Internship Project
Report

Respiratory Disease Classification Through Feature Engineering

Submitted by

Kushagra Singh
M.Sc. Data Science
Defence Institute of Advanced Technology
Pune, India

Under the guidance of

Pankaj Kumar
Scientist E
Information Technology Goverance

Solid State Physics Laboratory
Defence Research and Development Organisation (DRDO)
Delhi, India

July 2025

Acknowledgment

I would like to express my sincere gratitude to **Mr. Pankaj Kumar**, **Scientist E**, Inforface Technology Governance, SSPL, DRDO, for providing me with the invaluable opportunity to undertake this internship under his guidance. His exceptional mentorship, insightful feedback, and constant encouragement throughout the course of this research have been instrumental in shaping the quality and direction of my work.

I also extend my sincere appreciation to the entire team at SSPL, DRDO, for creating a dynamic and intellectually stimulating research environment. Their cooperation, shared technical expertise, and constructive discussions significantly enriched my learning experience and helped me navigate the challenges encountered during the project.

Lastly, I would like to acknowledge the unwavering support and motivation from my family and peers. Their patience, understanding, and positive reinforcement played a vital role in helping me remain focused and dedicated throughout the internship period.

Abstract

Respiratory diseases continue to pose a significant threat to public health across the globe, contributing to a high burden of morbidity and mortality. Early and accurate detection of these conditions is crucial to improving patient outcomes and reducing healthcare costs, particularly in resource-constrained settings. This internship project aims to explore automated classification of respiratory diseases using machine learning techniques applied to acoustic features extracted from patient breathing sounds.

The research is grounded on the ICBHI 2017 Respiratory Sound Database, comprising annotated audio recordings from real patients. The project implements both binary (one-vs-rest) and multiclass classification frameworks. Feature engineering techniques were employed to extract discriminative features such as Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, Mel Spectrograms, Spectral Contrast, and Tonnetz from each audio file using the Librosa library in Python.

In the binary classification phase, individual Convolutional Neural Network (CNN) models were trained for each respiratory condition—including Chronic Obstructive Pulmonary Disease (COPD), Bronchiectasis, Bronchiolitis, Upper Respiratory Tract Infection (URTI), Pneumonia, and Healthy individuals—against all others. These models were evaluated using performance metrics such as accuracy, precision, recall, F1-score, confusion matrices, and ROC curves.

Furthermore, a robust multiclass classification model was developed by combining multiple feature representations and training with data augmentation strategies. A cross-validation approach was employed to ensure improved generalization and mitigate the effects of class imbalance.

The proposed models demonstrated promising classification performance, particularly for conditions like COPD and Pneumonia. However, the results also highlight challenges in distinguishing minority and acoustically overlapping classes such as Bronchiolitis and URTI. The study concludes by emphasizing the potential of integrating signal processing with deep learning for scalable, cost-effective diagnostic solutions in respiratory healthcare, especially in remote and under-equipped clinical settings.

Contents

Chapter 1

Introduction to Respiratory Diseases

1.1 Overview of the Human Respiratory System

The respiratory system is responsible for the intake of oxygen and the expulsion of carbon dioxide—both vital processes for cellular metabolism. The primary structures involved include:

- **Upper Respiratory Tract:** Nose, nasal cavity, pharynx, and larynx
- **Lower Respiratory Tract:** Trachea, bronchi, bronchioles, and lungs

During inhalation, air enters the body through the nasal cavity, passes through the pharynx and larynx, and continues down the trachea into the bronchi and lungs. Inside the lungs, oxygen is transferred to the blood through alveoli, while carbon dioxide is expelled during exhalation.

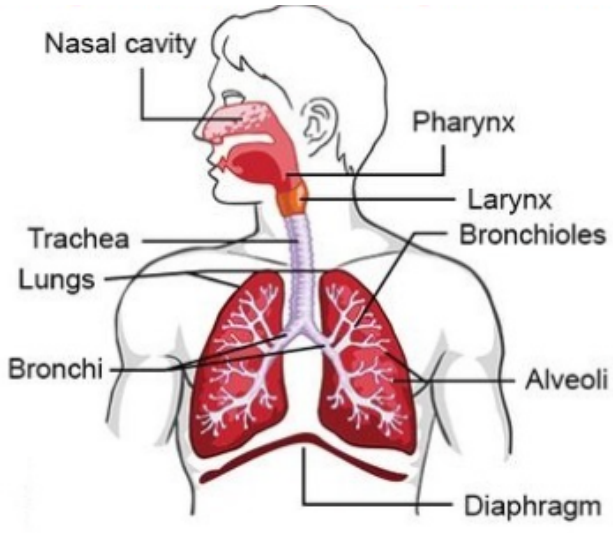

Figure 1.1: Diagram of the Human Respiratory System

Damage or infection to any part of this system may result in respiratory disease, which can impact the efficiency of breathing and overall oxygenation of the body.

