International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# PneumoSense: A Pneumonia detection model using Cough Sound analysis.

**Soham Mishra[1], Shubha Mishra[2], Riya Srivastava[3], Yarun Prakash Soni[4], Suchi Tripathi[5]**

*[1]Bachelor of Technology in Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*
*[2]Assistant Professor, Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*
*[3]Bachelor of Technology in Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*
*[4]Bachelor of Technology in Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*
*[5]Bachelor of Technology in Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** Pneumonia, a respiratory infection with significant global mortality rates, can be effectively screened using non-invasive methods like cough sound analysis. This study presents a deep learning-based classification model called ”PneumoSense”, leveraging a one-dimensional Convolutional Neural Network (1D-CNN), Bidirectional Gated Recurrent Units (Bi-GRU), and an Attention mechanism to detect pneumonia from cough audio signals. PneumoSense is trained on a Coswara dataset comprising 270 healthy and 82 pneumonia cough samples, with data augmentation techniques applied to improve generalization and model robustness. The proposed method achieved a test accuracy of 97.00 %, indicating its potential as a viable screening tool for respiratory illnesses using simple audio input.

***Index Terms***—Pneumonia, Cough Sounds, Deep Learning, Convolutional Neural Networks, Bi-GRU, Audio Classification, Healthcare

## 1. INTRODUCTION

Pneumonia is a life-threatening respiratory infection characterized by inflammation of the alveoli in one or both lungs. Symptoms such as persistent coughing, high fever, chest pain, and difficulty breathing are common, and if left untreated, the condition can quickly escalate, particularly in vulnerable populations. According to the World Health Organization, early and accurate diagnosis is crucial for effective treatment, yet access to standard diagnostic tools such as chest radiography, CT scans, or laboratory tests is often unavailable in rural and underdeveloped regions. PneumoSense, an AI-based diagnostic tool, detects pneumonia using only cough sound recordings. By leveraging the growing capabilities of deep learning and audio signal processing, PneumoSense offers a low-cost, non-invasive, and accessible alternative to traditional diagnostic methods. Research has shown that respiratory conditions—including asthma, COVID19, and pneumonia—alter the acoustic features of coughs in detectable ways. These variations, though subtle to the human ear, can be captured and interpreted by deep learning models trained on large datasets. The core of PneumoSense is a hybrid deep learning architecture comprising a 1D Convolutional Neural Network (1D-CNN), Bidirectional Gated Recurrent Units (Bi-GRU), and an Attention mechanism. This combination allows the model to capture spectral, and temporal characteristic of cough sounds, thereby improving classification accuracy. The system is trained on a public Coswara dataset containing cough samples from both pneumonia patients and healthy individuals. To enhance model robustness and reduce overfitting, we employ extensive audio feature extraction (e.g., MFCC, chroma, spectral contrast, tonnetz) and data augmentation techniques.

*A. Comparative Analysis*

To evaluate the performance of PneumoSense, we conducted a comparative analysis with several traditional machine learning (ML) models commonly used in audio-based classification tasks. These include:

- Support Vector Machine (SVM)
- Random Forest (RF)
- k-Nearest Neighbors (k-NN)
- Logistic Regression (LR)

Each baseline model was trained using the same feature set, derived from the audio preprocessing pipeline. The following table summarizes the performance comparison across key evaluation metrics:

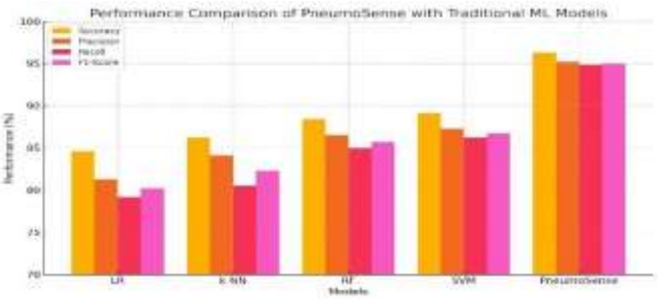


**Fig. 1**. Performance comparison of PneumoSense with traditional ML models in terms of Accuracy, Precision, Recall, and F1-Score.

