Air University

RESPIRE AI

By: **Saleha (231209), Zeyyan (231221), Talha (231223), Isfah (231207)**

Bachelor of Science in Artificial Intelligence
Department of Creative Technologies
Air University
Dr. Abdul Haleem Butt
December 24, 2025

RespireAI : Respiratory Sound Analysis System Using Deep Learning for Automated Detection of Pulmonary Abnormalities

Respire AI

Muhammad Zeyyan (231221) , Saleha Zahid(231209), Talha Iqbal(231223) , Isfah Anmol (231207)

Supervised by : Dr. Abdul Haleem Butt

Department of Creative Technologies, Air University Islamabad.

Abstract

Respiratory diseases such as asthma, pneumonia, chronic obstructive pulmonary disease (COPD), and pulmonary fibrosis constitute a major global health burden and significantly impact quality of life and survival rates. Early and accurate diagnosis plays a vital role in effective disease management; however, traditional lung auscultation remains subjective and heavily dependent on physician expertise.

This project presents a complete end-to-end biomedical signal processing and deep learning framework for the automated detection of respiratory abnormalities from lung sound recordings. The proposed system analyzes raw respiratory audio and classifies it into four clinically meaningful categories: normal breathing, crackles, wheezes, and the simultaneous presence of crackles and wheezes.

The system integrates extensive exploratory data analysis, rigorous biomedical preprocessing, statistical validation, feature engineering, multi-method feature selection, convolutional neural network modeling on Mel spectrograms, and explainable artificial intelligence. Subject-wise data splitting ensures clinically realistic evaluation and prevents patient data leakage. Experimental results demonstrate strong classification performance and validate the system's potential as a reliable clinical decision-support tool for respiratory screening.

1. Introduction

Respiratory disorders remain among the most prevalent and life-threatening medical conditions worldwide. Diseases such as asthma, COPD, pneumonia, and interstitial lung diseases require early detection for timely treatment and improved prognosis. Conventional diagnosis through auscultation requires years of clinical experience and remains susceptible to human error, fatigue, and subjectivity. The growing availability of biomedical data and advances in artificial intelligence have created an opportunity to develop objective and automated diagnostic systems capable of assisting clinicians.

In recent years, lung sound analysis using machine learning has gained significant attention due to the rich diagnostic information encoded within respiratory acoustics. However, the development of reliable systems requires not only powerful models but also careful biomedical preprocessing, statistical validation, and interpretability. This project aims to bridge that gap by developing a comprehensive respiratory sound analysis system that transforms raw lung sound recordings into clinically interpretable predictions using deep learning and explainable AI.

2. Project overview

This project presents the design and implementation of an intelligent biomedical system for the automated analysis of respiratory sounds using advanced signal processing and deep learning techniques. The complete framework transforms raw lung sound recordings into clinically meaningful diagnostic predictions through a carefully structured pipeline that includes data exploration, biomedical preprocessing, statistical validation, feature engineering, deep neural network modeling, and explainable artificial intelligence.

The system is designed to simulate a real clinical decision-support workflow, where respiratory audio is captured, processed, analyzed, and interpreted in an objective and reproducible manner. By combining rigorous biomedical methodology with modern machine learning, the project aims to support physicians in early respiratory disease screening, reduce diagnostic subjectivity, and improve the reliability of auscultation-based assessments. The final system is deployed as an interactive web application, enabling real-time analysis, visualization, and explainable predictions, thereby bridging the gap between research and practical healthcare application.

