IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.08

Respiratory Sound Classification Using Machine learning

Dr.Sapna M K
Dept. of AI & ML
Vemana Institute of Technology, Bangalore, India
sapnamk@vemanait.edu.in
https://orcid.org/0009-0000-6585-3432
Lolly Manuela R, Denisha V, M C Pallavi Shekhar, Chethana K
Dept. of AI & ML
Vemana Institute of Technology, Bangalore, India
lollyManuela4929@gmail.com, vdenisha5@gmail.com
pallavishekharmc@gmail.com, chethanakrishnappa9@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.12/Issue11/NVAE10087
Research Article| Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue11/NVAE10087
Received:22,October 2025,Revised:28,October 2025, Accepted:31, October 2025, Published Online: 21, November 2025.
https://www.ijirae.com/volumes/Vol12/iss-11/08.NVAE10087.pdf
Citation:Dr.Sapna,Lolly,Denisha,Pallavi,Chethana(2025),Respiratory Sound Classification Using Machine learning, IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 11 of 2025 pages 481-486
doi:>https://doi.org/10.26562/ijirae.2025.v1211.08
BibTeX Key: Dr.Sapna@2025Repiratory
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2025.v1211.08 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Respiratory diseases are high-priority global health issues. Stethoscope - associated auscultator examinations remain almost the only method a physician uses to diagnose respiratory problems. This has been possible with the introduction of computerized respiratory sound analysis (CRSA) that makes it easy to assess the respiratory sounds thanks to an objective, accurate, and quantitative manner. These advances in technologies that include sound analysis as well as medical imaging are very necessary in the face of the global scarcity of qualified healthcare workers. This study presents the authors' use of an elaborate dataset consisting of lung sound recordings and imaging studies either from patients with or without respiratory diseases. The lung sound recordings were made via electronic stethoscopes and omni directional microphones for real-time analysis, while the imaging data were collected from chest X-rays and CT scan images. The study presents a primary focus on deep learning, specifically the extraction of Mel Frequency Cepstral Coefficients (MFCCs) and the Convolutional Neural Network (CNN) model. As part of the study, digital signal processing (DSP) is also implemented with the help of the ESP32. The intention behind the method is to enhance the detection and classification of lung diseases so that the diagnostic accuracy and efficiency would be improved in respiratory care.
Keywords: Respiratory disease, omni directional microphone, convolutional neural network (CNN), digital signal processing (DSP), classification, efficiency.

I. INTRODUCTION

Whatever respiratory disease it may be, it is becoming one of the paramount threats to global health now and for the future. Many of these chronic diseases lead among the causes of mortality and morbidity worldwide: chronic obstructive pulmonary disease (copd), pneumonia, lung cancer, and tuberculosis (tb) [1, 2, 5]. who categorized copd among the leading causes of death in the world [2]. The extent of respiratory complications after covid-19 makes the burden owe an extra dimension to the healthcare systems [1, 8]. Auscultation has been the most prominent non-invasive diagnostic tool for these diseases over centuries. it is the procedure of listening internally with a stethoscope to a patient's own body sounds, chiefly that of lung sounds. Clinicians listen for adventitious (abnormal) sounds such as wheezes, crackles (rales), and rhonchi, which are indicative of pathologies such as narrowing airways, fluid in the lungs, or inflammation [7, 8]. However, the method suffers significant drawbacks. interpretation of the sounds is highly subjective and dependent on the clinician's experience and audio genic purposes; there is much inter-listener variation, with subtle acoustic differences likely to be missed, leading to delayed or inaccurate diagnoses [7]. Recognition of these challenges has put computerized respiratory sound analysis(crsa) into an intense research field [8].crsa evaluates digitization and processing of lung sounds to give an objective, quantifiable, and repeatable technique for diagnosis.