As shown in Figure 1, PneumoSense consistently outperforms traditional ML models across all evaluation metrics. While classical approaches such as SVM and RF achieve respectable results, they are limited in capturing the sequential and hierarchical patterns present in time-series audio data. In contrast, PneumoSense benefits from the deep feature learning capability of CNN layers, the contextual memory of Bi-GRU, and the interpretability provided by the attention mechanism. In conclusion, PneumoSense demonstrates the potential of deep learning in transforming cough analysis into a reliable diagnostic tool for pneumonia detection. By offering an accessible, non-invasive, and accurate solution, it holds promise for supporting frontline healthcare workers, particularly in remote or resource-constrained environments. This work contributes to the broader field of AI-powered healthcare, aiming to reduce diagnostic disparities and improve health outcomes worldwide.

## II. RELATED WORK

The use of cough sound analysis for diagnosing respiratory diseases, particularly pneumonia, has gained significant attention in recent years, primarily due to its non-invasive nature and potential to be implemented in resource-limited settings. Several studies have demonstrated the effectiveness of audio signal processing combined with machine learning models for classifying various respiratory conditions based on cough sounds. Kapetanidis et al. (2024) highlighted the growing importance of audio analysis and AI in diagnosing respiratory diseases. Their systematic review emphasized the value of cough sound detection, particularly in identifying conditions like asthma and COVID-19. They also pointed out the challenges associated with environmental noise, data quality, and the need for improved diagnostic accuracy through advanced machine learning techniques. This aligns with the motivation behind PneumoSense, which leverages deep learning to address these challenges and provides a reliable diagnostic tool using only cough sounds.[1]

Nguyen and Pernkopf (2022) developed a lung sound classification model using advanced transfer learning techniques, which improved the classification of lung sounds, including adventitious sounds. Their work underlines the importance of handling class imbalances and variations in recording devices—issues that PneumoSense also addresses through data augmentation and feature extraction techniques, such as MFCC, chroma, and spectral contrast.[2]

Ghrabli, Elgendi, and Menon (2024) also explored cough sound analysis for diagnosing respiratory diseases, including pneumonia. They used spectral analysis and machine learning classifiers to identify disease-specific frequencies, which highlights the relevance of audio-based diagnosis in detecting conditions like pneumonia. PneumoSense similarly employs feature extraction to capture the spectral and temporal characteristics of cough sounds, aiding in distinguishing between healthy individuals and pneumonia patients.[3]

Srinivasan and Soni (2021) discussed the potential of ML and DL techniques in diagnosing respiratory diseases through cough sound analysis, noting the challenge of handling diverse datasets and background noise. PneumoSense directly addresses these challenges by employing a hybrid deep learning architecture, combining CNN, Bi-GRU, and an attention mechanism, which allows the model to focus on the most relevant parts of the audio sequence while mitigating issues related to noise and variability in the data.[4]

The work of Amoh and Odame (2016) introduced deep neural networks (DNNs) for cough sound detection and compared CNN and RNN architectures. Their study demonstrated the advantages of DNNs over traditional methods, with CNN achieving higher specificity and RNN excelling in sensitivity. PneumoSense builds on these advancements, using a CNN for feature learning and a Bi-GRU for capturing temporal dependencies, thus improving classification accuracy for pneumonia detection.[9]

Similarly, Sharan and Rahimi-Ardabili (2023) highlighted the performance of deep learning algorithms in analyzing cough sounds for respiratory diseases, including pneumonia. Their findings show that machine learning models can achieve high accuracy in detecting respiratory conditions, though they also emphasized the need for larger datasets and standardized methodologies. PneumoSense benefits from these insights by leveraging a large public dataset and employing extensive data augmentation to improve generalization and model robustness.[10]

In comparison to traditional machine learning models like SVM, k-NN, and random forests, PneumoSense outperforms these models across evaluation metrics such as accuracy, precision, recall, and F1-score. Traditional models often struggle to capture the sequential and hierarchical patterns in time-series audio data, whereas PneumoSense, with its hybrid architecture, benefits from the deep feature learning capabilities of CNN layers, the temporal memory of Bi-GRU, and the interpretability offered by the attention mechanism. In conclusion, PneumoSense contributes to the growing body of research on cough sound-based diagnostic tools, offering a novel, non-invasive solution for pneumonia detection. By combining state-of-the-art deep learning techniques with effective feature extraction methods, it provides a more accurate and robust alternative to traditional diagnostic tools, particularly in settings where access to advanced medical equipment is limited.