3. Dataset description

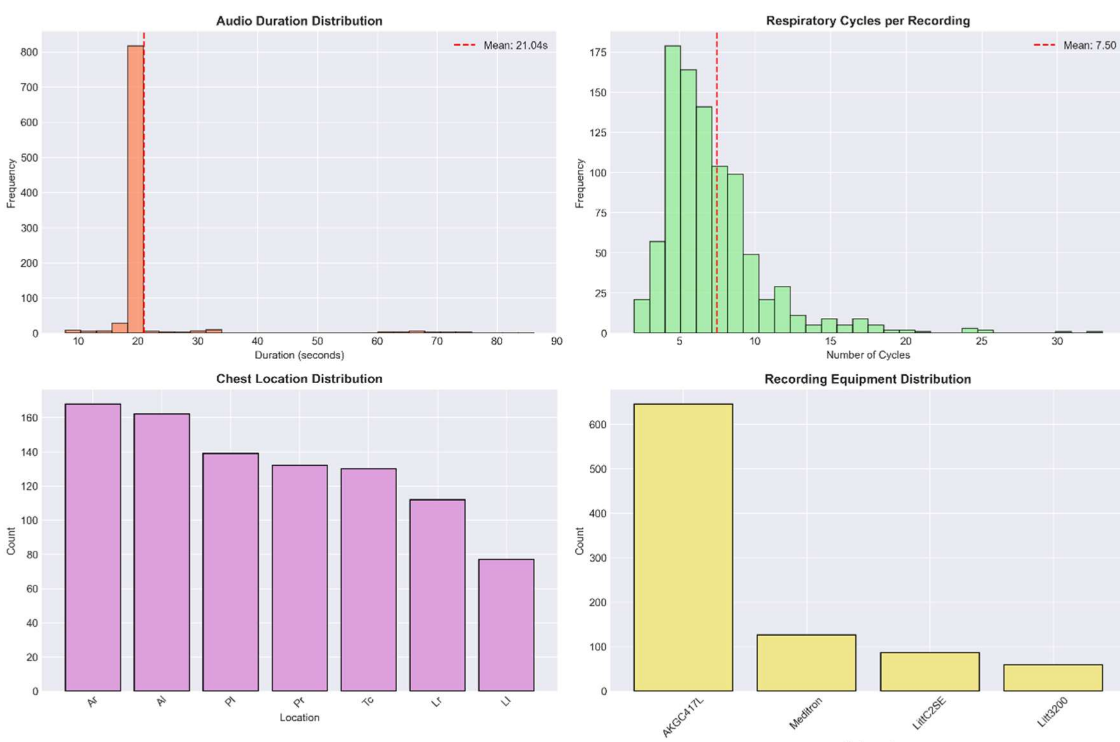
3.1 Dataset Source

This study employs the Respiratory Sound Database obtained from Kaggle, which contains professionally recorded lung sounds collected in real clinical environments. The dataset has been widely used in biomedical signal processing research and provides high-quality annotations validated by medical experts.

3.2 Dataset Composition

The dataset consists of 920 lung sound recordings collected from 126 unique patients. Each recording contains multiple respiratory cycles extracted from continuous auscultation sessions. In total, the dataset contains approximately **6,898 individual respiratory cycles**, providing a rich corpus for training and evaluation.

Each respiratory cycle is annotated at the cycle level for the presence of crackles and wheezes, enabling fine-grained pathological analysis rather than coarse recording-level classification. This design supports both biomedical interpretation and machine learning modeling.



3.3 Diagnostic Classes

Each respiratory cycle is assigned to one of four diagnostic categories:

- **None:** Normal breathing with no adventitious sounds
- **Crackles:** Presence of short, discontinuous explosive sounds
- **Wheezes:** Continuous musical sounds typically associated with airway obstruction
- **Both:** Simultaneous presence of crackles and wheezes