The rise of machine learning (ml), and more recently deep learning (dl), has revolutionized this field. dl models tend to be more powerful than ml models in that they can learn quite complex hierarchical patterns directly from raw data. This property makes dl models exceptionally applicable in analyzing complex audio signals. The image perspective of some researchers on depth learning focuses on medical images such as chest x-rays (cxrs) [1, 3] but there is a huge amount of ongoing work that concentrates on the research of auscultation sounds. This audio-based approach is low-cost, non-invasive, light weight in terms of equipment required (often requiring just a digital stethoscope or smart phone), and free of ionizing radiation. deep learning model scan analyze audio features, such as mel-spectrograms or mel-frequency cepstral coefficients (mfccs), to classify diseases with an accuracy that can meet or even exceed that of trained medical professionals [5, 6]. This paper consolidates findings from a number of recent studies into a general framework for the automatic classification of respiratory diseases from acoustic sounds. We shall discuss key components of the framework, including feature extraction techniques, strategies for data handling, and the various deep learning architectures from standalone CNN's to hybrid RNN-CNN models-that have proven most effective.

II. LITERATURE REVIEW

Recent new methods have been hacked to various feature extraction methods and model architectures towards achieving high accuracy in the classification of the respiratory sounds.

1.Feature Extraction Techniques

Before the deep learning models may work on audio data, raw sound waves should be converted first to the corresponding numerical representation.

Mel Frequency Cepstral Coefficients (MFCCs): This is one of the most dominant and highly powerful features for analyzing audio. The MFCCs extract those ideal parameters of the audio signal just like simulating human auditory perception. They are proven robust and have been successfully used for classifying COPD by cough sounds[2], wheezing detection [7], or many pulmonary disease classifications [5, 8]. Mel-Spectrograms: A spectrogram is a visual representation of the change in time of frequencies contained in a sound. A Mel-spectrogram is a spectrogram where the frequencies are transformed into the Mel scale. This kind of representation, look in the sense of image, is perfect to be used as input for Convolutional Neural Networks (CNNs) [6, 9]. Other Features: Other features have been studied by researchers. Spectral Contrast turned out to be very effective in wheezing detection, especially when combined with a Support Vector Machine (SVM) [7]. Other features, such as Chroma and Tonnetz, are usually more studied in music analysis; however, they have also been analyzed in a multi- feature fusion approach for COPD classification [2]. Mortgage amounts usually range from several studies recommending that information from more than one type of feature source (such as MFCC and Mel) may yield models providing more robust performance through enriching classifiers with complementary information [2].

2. Architecture of Deep Learning Models

The architecture model will play a vital role in the exploitation of the complex patterns formed in respiratory audio. Convolutional Neural Networks (CNN): CNNs are very suitable because they conveniently regard the images as images in two dimensions; thus, they are automatically able to learn spatial hierarchies of features-that is, which band of frequencies and harmonic structures associate with wheezes or crackles [5]. Transfer has been proven successful using other pre trained models such as VGG16 [8] or Res Net [6]that permit training of such powerful models using a scanty amount of medical data.

Recurrent Neural Networks (RNN): Model such as Long Short-Term Memory (LSTM) or, Gated Recurrent Units (GRU), especially on Bi-directional (BiGRUs) configurations, could create new advantages in the processing world of sequential data. Very good in modeling through the temporal dependencies of sounds-in which those sounds construct [2]-has been achieved in the classification of coughs [9].

Hybrid and Cascade Models: The most progressive methodology usually adopts using the architectures combined with all their advantages. For instance a CNN at the start of a cascade will extract relevant features from spectrograms and then serve these as inputs to the RNN that will model their sequential nature. This spatial-then-temporal approach has achieved state-of-the-art results in VGG16-LSTM [8] and BiG RU-CNN cascade [9] models. Other hybrid methods in development are those that include CNNs with Graph Convolutional Networks (GCNs),which can capture complex relational patterns within the audio [4].

3.Data Handling and Augmentation

Medical datasets, notoriously hard to come by, are often very imbalanced (plainly speaking: far more healthy samples than diseased ones).

Data Balancing: To counter act the balance of data, one would employ techniques such as SMOTE (Synthetic Minority Over-sampling Technique) [2], or simply slice and augment data (e.g., time-stretching, flipping spectrograms) [6, 9].