## III. METHODOLOGY

*A. Dataset*

The data set comprises cough audio samples classified into two classes: healthy individuals and those affected by pneumonia. The data set initially contained 270 healthy samples and 82 pneumonia samples. The samples were recorded under a variety of conditions, with some background noise present, which provides a more realistic setting to test the robustness of the model. However, it is important to note that the data set is relatively small, which is a limitation that can impact the generalization of the model. Each audio sample was recorded in WAV format and the audio signals were standardized to a sampling rate of 22,050 Hz, to ensure consistency during preprocess and feature extraction.

**Fig. 2**. Healthy Cough Sounds in .WAV Format

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

**Fig. 3**. Pneumonia Cough Sounds in .WAV Format

*B. Data Augmentation*

Data augmentation is an important technique for improving the generalization ability of the model. In this study, the following augmentation techniques are applied to the audio samples: - Pitch Shifting: The pitch of the audio is slightly shifted to simulate variations in cough sounds. - Time Stretching: The speed of the audio is altered without changing the pitch, which helps the model to generalize to different speaking or coughing speeds. - Additive Noise: Noise is introduced into the audio to simulate real-world conditions, such as background chatter or environmental sounds. These techniques help the model become more robust and improve its performance on unseen data.

**Fig. 4.** Augmented Healthy Cough Sounds in .WAV Format

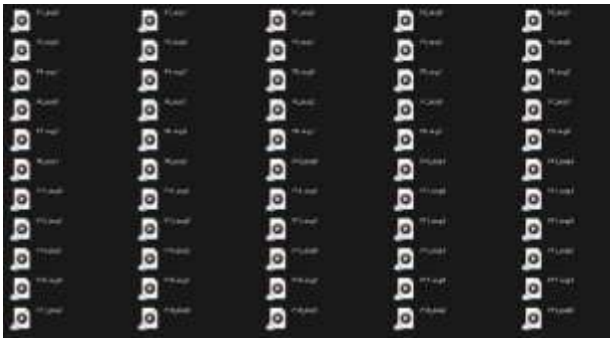

**Fig. 5.** Augmented Pneumonia Cough Sounds in .WAV Format

*C. Data Pre-processing*

To prepare the audio data for deep learning, several preprocessing steps are performed: - Noise Removal: Background noise is filtered out to improve the quality of the audio signal and ensure that the model focuses on the relevant cough sound features. Feature extraction: Several audio features are extracted from each cough sample: - MFCC (MelFrequency Cepstral Coefficients): These coefficients capture the spectral characteristic of the audios signals and are commonly used in speech and audio processing tasks. - Chroma: This feature captures harmonic content, making it useful for tonal analysis. - Spectral Contrast: This feature measures the difference in amplitude between peaks and valleys in the audio spectrum, providing insights into the texture of the sound. - Tonnetz: This feature measures the harmonic relations in the signal, often used in music analysis but also applicable to sound-based tasks. - Zero Crossing Rate (ZCR): This feature counts the number of times the audio signal crosses zero, the amplitudes line, which can be useful in distinguishing different types of sounds. - Normalization: The extracted features are normalized to have zero mean and unit variance, which aids the model in converging faster during training.

*D. Model Architecture*

The proposed pneumonia detection model utilizes a hybrid deep learning structure comprising a 1D Convolutional Neural Network (CNN), a Bidirectional Gated Recurrent Unit (BiGRU), and an Attention mechanism. This combination is designed to extract local features, model sequential dependencies, and highlight critical parts of the cough sound signal.