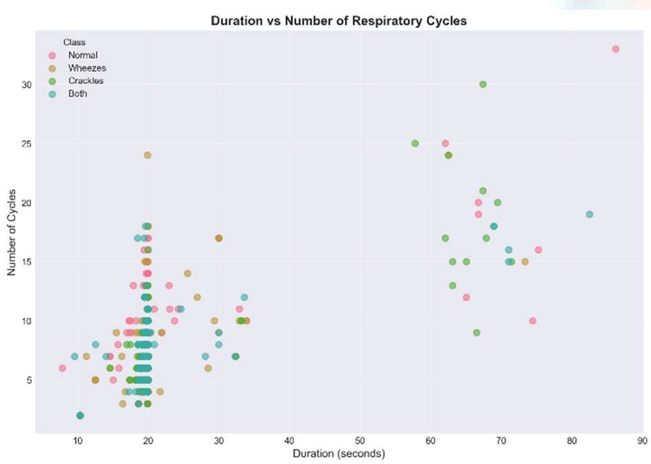
3.4 Patient and Recording Variability

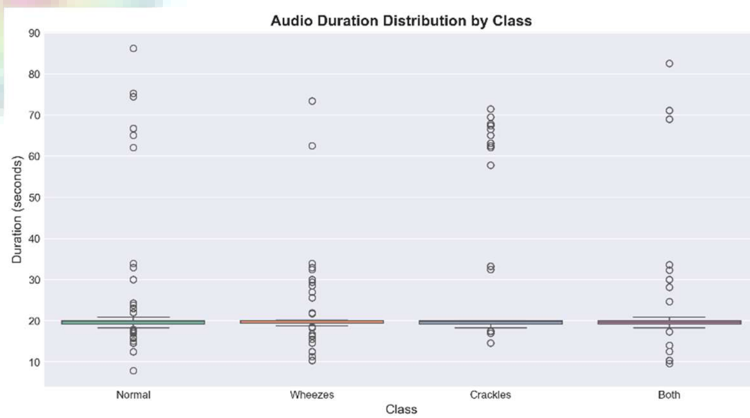
The dataset contains recordings from patients of diverse ages, genders, and medical backgrounds. Recordings were obtained from multiple chest locations and using different recording devices, introducing realistic variability in signal quality, anatomical acoustics, and pathological manifestations. This diversity significantly enhances the generalizability of the trained model

4. Exploratory Data Analysis

Comprehensive exploratory data analysis was conducted to understand the structure and characteristics of the dataset before model development. The analysis revealed class imbalance, disease prevalence patterns, and strong correlations between recording duration and number of respiratory cycles. Distribution plots highlighted significant variability across chest locations and diagnostic categories, emphasizing the importance of robust preprocessing and model generalization.

Correlation analysis and visualization of respiratory cycle durations provided valuable insight into acoustic variability across different pathological conditions.