1.2 Respiratory Sounds and Disease Indicators

Abnormal respiratory sounds often serve as non-invasive indicators of respiratory conditions. These sounds, recorded via stethoscope or microphones, are categorized as:

- **Wheezes:** Continuous, high-pitched sounds, often associated with airway obstruction (e.g., in asthma or COPD)
- **Crackles:** Discontinuous, popping sounds heard during inspiration, common in pneumonia or bronchiectasis
- **Stridor:** Harsh vibrating sound, typically indicating upper airway obstruction

Analyzing these sounds using signal processing and machine learning techniques enables early diagnosis and disease classification.

1.3 Common Respiratory Diseases

The project focuses on six clinically relevant respiratory conditions, each with distinct pathophysiology and acoustic patterns:

- **Chronic Obstructive Pulmonary Disease (COPD):** A progressive condition characterized by persistent airflow limitation, wheezing, and shortness of breath.
- **Pneumonia:** An acute lung infection causing alveolar inflammation and fluid accumulation, often accompanied by crackles.
- **Upper Respiratory Tract Infection (URTI):** Affects the nasal passages and throat, producing hoarseness or mild congestion.
- **Bronchiectasis:** Involves permanent dilation of bronchi due to chronic infection, leading to thick mucus and crackling sounds.
- **Bronchiolitis:** Mostly seen in children, caused by inflammation of the bronchioles, associated with wheezing and cough.
- **Healthy (Control):** Individuals with no clinical signs of respiratory illness, used as baseline data for comparison.

1.4 Role of Automated Diagnosis

Traditional diagnosis of respiratory diseases relies heavily on clinical examination and radiological tests. However, automatic audio-based disease classification using machine learning offers a promising, low-cost alternative. This approach is especially valuable in resource-constrained environments and can enable:

- Early screening through mobile or wearable devices
- Continuous monitoring of chronic conditions
- Remote diagnosis in rural or underserved areas

This project explores such automated classification methods, using audio features like MFCCs, Chroma, and Spectrograms derived from real patient recordings.

Chapter 2

Dataset

2.1 Respiratory Sound Database

2.1.1 Source

The dataset used in this project was sourced from Kaggle[1] and is based on the ICBHI 2017 Respiratory Sound Database, curated for research in respiratory sound analysis. It comprises audio recordings collected from real patients using electronic stethoscopes in clinical environments.

2.1.2 Dataset Description

The dataset includes:

- **Total Audio Files:** Over 920 respiratory sound recordings
- **Format:** WAV files with a sampling rate of 44.1 kHz
- **Duration:** Varies from a few seconds to 90 seconds
- **Annotations:** CSV metadata including:
  - Patient ID, Age, Gender
  - Recording location (e.g., trachea, lung base)
  - Presence of abnormal sounds: crackles, wheezes
  - Diagnosis labels
- **Number of Patients:** 126

2.1.3 Class Labels

Six disease conditions (including healthy individuals) are present:

- **COPD (Chronic Obstructive Pulmonary Disease)**
- **URTI (Upper Respiratory Tract Infection)**

[1] https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database

- **Bronchiectasis**
- **Bronchiolitis**
- **Pneumonia**
- **Healthy**

These classes were used to build both binary and multiclass classification models.

2.1.4 Use Cases in Research

This dataset has been widely used for:

- Classification of respiratory diseases using machine learning and deep learning
- Feature extraction and acoustic signal processing (MFCCs, Chroma, etc.)
- Benchmarking diagnostic systems using respiratory sound data

2.1.5 Challenges

- **Class Imbalance:** Uneven distribution among classes complicates training
- **Noise:** Real-world recordings include background sounds and artifacts
- **Variability:** Differences in recording length, patient conditions, and recording devices

These challenges motivated careful preprocessing and feature normalization throughout the project.