Input Layer: The input to the model is a time-series feature array extracted from preprocessed cough audio clips. Each audio sample is converted into a fixed-length vector using a combination of 40 Mel-Frequency Cepstral Coefficients (MFCCs), Chroma features, Spectral Contrast, and Tonnetz. These features are stacked and reshaped into a 2D array suitable for the convolutional layer. 1D Convolutional Layer: A 1D CNN layer with 64 filters and a kernel size of 3 is applied to capture short-range acoustic patterns. This helps in identifying characteristic frequency modulations associated with pneumonia-related coughs. Batch Normalization and ReLU Activation: Batch normalization is performed to stabilize the learning process and improve convergence. A ReLU activation function introduces non-linearity, allowing the model to learn complex patterns in the data. MaxPooling1D and Dropout Layer: A MaxPooling1D layer with a pool size of 2 reduces the dimensionality of the feature maps, followed by a Dropout layer with a rate of 0.3 to reduce overfitting by randomly deactivating neurons during training. Bidirectional GRU Layer: The output from the CNN is passed to a Bidirectional GRU layer with 64 units. This recurrent layer captures both past and future context in the time-series data, enabling the model to understand temporal dynamics in cough sounds from both directions. Attention Layer: An Attention mechanism follows the Bi-GRU, which assigns weights to each timestep, allowing the model to focus on the most relevant audio segments that contribute to pneumonia classification. Fully Connected Dense Layer: A Dense layer with 64 units aggregates the weighted outputs from the Attention layer. This layer performs higherlevel reasoning based on the extracted features. Output Layer: The final layer is a Dense layer with 1 neuron and a sigmoid activation function, producing a probability score that indicates the presence or absence of pneumonia. This architecture efficiently combines spatial, temporal, and contextual information, making it highly effective for binary classification of cough audio signals into pneumonia and healthy categories.

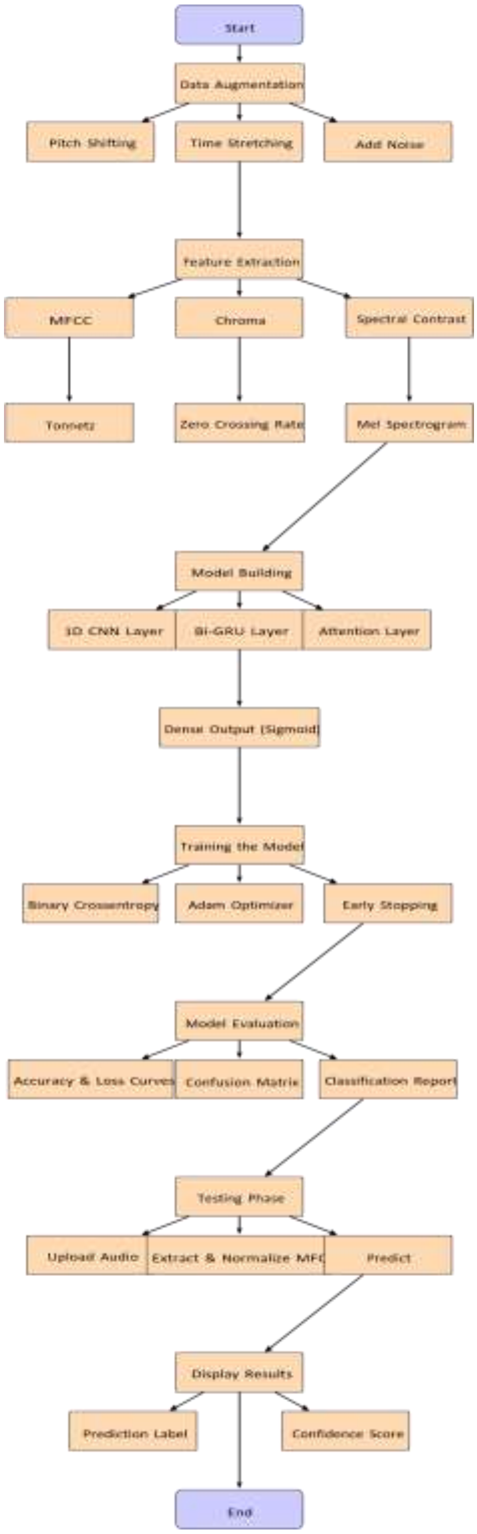


**Fig. 6.** Flowchart of the Pneumonia Detection Pipeline

## IV. RESULTS

*A. Feature Extraction*

The extracted features include MFCC, Chroma, Spectral Contrast, Tonnetz, and Zero Crossing Rate. These features are critical for differentiating between healthy and pneumonia cough sounds, as they capture distinct acoustic properties related to the nature of the cough. An example of the feature extraction grid is shown below:

