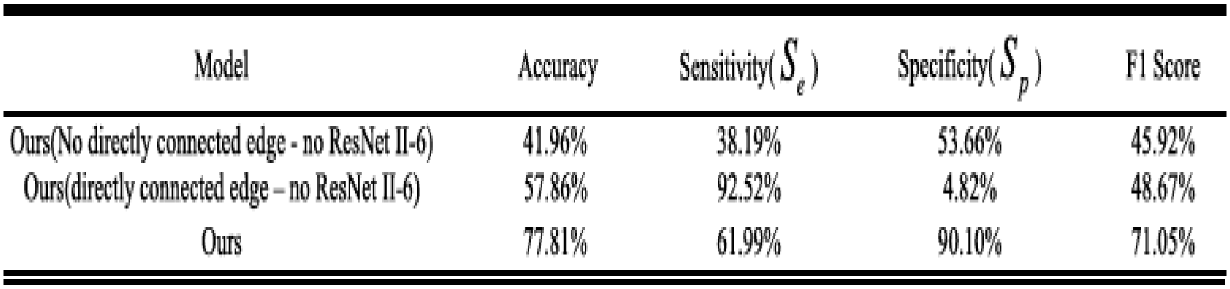


Model	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
LungAttn[14]	N/A	36.36%	71.44%	53.09%
Bi-ResNet[15]	52.79%	31.12%	69.20%	50.16%
ResNet50(Co-Tuning, Log-Mel)[17]	N/A	37.24%	79.34%	50.58%
CNN-DNN(Log-Mel)[38]	N/A	30.00%	69.00%	46.00%
C-DNN+Autoencoder (Gammatonegram) [39]	N/A	30.00%	69.00%	42.00%
CNN-MoE[40]	N/A	26.00%	68.00%	47.00%
Ours	77.81%	61.99%	90.10%	71.05%

N/A: Not mentioned in the reference papers.

Figure 5.2: Results of experiments comparing methods related to lung sounds.

- Accuracy: 77.81% — significantly higher than Bi-ResNet (52.79%).
- Sensitivity: 61.99% — detects abnormal sounds more reliably.
- Specificity: 90.10% — excellent at ruling out false positives.
- F1 Score: 71.05% — strong balance between precision and recall.
- Other models show limited sensitivity (≤ 37.24%) and moderate specificity (≤ 79.34%), indicating weaker detection of abnormal cases and less reliable classification.
- Missing accuracy values for most models suggest either unavailable data or non-reporting in source papers, making your model's full metric set even more valuable.



Model	Accuracy	Sensitivity(S_e)	Specificity(S_p)	F1 Score
Ours(No directly connected edge - no ResNet II-6)	41.96%	38.19%	53.66%	45.92%
Ours(directly connected edge – no ResNet II-6)	57.86%	92.52%	4.82%	48.67%
Ours	77.81%	61.99%	90.10%	71.05%

Figure 5.3: Results of ablation experiments

The addition of ResNet II-6 and the correct edge connectivity significantly improves classification reliability, making your final model robust for clinical use.

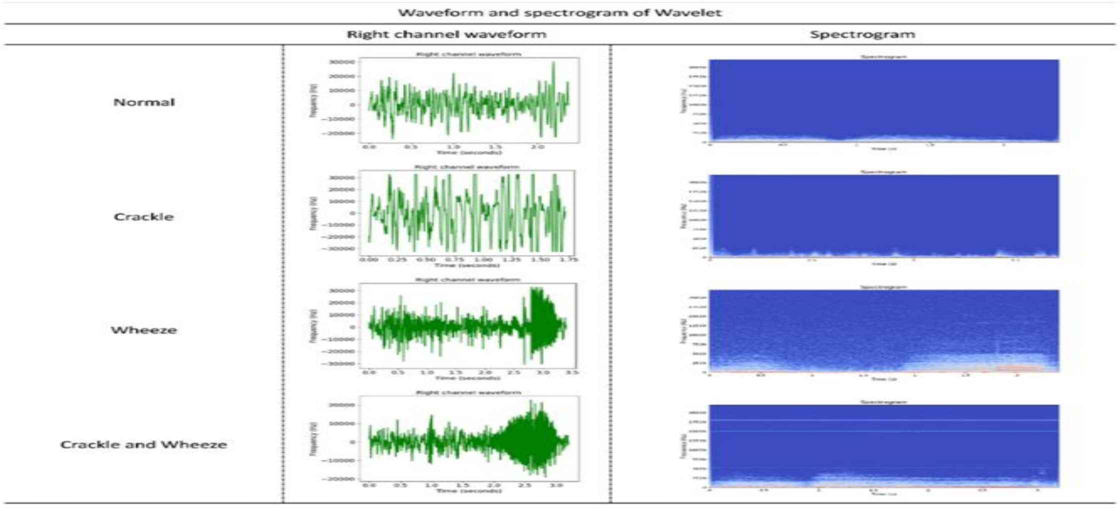


Figure 5.4: Visualization of Wavelet raw lung sound left and right channel waveforms and spectrograms.

V. CONCLUSION AND FUTURE WORK

This project demonstrates how deep learning and signal processing can transform respiratory diagnostics. By combining STFT spectrograms, Wavelet decomposition, and an Improved Bi-ResNet model, it delivers a hybrid, accurate pipeline for classifying abnormal lung sounds.

Key Contributions

- Automated classification into Normal, Crackle, Wheeze, and Crackle+Wheeze
- Visual interpretation of time–frequency and multi-resolution features
- Strong performance (Accuracy: 77.81%, F1: 71.05%)
- Clinician-friendly interface for uploading and visualizing results

The system enhances early detection of respiratory disorders, reduces diagnostic subjectivity, and supports telemedicine workflows. Its hybrid feature extraction and deep residual learning provide superior reliability, laying the foundation for scalable AI-powered respiratory screening tools.

Future Enhancements

- Integration with IoT stethoscopes for real-time monitoring
- Mobile application for clinical deployment
- Expansion to larger datasets for better generalization
- Fusion with multimodal data (e.g., imaging, patient history)

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, Dec. 2015, doi: 10.48550/arXiv.1512.03385.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," arXiv preprint arXiv:1505.04597, May 2015, doi: 10.48550/arXiv.1505.04597.

[3] M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, and E. Perantoni, "An open access database for the evaluation of respiratory sound classification algorithms," Physiological Measurement, vol. 40, no. 3, p. 035001, Mar. 2019, doi: 10.1088/1361-6579/ab03ea.

[4] T. Bikku, S. K. P. N. V., S. Thota, et al., "Deep learning-driven early diagnosis of respiratory diseases using CNN-RNN fusion on lung sound data," Scientific Reports, vol. 15, p. 45233, 2025, doi: 10.1038/s41598-025-28832-7.

[5] G. Petmezas, G.-A. Cheimariotis, L. Stefanopoulos, B. Rocha, R. P. Paiva, A. K. Katsaggelos, and N. Maglaveras, "Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function," Sensors, vol. 22, no. 3, p. 1232, Feb. 2022, doi: 10.3390/s22031232.

[6] O. Panahi, "Deep learning in diagnostics," Journal of Medical Discoveries, vol. 2, no. 1, Apr. 2025, doi: 10.48550/arXiv.390520730.

[7] A. Crisdayanti, S. W. Nam, S. K. Jung, and S.-E. Kim, "Attention feature fusion network via knowledge propagation for automated respiratory sound classification," IEEE Open Journal of Engineering in Medicine and Biology, vol. 5, pp. 383–392, 2024, doi: 10.1109/OJEMB.2024.3402139.