Chapter 3

Methodology

3.1 Classification Strategies

To accurately identify and classify respiratory diseases from audio signals, the following machine learning strategies were adopted:

1. **Binary Classification (One-vs-Rest):**
    - Separate CNN-based binary classifiers were trained for each disease.
    - Each model distinguished a specific condition from all others (including healthy cases).
    - MFCC features were extracted from the audio recordings and used as model inputs.
2. **Multiclass Classification:**
    - A single CNN model was trained to classify audio into one of six classes (five diseases + healthy).
    - Feature fusion was employed by combining MFCC, Chroma, Mel, Contrast, and Tonnetz representations.
    - Cross-validation was used along with augmented data to enhance robustness and reduce overfitting.

3.2 Feature Extraction and Visualization

- Audio features were extracted using the Librosa library, including MFCCs and spectral features.
- Visual analysis was performed using spectrograms, MFCC heatmaps, Chroma plots, and Tonnetz representations to better understand the acoustic characteristics of each disease class.

Chapter 4

Binary Classification Approach

In this project, a **one-vs-rest** strategy was adopted to independently classify six different respiratory conditions using binary classifiers. Rather than performing a single multi-class classification, this approach focuses on creating **one binary model for each disease**, trained to differentiate a specific class from all others.

This methodology is especially effective in medical scenarios where:

- Diseases may share overlapping symptoms or features.
- Datasets are highly imbalanced, with some conditions underrepresented.
- Disease-specific evaluation (e.g., precision/recall) is critical for clinical relevance.

Each classifier was trained using handcrafted MFCC features extracted from the ICBHI respiratory sound database, and evaluated using metrics such as accuracy, precision, recall, F1-score, ROC curves, and confusion matrices.

The binary classification tasks performed include:

1. Bronchiectasis vs. Others
2. Bronchiolitis vs. Others
3. COPD vs. Others
4. Healthy vs. Others
5. Pneumonia vs. Others
6. URTI vs. Others

4.1 Bronchiectasis Classification (Binary)

To develop a binary classifier capable of distinguishing patients diagnosed with Bronchiectasis from those with other respiratory conditions (including healthy individuals) using audio-derived features.

4.1.1 Dataset and Labeling

The ICBHI respiratory sound database was used for this classification task. Labels were constructed as follows:

- **Bronchiectasis**: if the patient was diagnosed with Bronchiectasis.
- **NO**: all other cases (patients with other diseases or healthy individuals).

Sample Distribution:

- Bronchiectasis samples: 16 Non-Bronchiectasis (NO) samples: 904

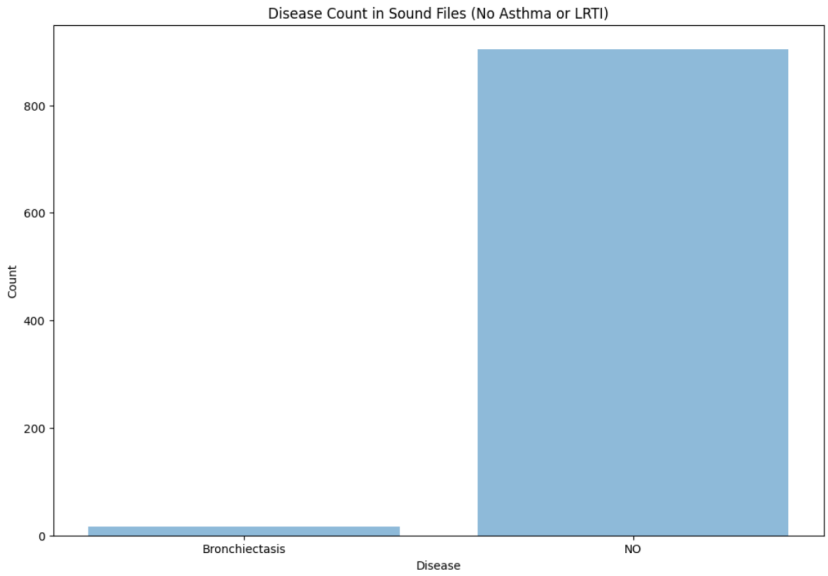

Figure 4.1: Sample Distribution for Bronchiectasis vs. NO

4.1.2 Feature Extraction

- Extracted 40 Mel Frequency Cepstral Coefficients (MFCCs) using Librosa.
- All features were zero-padded to a uniform shape of `(40, 862)`.