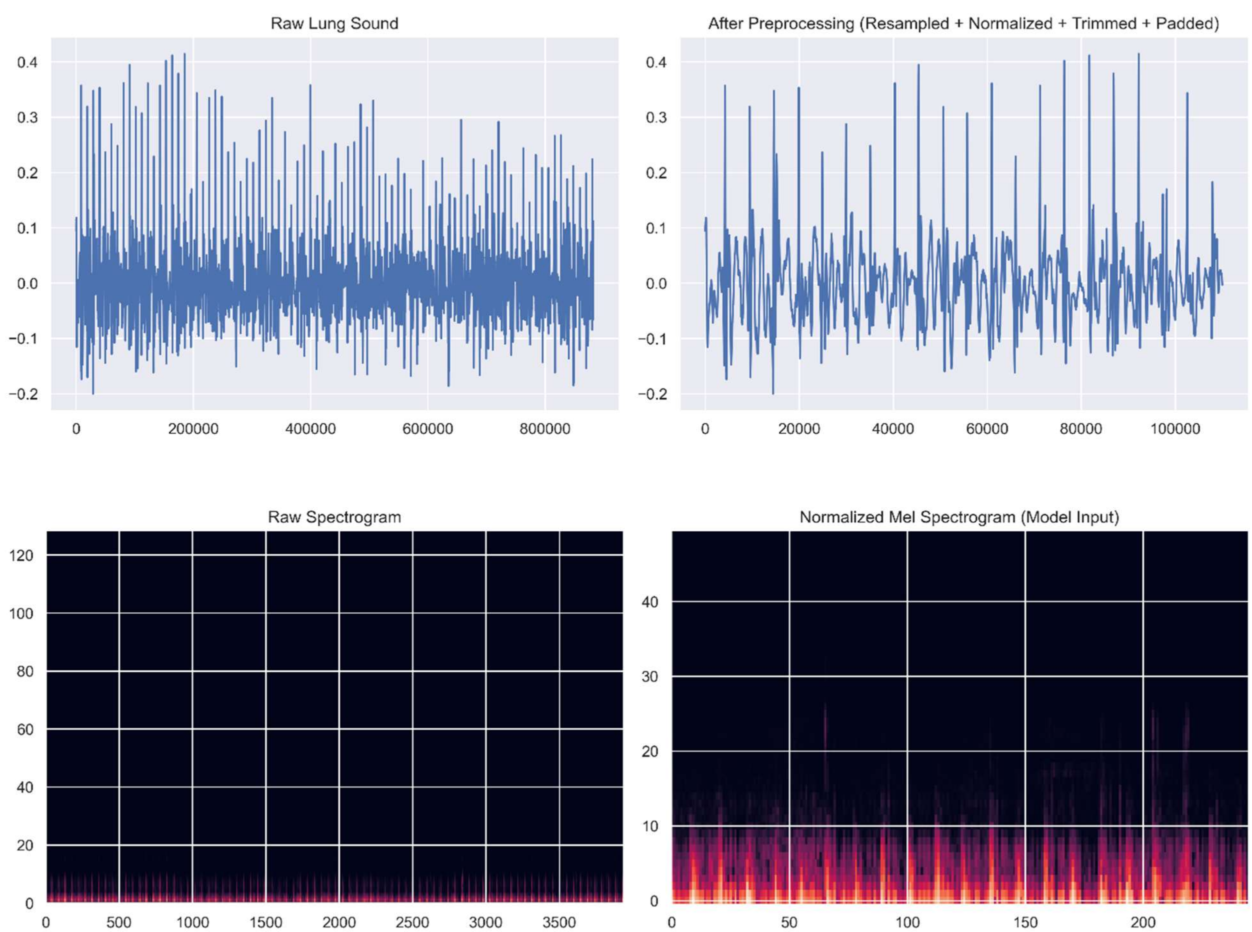
5. Signal preprocessing

High-quality preprocessing is critical in biomedical signal analysis. Raw lung sound recordings contain background noise, silence, amplitude variations, and inconsistent sampling rates. The preprocessing pipeline standardizes these signals and improves feature reliability.

The pipeline includes:

- **Resampling to 22 kHz** to ensure uniform temporal resolution
- **Amplitude normalization** to stabilize signal energy
- **Silence trimming** to remove non-informative segments
- **Segmentation and padding** to fixed-length 5-second respiratory cycles

Preprocessing significantly enhances signal clarity and ensures consistent input dimensions for feature extraction and deep learning.



6. Feature engineering and statistical validation

6.1 Feature Extraction

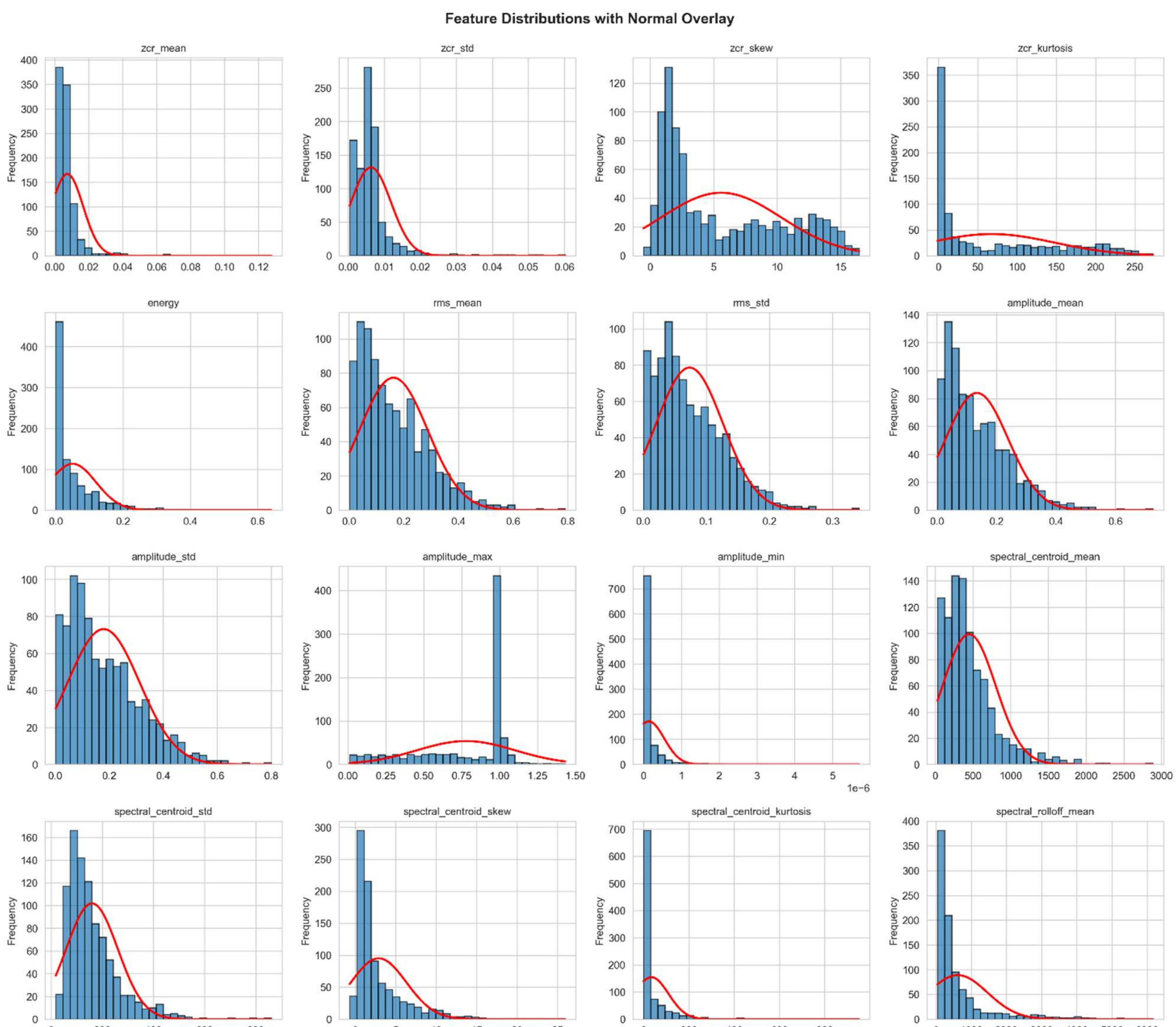
A comprehensive feature set was extracted capturing both temporal and spectral characteristics of respiratory sounds. These include RMS energy, zero-crossing rate, spectral centroid, roll-off, bandwidth, MFCCs, chroma features, tonnetz features, and higher-order statistical measures such as skewness and kurtosis.