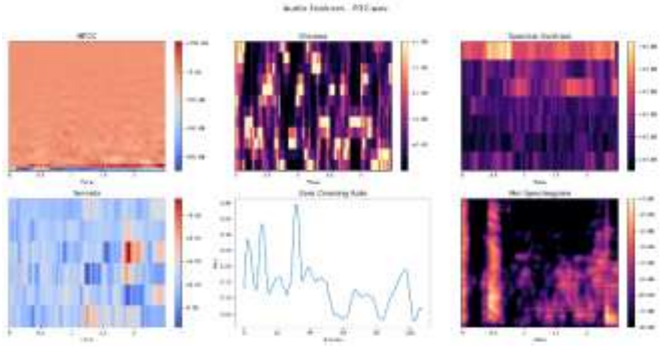

**Fig. 7.** Feature extraction grid showing various features for a sample cough sound.

The results of this study affirm that cough audio signals carry sufficient information to differentiate between healthy individuals and pneumonia patients. The use of Bi-GRU layers allowed the model to retain contextual information from both directions of the sequence, which is crucial for understanding temporal patterns in audio. The attention mechanism further improved model focus by dynamically weighting the most relevant time steps. Data augmentation played a vital role in addressing class imbalance and enhancing generalization. Despite the promising results, the dataset size remains a limitation, and performance could be further improved with larger and more diverse data. Moreover, real-world implementation would require robustness to noise and variability in recording environments.

*B. Model Performance*

The proposed deep learning model achieved a training accuracy of 96.90 percent and a final test accuracy of 97.00 percent. The training loss was observed to be 0.4698, indicating effective convergence. Evaluation metrics including precision, recall, and F1-score demonstrated high performance for both healthy and pneumonia classes, with the model particularly excelling in recall for pneumonia detection. The confusion matrix showed a clear distinction between the two classes, with minimal false positives and false negatives, highlighting the model's robustness and reliability. The following figure shows the precision, recall, and F1-score for both classes:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 0.94 | 0.96 | 54 |
| 1 | 0.96 | 0.98 | 0.97 | 66 |
| accuracy | | | 0.97 | 120 |
| macro avg | 0.97 | 0.96 | 0.97 | 120 |
| weighted avg | 0.97 | 0.97 | 0.97 | 120 |

**Fig. 8.** Confusion Matrix- precision, recall, f1-score, support

The following table shows the model's training accuracy, loss, and test accuracy:

| Metric | Value |
|---|---|
| Training Accuracy | 0.9690 |
| Training Loss | 0.4698 |
| Test Accuracy | 0.97 |

**Table 1.** Model performance metrics.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

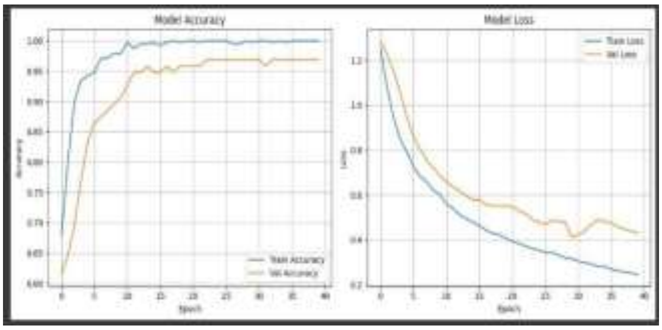


**Fig. 9.** Model Accuracy vs Model Loss

*C. Model loading and testing*

The model is loaded and then the patient's cough sound is uploaded. Model successfully predicts whether the patient has Pneumonia or not.