Device Variability: Another factor of interest is the recorded variability presented by the ICBHI dataset, a common benchmark for various stethoscopes [6, 8, 9]. Significant deviation occurs in the significance of variability, spectrum correction therefore can be adopted during preprocessing to normalize the differential frequency responses among devices, thereby enhancing model generalization [6].

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.08

These older techniques exhibited remarkable efficacy, with several studies reporting multi-class classification accuracies on the ICBHI dataset exceeding 90% [5, 6, 8, 9].

III. PROPOSED CLASSIFICATION FRAMEWORK

Drawing findings from literature, we propose a generalized data flow framework on the fact that an entire process has to be implemented to create an automated respiratory sound classification system. This is illustrated further in the subsequent diagrams, where in best practices of data handling, feature extraction, and model-based classification are integrated.

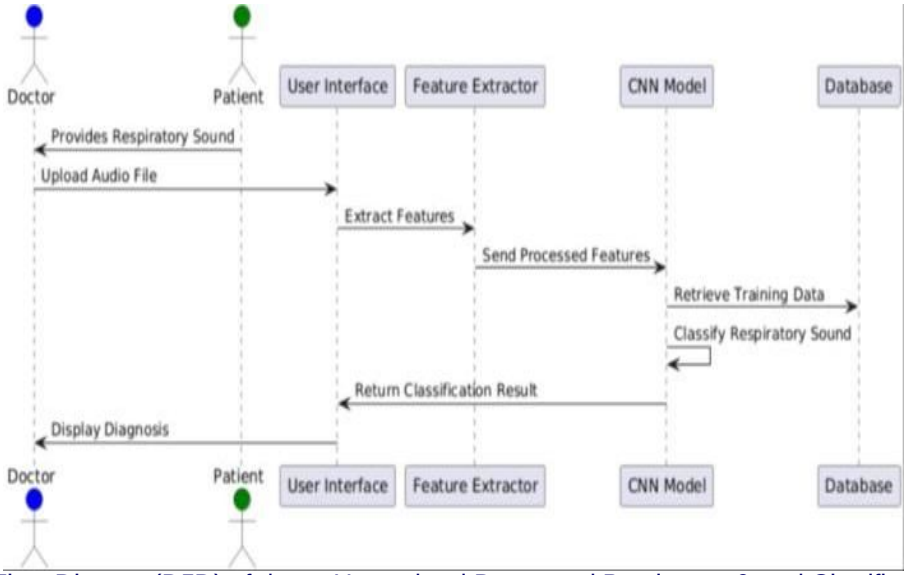


Fig 1: Data Flow Diagram (DFD) of the an Upward and Downward Respiratory Sound Classification System. (Reproduced from[10])

Figure1provides a high-level overview of the entire process, starting at patient input, and ending in the final display of diagnosis.