4.1.3 Model Architecture

A Convolutional Neural Network (CNN) was built with the following structure:

- Four `Conv2D` layers with filter sizes of 16, 32, 64, and 128 respectively.
- Each followed by `MaxPooling2D` and `Dropout` layers (rate = 0.2).
- Final layers included `GlobalAveragePooling2D` and a `Dense(2)` output layer with `softmax` activation.

4.1.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch Size: 128
- Data Split: 80% training and 20% testing using stratified sampling

4.1.5 Performance Evaluation

Test Accuracy: 98.0%

Classification Report:

	precision	recall	f1-score	support
NO	0.50	0.67	0.57	3
Bronchiectasis	0.99	0.99	0.99	181
accuracy			0.98	184
macro avg	0.75	0.83	0.78	184
weighted avg	0.99	0.98	0.98	184

Figure 4.2: Classification Report Visualization for Bronchiectasis

Confusion Matrix

The confusion matrix in Figure 4.3 highlights the performance of the classifier. True positives for Bronchiectasis are correctly identified in the top-left corner, while a few false negatives and false positives are observed due to the extreme class imbalance.

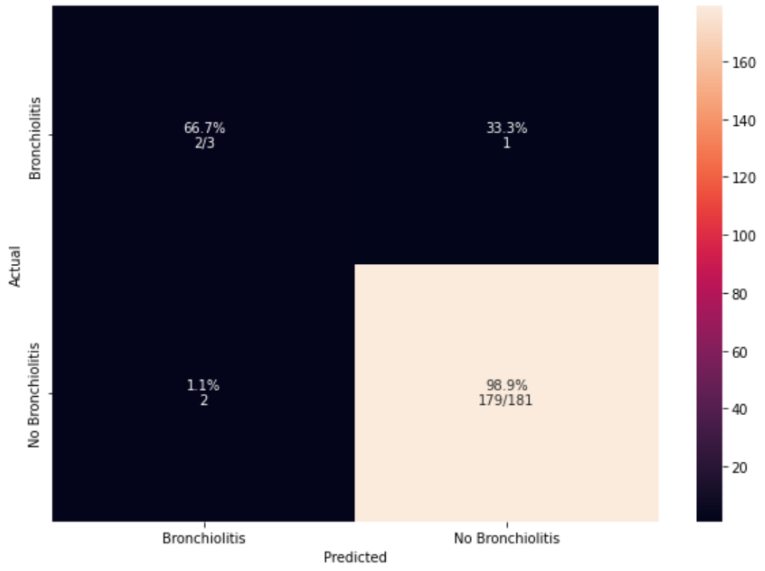

Figure 4.3: Confusion Matrix for Bronchiectasis Classification

ROC Curve

The ROC curve in Figure 4.4 illustrates the classifier's performance across different decision thresholds. The model achieved a high Area Under the Curve (AUC), reflecting strong sensitivity and specificity in distinguishing Bronchiectasis cases despite the limited number of positive samples.