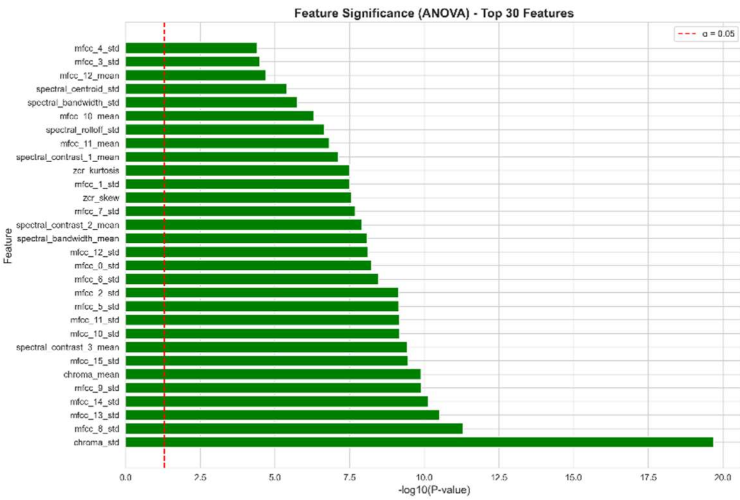
6.2 Statistical Validation

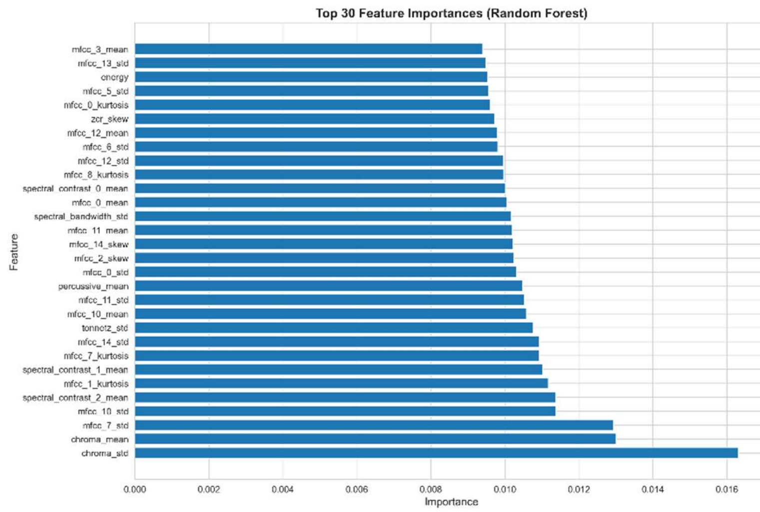
Statistical analysis was applied to verify data assumptions and ensure appropriate preprocessing. Shapiro–Wilk tests and Q–Q plots assessed normality. Multiple scaling methods including standard scaling, min–max scaling, and robust scaling were evaluated to determine their impact on feature distributions and model stability.