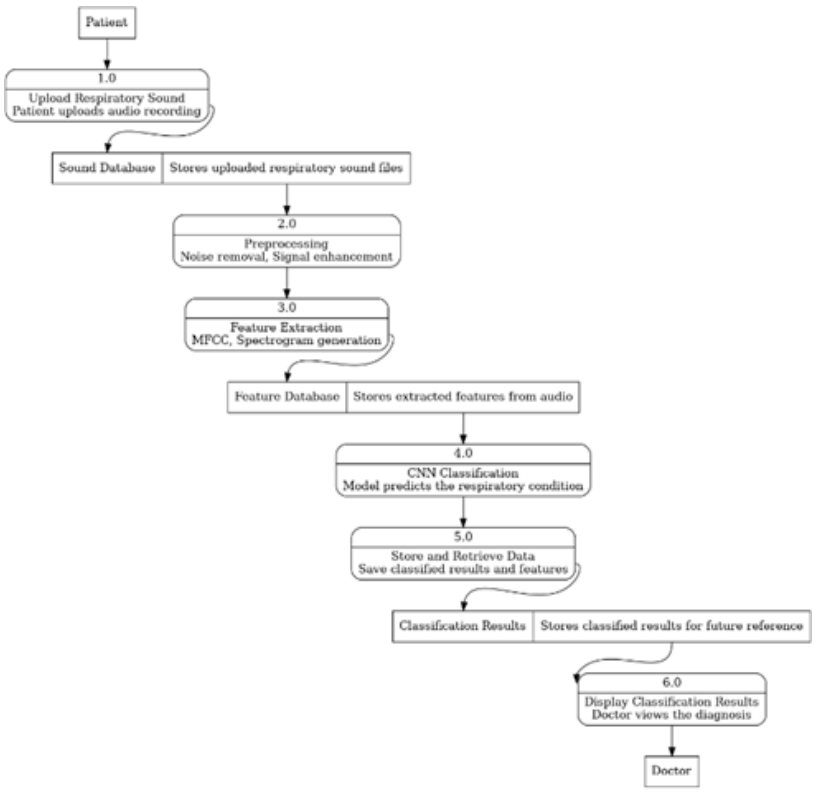


Fig 2: Sequence Diagram.(Reproduced from[10])

1. Patient Upload: The process is initiated when the user (patient or clinician) uploads an audio recording of a respiratory sound (eg: breathing sound, cough) [8].
   Sound Database: This raw audio file is stored in a Sound Database securely for later retrieval and maintenance [4].
2. Preprocessing: The system fetches the file and subjects it to necessary preprocessing steps that include noise suppression (e.g., filtering out background sounds or heart sounds) and signal enhancement to improve the extraction of the respiratory sounds [6, 8].
3. Feature Extraction: It is then converted into a robust, feature-rich representation. The major features that literature specifies are MFCCs [2, 5, 8] and Mel-spectrograms [6, 9]. This is how a"sound finger print"comes into existence for the model's assessment.
   Feature Database: The additional features are stored in their specific database.
4. Classifier model: The features are processed into the core of the system, a trained deep-learning model. The diagram contains a reference to CNN; however, it must be noted that such an engine would work most efficiently as a hybrid architecture (e.g., CNN-LSTM [8] or CNN-BiGRU [9]) which enables both spectral and temporal analysis. The model predicts the condition of respiration, each represented with a letter, that is either healthy, COPD, Pneumonia, or Bronchiolitis [5, 9].
5. Storage and Display: Now the model prediction is stored under Classification Results into a database such that it is delivered to clinicians (e.g., a doctor) in a clear, interpretable format. It has now completed the diagnostic loop [10].

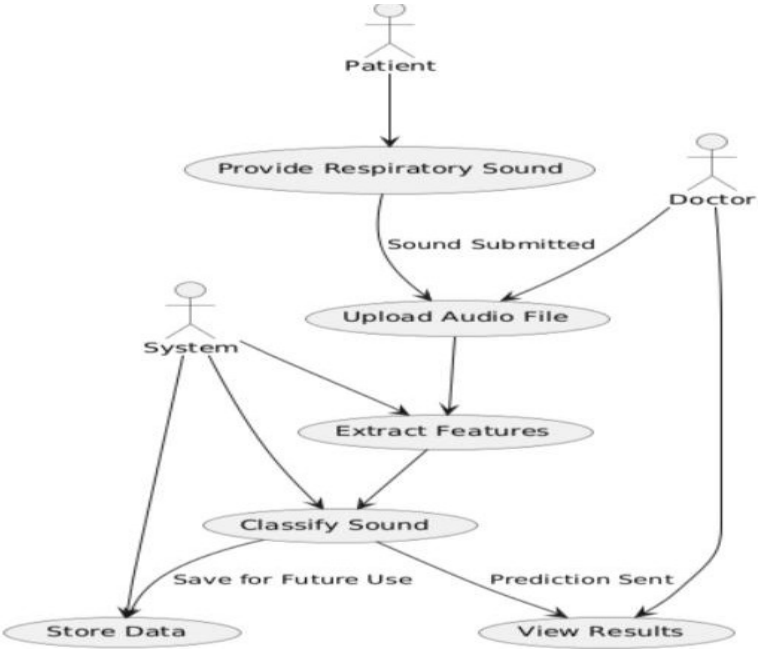


Fig 3: Use Case Diagram.(Reproduced from[10])

Figure 2 gives the primary actors and interactions. The Patient is the one that provides the sound input for processing by the System. The end-user is the doctor (or medical user), as they request the analysis and finally view the classified results.