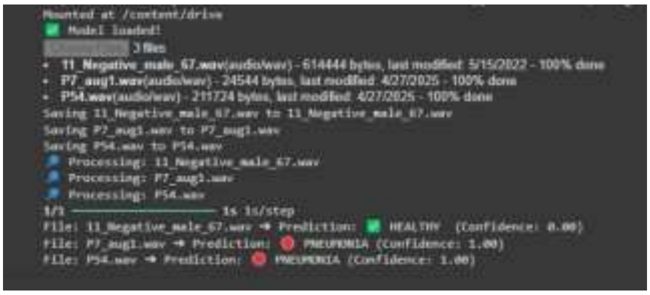


**Fig. 10.** Model Loaded and Tested on Test Cough Sound

*D. Discussion*

The experimental results strongly support the hypothesis that cough audio signals encapsulate sufficient discriminative features to distinguish between healthy individuals and those affected by pneumonia. This finding aligns with prior research by Ghrabli et al. and Kapetanidis et al.,[1],[3] who emphasized that audio-based biomarkers, particularly cough characteristics, can be leveraged for respiratory disease classification. PneumoSense builds upon these insights by employing a deep learning architecture specifically designed to extract and interpret both spectral and temporal nuances within cough recordings. The integration of Bidirectional Gated Recurrent Units (Bi-GRU) played a critical role in modeling temporal dependencies in both forward and backward directions—essential for analysing non-stationary biomedical audio signals such as coughs. This approach echoes findings from Amoh and Odame[9], who demonstrated the importance of temporal modeling in respiratory sound analysis. Additionally, the inclusion of an attention mechanism further refined performance by enabling the model to dynamically focus on the most informative time segments within the audio stream. This dynamic weighting improved both the interpretability and classification accuracy of the system, in line with the strategies discussed by Nguyen and Pernkopf[2] in their work on lung sound classification using transfer learning and attention mechanisms. To address the dataset's class imbalance—270 healthy vs. 82 pneumonia cases—PneumoSense incorporates multiple data augmentation techniques including pitch shifting, time stretching, and additive noise. These strategies not only increased data diversity but also enhanced the model's generalization capability, a technique widely recognized in recent literature for mitigating overfitting in small medical datasets. This approach parallels the recommendations made in multiple reviews, such as those by Srinivasan and Soni,[5] which highlight the importance of data augmentation in improving model robustness and resilience to noise and variability. Despite the high classification accuracy achieved (96.90 percent), limitations persist due to the relatively small and imbalanced dataset. While augmentation helped alleviate some of these challenges, the dataset's size and lack of environmental variability may hinder real-world generalizability. This concern is echoed in the literature, where authors like Sharan and Rahimi-Ardabili call for larger, standardized datasets to improve clinical applicability. Future work should prioritize expanding the dataset through multiinstitutional collaborations and explore semi-supervised or self-supervised learning frameworks to better model realworld recording diversity and patient heterogeneity. Furthermore, to facilitate clinical deployment, rigorous testing under varied environmental conditions—including background noise, device variability, and diverse patient profiles—is essential. Ensuring robustness to these factors will be key to maintaining diagnostic accuracy and reliability outside controlled laboratory environments, especially in resource-constrained or remote settings where PneumoSense could have the greatest impact.

## V. CONCLUSION

This study introduces *PneumoSense*, an AI-driven diagnostic system designed to detect pneumonia through cough sound analysis using a hybrid deep learning architecture that integrates 1D Convolutional Neural Networks (CNN), Bidirectional Gated Recurrent Units (Bi-GRU), and an attention mechanism. By combining spectral feature extraction, bidirectional temporal modelling, and attention-based interpretability, the model demonstrates high classification performance, even when trained on a moderately sized and imbalanced dataset. The results validate the effectiveness of using non-invasive acoustic biomarkers— specifically cough characteristics—for pneumonia detection, reinforcing insights from prior research on respiratory sound-based diagnostics. PneumoSense not only outperforms traditional machine learning approaches in terms of accuracy, precision, recall, and F1-score, but also highlights the potential of deep learning to transform respiratory screening into a scalable, low-cost, and accessible solution for underserved regions. Despite these promising outcomes, limitations remain—chiefly the dataset's size, class imbalance, and lack of environmental variability. Future research should focus on acquiring larger, more diverse datasets, improving noise robustness for real-world deployment, and optimizing inference efficiency for mobile and edge devices. Additionally, expanding the model to support multi-label classification of other co-occurring respiratory conditions such as asthma, bronchitis, and COVID-19 could further broaden its clinical utility. In conclusion, PneumoSense represents a significant step toward democratizing pneumonia screening, with the potential to aid early diagnosis and support remote healthcare delivery—especially in resource-constrained environments where timely medical intervention can be life-saving.