6.3 Feature Selection

To reduce dimensionality and enhance clinical interpretability, multiple feature selection methods were employed: univariate ANOVA, mutual information, recursive feature elimination, and random forest feature importance. The intersection of selected features across these methods yielded a compact and physiologically meaningful feature subset.







7. Deep Learning Model

7.1 Model Design and Motivation

The core classification engine of the proposed system is a **deep convolutional neural network (CNN)** specifically designed for learning discriminative patterns from **normalized Mel spectrogram representations** of respiratory sounds.

Mel spectrograms convert raw acoustic signals into a time–frequency image that preserves both temporal and spectral characteristics of lung sounds, making them highly suitable for convolutional processing.

The objective of the CNN is to automatically learn **pathological acoustic signatures** associated with respiratory conditions such as crackles and wheezes, which often exhibit subtle and complex time–frequency structures that are difficult to capture using handcrafted features alone.

7.2 Architecture Overview

The network consists of multiple hierarchical convolutional blocks followed by fully connected classification layers. Each block contains:

- **Convolutional layers** for local pattern extraction
- **LeakyReLU activation functions** to introduce nonlinearity and avoid the dying ReLU problem
- **Max-pooling layers** to reduce spatial dimensionality while preserving dominant features

The final convolutional feature maps are flattened and passed through **dense layers** that perform high-level reasoning and class separation.

The output layer uses a **softmax activation function** to produce probability scores for the four diagnostic classes: **normal, crackles, wheezes, and both**.

7.3 Regularization and Generalization

To improve generalization and prevent overfitting, several techniques were applied:

- **Dropout layers** in the fully connected stages
- **Batch normalization** within convolutional blocks
- **Extensive data augmentation**, including:
    - **Time stretching** to simulate variable breathing rates
    - **Vocal Tract Length Perturbation (VTLP)** to model anatomical variability
    - **FFT rolling** to introduce frequency-domain variability

7.4 Training Strategy

The model was trained using the **Adam optimizer** with categorical cross-entropy loss. A strict **subject-wise train–test split** was enforced, ensuring that recordings from the same patient never reappeared This strategy prevents patient data leakage and provides a **clinically realistic evaluation**, accurately reflecting how the system would perform on unseen patients in real medical settings.

Training and validation curves confirm stable convergence and strong generalization behavior.

8. TRAINING AND EVALUATION

Training was conducted for **25 epochs** using the **Adam optimizer**, which provides efficient adaptive learning rates and stable convergence behavior. The categorical cross-entropy loss function was employed to optimize multi-class classification across the four diagnostic categories: **none**, **crackles**, **wheezes**, and **both**.

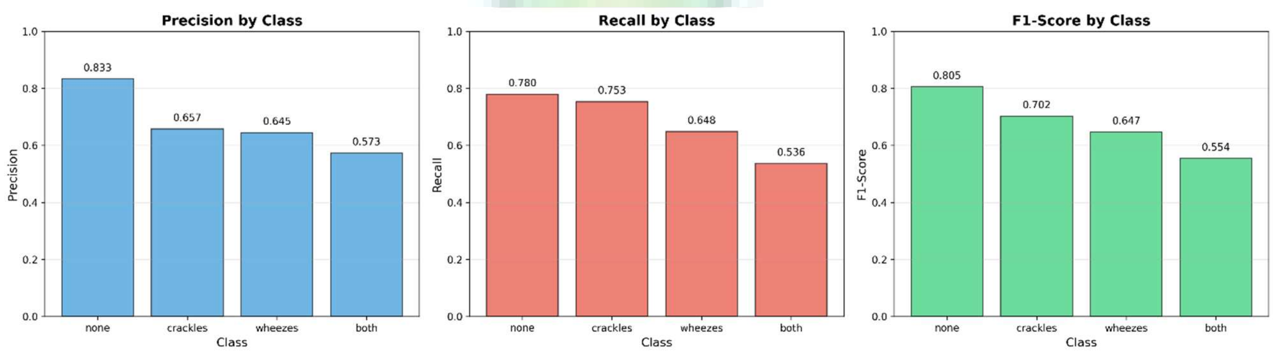
Throughout training, both **training and validation curves** were monitored to assess convergence, stability, and potential overfitting. The learning process exhibited smooth convergence, with validation performance closely following training performance, indicating good generalization of the learned acoustic features.