Figure 3 shows the flow of events over time. Upon uploading a patient file via the user interface, the feature extractor gets triggered. The extracted features are processed through the CNN Model (i.e.our classification engine), which retrieves stored patterns (or even patient histories) from a database. A classification result is returned, which is finally displayed to the doctor via the user interface.

IV. HARDWARE AND SYSTEMS COMPONENTS

To implement the framework described above, hardware can be classified into two broad categories: data acquisition hardware to capture sounds and computational hardware to process the data and train the models.

Data Acquisition Hardware: The initial data capture (Step1.0) has been done with high-fidelity audio recording devices. For common reference are digital stethoscopes-composite types such as the Littmann 3200 [2]-or with other electronic stethoscopes [7]. Some have even used high-quality electret microphones [4]. Choice of hardware is crucial since the principal challenge is device variability; this will practically also be extended to high-quality microphones evaluated within smart phones [8]. integrated Neural Processing Unit (NPU)) for faster, offline analysis [8].

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.08

Fig 4: omni directional microphone

Computational Hardware: The highly computing-intensive steps, particularly Feature Extraction (Step 3.0) and Model Classification (Step 4.0), will demand considerable computational power.

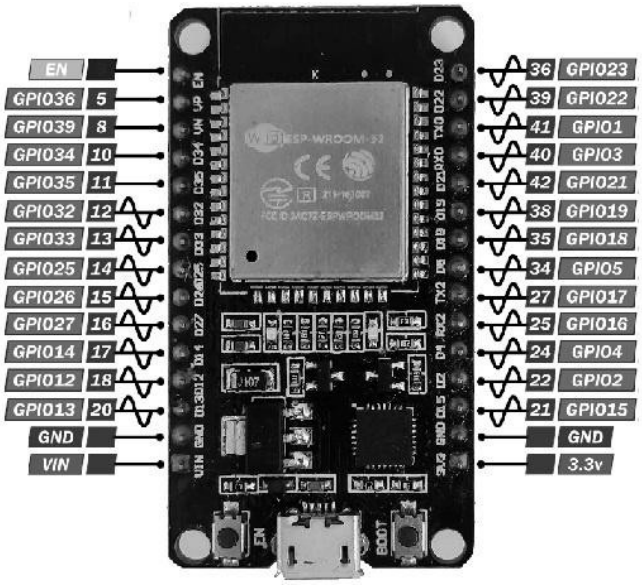


Fig 5: ESP32

Model training: Training is usually performed offline on a high-performance computing infrastructure, e.g., servers or workstations with powerful GPUs specifically created to accelerate the complex matrix operations involved in training deep neural networks [1, 5, 8]. Model Deployment (Inference): The trained model for the system to become useful must be deployed on reasonably accessible hardware. The frame work is meant to be deployed as an online tool or a mobile application [8]. In this case, the User Interface (Figure3) would run on the user's local device (e.g., a smart phone or PC). The audio file is uploaded to a central server where the inference takes place by the trained model (running on CPUs or GPUs), the result is sent back to the User Interface. Future iterations aim for lightweight models that can run directly on-device (e.g.,on a smart phone's