## VI. REFERENCES

[1] Kapetanidis, P., Kalioras, F., Tsakonas, C., Tzam

Kontogiannis, G., Karamanidou, T., Stavropoulos, T. G., and Nik S. (2024). Respiratory diseases diagnosis using audio analy artificial intelligence: A systematic review. *Sensors, 24*(4) https://doi.org/10.3390/s24041173

[2] Nguyen, T., and Pernkopf, F. (2022). Lung sound classification using co-tuning and stochastic normalization. IEEE Transactions on Biomedical Engineering, 69(9), 2872–2882. https://doi.org/10.1109/TBME.2022.3156293

[3] Ghrabli, S., Elgendi, M., & Menon, C. (2024). Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. Scientific Reports, 14, Article 593. https://doi.org/10.1038/s41598-023-50371-2

[4] Sharan, R. V., & Rahimi-Ardabili, H. (2023). Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review. International Journal of Medical Informatics, 176, 105093. https://doi.org/10.1016/j.ijmedinf.2023.105093

[5] Srinivasan, S., & Soni, S. (2021). Towards using cough for respiratory disease diagnosis by leveraging machine learning and deep learning techniques: A survey. Journal of King Saud University - Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2021.07.014

[6] Sharan, P. (2022). Automated discrimination of cough in audio recordings: A scoping review. Frontiers in Signal Processing, 2, Article 759684. https://doi.org/10.3389/frsip.2022.759684

[7] Alqudaihi, K. S., Aslam, N., Khan, I. U., Almuhaideb, A. M., Alsunaidi, S. J., Ibrahim, N. M. A. R., Alhaidari, F. A., Shaikh, F. S., Alsenbel, Y. M., Alalharith, D. M., Alharthi, H. M., Alghamdi, W. M., & Alshahrani, M. S. (2021). Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities. IEEE Access, 9, 102327–102344. https://doi.org/10.1109/ACCESS.2021.3097559

[8] Grant, D. (2024). Towards deployment of aud COVID-19 detection tools (Doctoral dissertation). Johns University. https://jscholarship.library.jhu.edu/items/3efe6ee4c28-9d27-2500068ec076

[9] Amoh, J., & Odame, K. (2016). Deep neural networks for identifying cough sounds. IEEE Transactions on Biomedical Circuits and Systems, 10(5), 1003–1011. https://doi.org/10.1109/TBCAS.2016.2598794

[10] Sharan, R. V. (2023). Cough sound detection from raw waveform using SincNet and bidirectional GRU. Biomedical Signal Processing and Control, 82, 104580. https://doi.org/10.1016/j.bspc.2023.104580. https://www.researchgate.net/publication/362390598_Deep_learning_based_pneumonia_classification_based_on_respiratory_sounds

[11] Abeyratne, U. R., Swarnkar, V., Setyati, A., & Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. Annals of Biomedical Engineering, 41, 2448–2462. https://doi.org/10.1007/s10439-013-0836-0

[12] Saba, E. (2018). Techniques for cough sound analysis (Doctoral dissertation, University of Washington). University of Washington Libraries. https://digital.lib.washington.edu/researchworks/items/67867a38-fadd-4f9c-82c5-1392ee3c0019

[13] Naqvi, S. Z. H., & Choudhry, M. A. (2020). An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis. Sensors, 20(22), 6512. https://doi.org/10.3390/s20226512

[14] Chung, Y., Jin, J., Jo, H. I., Lee, H., Kim, S.-H., Chung, S. J., Yoon, H. J., Park, J., & Jeon, J. Y. (2021). Diagnosis of pneumonia by cough sounds analyzed with statistical features and AI. Sensors, 21(21), 7036. https://doi.org/10.3390/s21217036

[15] Pahar, M., Miranda, I., Diacon, A., & Niesler, T. (2022). Automatic non-invasive cough detection based on accelerometer and audio signals. Journal of Signal Processing Systems, 94(12), 1195–1205. https://doi.org/10.1007/s11265-022-01748-5

[16] Amoh, J., & Odame, K. (2016). Deep neural networks for identifying cough sounds. *IEEE Transactions on Biomedical Circuits and Systems, 10*(5), 1003–1011. https://doi.org/10.1109/TBCAS.2016.2598794