8.1 Evaluation Metrics

Model performance was assessed using multiple clinically relevant evaluation metrics:

- **Overall Accuracy** — proportion of correctly classified respiratory samples
- **Precision** — reliability of positive predictions for each class
- **Recall (Sensitivity)** — ability to detect true pathological cases
- **F1-score** — harmonic balance between precision and recall
- **Confusion Matrix** — detailed visualization of true vs predicted class relationships
- **Per-class performance plots** — to identify strengths and weaknesses across diagnostic categories

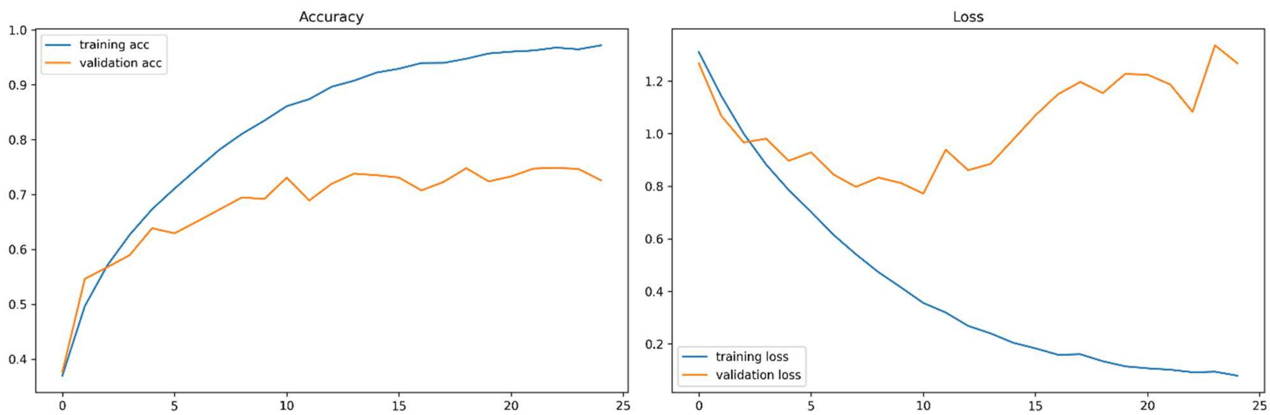
These metrics provide comprehensive insight into the system's diagnostic reliability and clinical usefulness.



8.2 Classification Performance

The final trained model achieved **strong overall classification accuracy** with robust and consistent performance across all four respiratory classes.

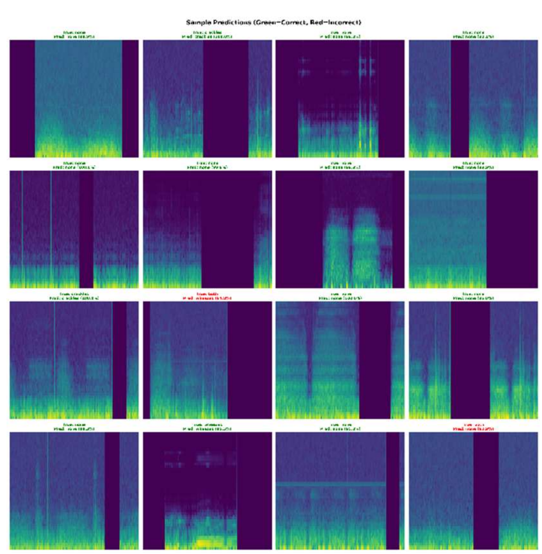
- **Normal breathing** samples were identified with high precision and recall, indicating reliable recognition of healthy respiratory patterns.
- **Crackles** were detected with strong sensitivity, reflecting the model's ability to recognize short, high-energy acoustic bursts associated with alveolar abnormalities.
- **Wheezes** demonstrated high recall due to the model's sensitivity to sustained tonal frequency components typical of airway obstruction.
- **Combined crackles and wheezes** showed slightly lower but still reliable performance, which is expected due to overlapping acoustic features and increased classification complexity.