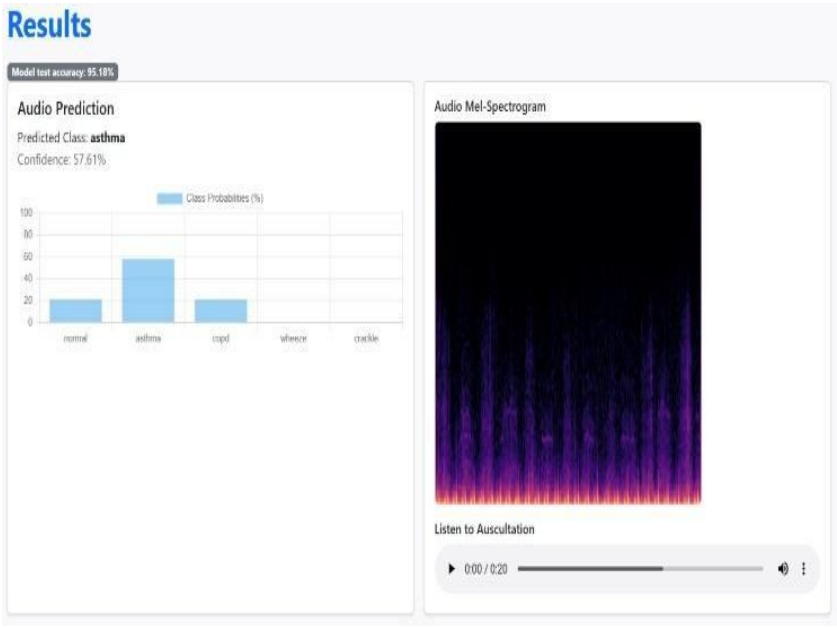
V. DISCUSSION

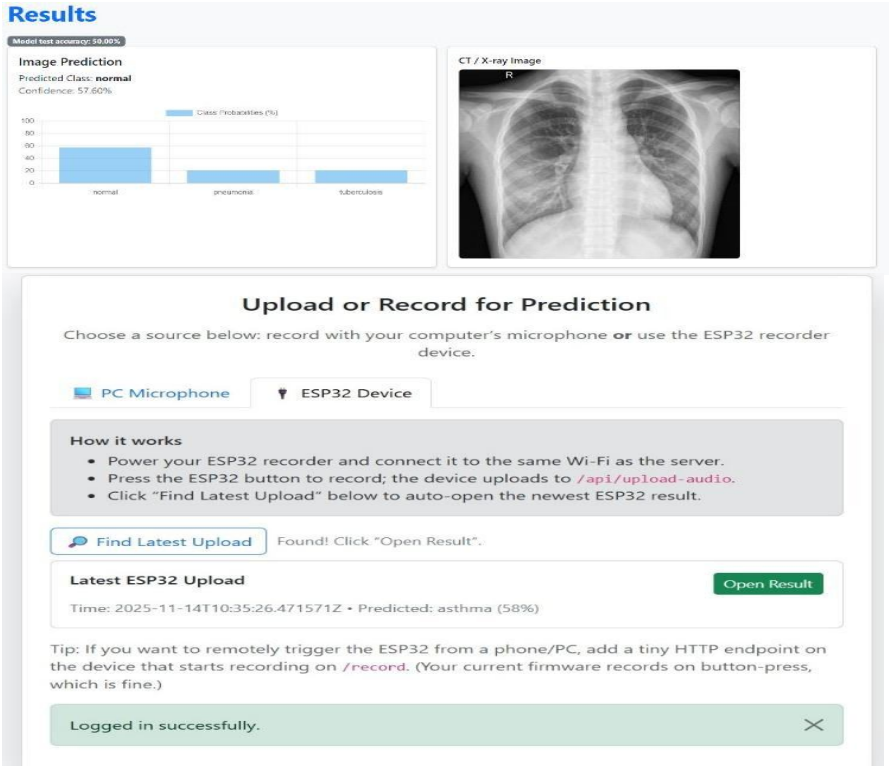
The synthesized frame work presented in Section3 offers a robust and non-invasive pathway for respiratory disease diagnosis. The beauty of this approach to deep learning lies in its ability to objectively discern complex acoustic patterns that human ears often overlook. The literature review has consistently shown that model architecture is a critical parameter for performance. While simple CNNs work perfectly [5], the top performance attained by hybrid model architectures, such as VGG16-LSTM [8], and BiGRU-CNN [9], prove that both "what" (spectral features via CNN) and "when"(temporal sequences via RNN) aspects of respiratory sounds need to be modeled. Data-centric aspects simply cannot be overemphasized. The high accuracies discussed in the papers are not only the results of performing advanced modelling, but very much the result of careful data handling procedures.