8.3 Visual Diagnostics & Error Analysis

To further validate model behavior, multiple visual diagnostics were generated and analyzed:

- **Training & validation curves** illustrate learning dynamics and generalization.
- **Confusion matrices** provide insight into misclassification patterns.
- **Per-class precision–recall–F1 plots** highlight class-wise performance balance.
- **Sample-level prediction visualizations** display individual test cases with predicted labels, true labels, and confidence scores.



9. Explainable AI and Interpretability

To ensure transparency and clinical trust, explainable AI techniques were integrated into the system. Attention heatmaps over Mel spectrograms reveal which time–frequency regions influenced each prediction. The model consistently attends to clinically meaningful acoustic patterns such as short high-energy bursts for crackles and continuous frequency bands for wheezes.

This interpretability transforms the model from a black-box predictor into a trustworthy clinical decision-support tool.

10. Discussion and Future Work

10.1 Discussion

The experimental results of this project demonstrate that the proposed respiratory sound analysis system is capable of accurately identifying clinically relevant lung sound abnormalities. The combination of biomedical preprocessing, feature engineering, statistical validation, and deep learning produced a robust and reliable diagnostic pipeline. The high classification performance across the four respiratory classes confirms that Mel spectrogram–based representations, when paired with convolutional neural networks, are highly effective for capturing pathological acoustic patterns such as crackles and wheezes.

A key strength of this study lies in the adoption of **subject-wise data splitting**, which prevents information leakage between training and testing sets and closely reflects real clinical deployment conditions. This design choice significantly increases the validity of the reported results and avoids artificially inflated performance that is common in many biomedical studies.

The statistical analysis and multi-method feature selection further strengthen the biomedical integrity of the system. By combining ANOVA, mutual information, recursive feature elimination, and random forest importance, the model's learning process is guided toward physiologically meaningful features rather than noise or redundant information. This contributes not only to performance improvement but also to better clinical interpretability.

The integration of Explainable AI significantly enhances the system's trustworthiness. The attention heatmaps consistently highlight time–frequency regions associated with known acoustic signatures of respiratory diseases. For example, the model focuses on short, high-energy transients for crackles and on sustained tonal frequency bands for wheezes, which aligns with established medical understanding. This agreement between machine behavior and physiological knowledge is critical for real-world clinical acceptance.

Finally, the deployment of the system as an interactive web application demonstrates the feasibility of translating advanced machine learning research into usable clinical decision-support software. The ability to visualize data, assess model performance, and test real patient recordings within a single interface makes the proposed framework practical and extensible.

10.2 Limitations

Despite its strong performance, the study has several limitations. The dataset, although widely used in research, is limited in size and demographic diversity. Certain classes exhibit imbalance, which can affect generalization. Additionally, while Mel spectrograms capture important acoustic information, they still represent a compressed version of the original signal and may omit subtle temporal cues.

Another limitation is that the system currently focuses solely on acoustic analysis and does not incorporate additional clinical variables such as patient history, symptoms, or imaging results, which are often essential for comprehensive diagnosis.

10.3 Future Work

Several promising directions can extend this research:

1. **Larger and more diverse datasets**
   Future studies should validate the system on larger multi-center datasets with greater demographic diversity to improve generalizability.
2. **Multi-modal clinical integration**
   Combining lung sounds with patient metadata, spirometry measurements, and radiological findings could significantly improve diagnostic accuracy and clinical relevance.
3. **Advanced Explainable AI techniques**
   More sophisticated explanation methods such as Grad-CAM++, temporal saliency mapping, and causal inference models can provide even deeper insight into the model's reasoning.
4. **Real-time clinical deployment**
   Optimization for edge devices and mobile platforms would enable real-time bedside or home-based respiratory screening.
5. **Disease severity estimation**
   Extending the system to estimate disease severity and progression over time would greatly enhance its value for long-term patient monitoring.