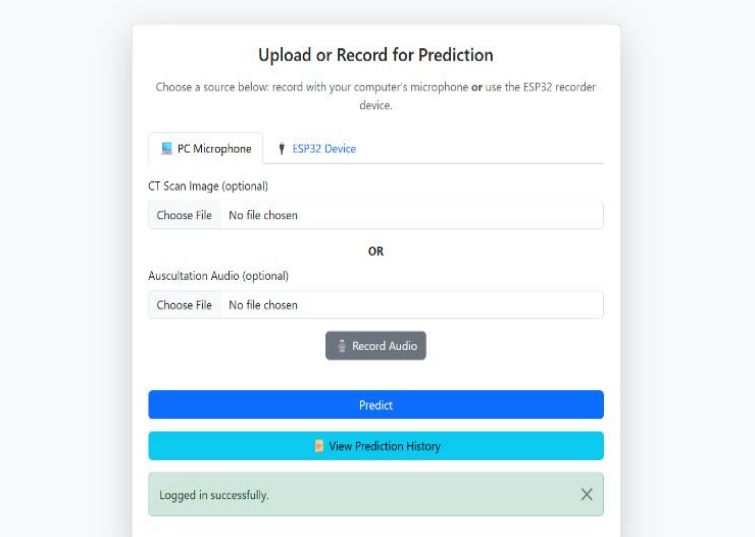
IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.08

Techniques to deal with class imbalance, like SMOTE [2] and data slicing [9],are crucial to ensure the model learns to identify rare pathologies well. Likewise, applying spectrum corrections for different recording devices represents an equally crucial procedure to obtain a robust model that can be placed into clinical practice with a range of different equipment [6].With successful implementation, such a frame work will carry out the transition of respiratory diagnosis from subjective art into objective data-driven science. A tool built on this framework could serve as a powerful decision-support system for clinicians, provide a low-cost screening tool for remote and underserved areas via tele health platforms, and enable patients with chronic conditions (like COPD or asthma) to monitor their status from home [8].

VI.RESULTS:







VII. CONCLUSION

This paper synthesized recent advancements in deep learning to propose a unified framework for the automated classification of respiratory diseases from auscultation sounds. By leveraging robust preprocessing, advanced feature extraction (primarily MFCCs and Mel-spectrograms), and powerful hybrid deep learning models (such as CNN-LSTM hybrids), it is possible to create systems that can classify a wide range of pulmonary conditions with high accuracy. The illustrated dataflow, use case, and sequence diagrams provide a clear blueprint for such a system. The consistent success reported in the literature, with accuracies often surpassing 90% [5, 8, 9], confirms the viability of this approach. This technology is not intended to replace clinicians, but to augment their capabilities, providing an objective, cost-effective, and highly accessible tool to aid in the early and accurate diagnosis of respiratory diseases, ultimately leading to better patient outcomes.

REFERENCES

1. S.Ashwini,J.R.Arunkumar,R.Thandaiah Prabu,N.H. Singh,andN.P.Singh,"Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network," Soft Computing, 2023.
2. P. J. Patel et al., "Multi feature fusion for COPD classification using Deep Learning algorithms," Journal of Integrated Science and Technology, 2024.
3. R.N.Sadoon and A.M.Chaid,"Deep Learning Technique for The Classification of Lung Diseases from X-ray Images,"Journal of Al-Qadisiyah for Computer Science and Mathematics, 2024.
4. R.Anaadumba, N.A.Belabbaci, and M.A.U. Alam, "Enhancing Respiratory Disease Diagnosis with Graph Convolutional Networks Through Non-Speech Audio Synthesis,"2025 IEEE/ACM Conference on Connected Health (CHASE), 2025.
5. O.Raju and D.Ramesh, "An Enhanced deep learning approachforDiseaseClassificationfromRespiratorysound,"2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT), 2024.
6. T.Nguyen and F.Pernkopf, "Lung Sound Classification Using Co-Tuning and Stochastic Normalization,"IEEE Transactions on Biomedical Engineering, 2022.
7. H.J.Moonetal.," Machine Learning-Driven Strategies for Enhanced Pediatric Wheezing Detection," Research Square Preprint, 2024.
8. A.Metal.,"Predicting the Severity of Pulmonary Disease from Respiratory Sounds using ML Algorithms,"2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), 2025.
9. Y.Zhu and W.Xu, "Research on Classification of Respiratory Diseases Based on BiGRU and CNN Cascade Neural Network,"2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2021.