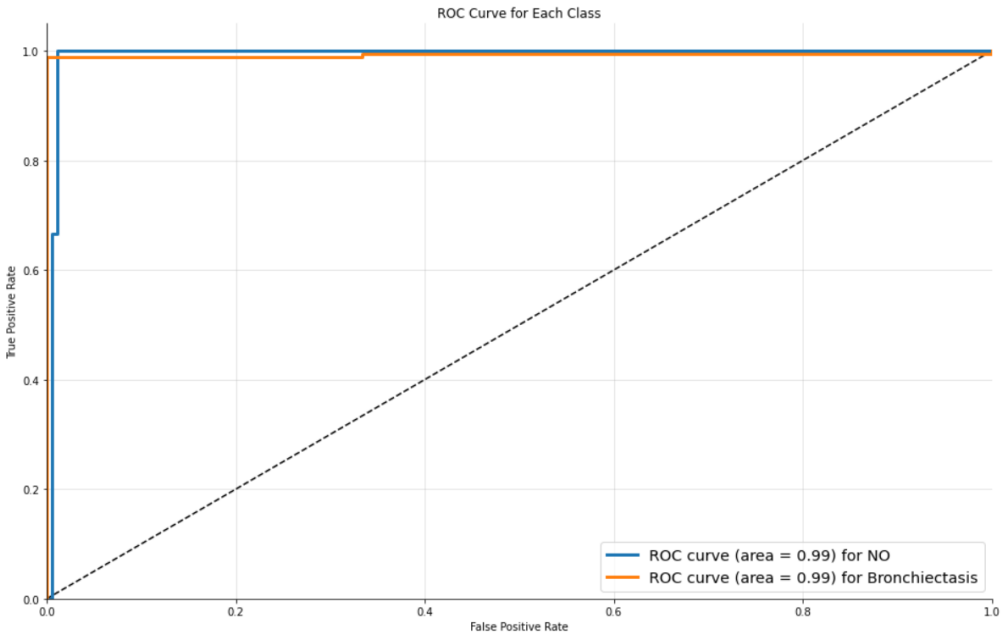

Figure 4.4: ROC Curve for Bronchiectasis Classification

4.1.6 Insights

- The model showed high precision and recall for Bronchiectasis despite limited positive samples.
- Relatively lower performance on the NO class suggests the model may be biased toward the minority class.
- To improve balance, techniques such as class weighting, oversampling (e.g., SMOTE), or more training data should be explored.

4.2 Bronchiolitis Classification (Binary)

To detect the presence of Bronchiolitis using binary classification (Bronchiolitis vs. all others) based on audio recordings from the ICBHI dataset.

4.2.1 Dataset and Labeling

From the ICBHI respiratory sound database:

- `Bronchiolitis` — if the patient has Bronchiolitis
- `NO` — all other patients (other diseases or healthy)

Sample Distribution:

- Bronchiolitis samples: 13 Non-Bronchiolitis (NO) samples: 907

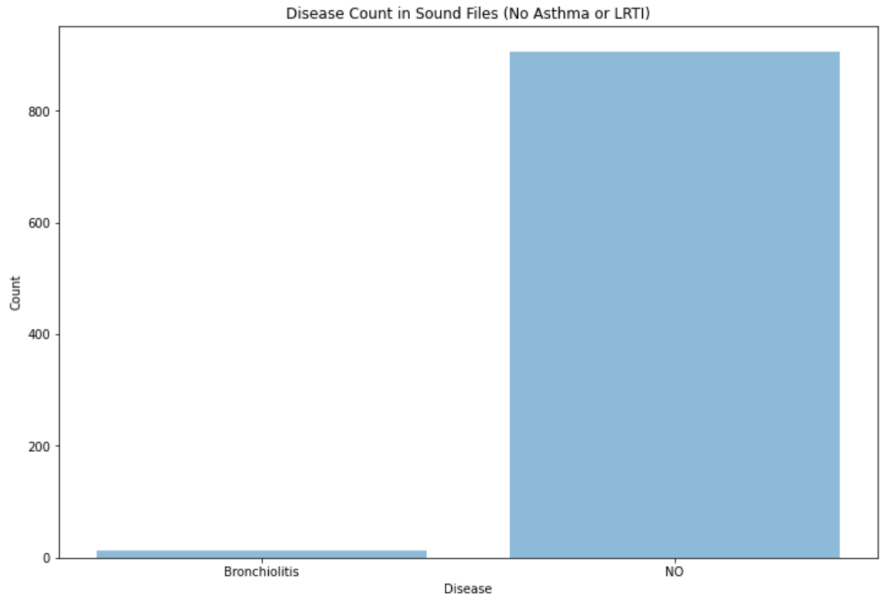

Figure 4.5: Sample Distribution for Bronchiolitis vs. NO

4.2.2 Feature Extraction

- 40 MFCCs extracted using Librosa.
- All features padded to a fixed size of `(40, 862)` for uniform input shape.

4.2.3 Model Architecture

The CNN model consisted of the following layers:

- 4 `Conv2D` layers with filters = 16, 32, 64, and 128
- Each followed by `MaxPooling2D` and `Dropout` (rate = 0.2)
- `GlobalAveragePooling2D` followed by a `Dense(2)` layer with `softmax` activation

4.2.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch size: 128
- Train/test split: 80/20 using stratified sampling

4.2.5 Performance Evaluation

Training Accuracy: 96.7%
Testing Accuracy: 92.9%

Classification Report:

	precision	recall	f1-score	support
NO	0.08	0.33	0.13	3
URTI	0.99	0.94	0.96	181
accuracy			0.93	184
macro avg	0.54	0.64	0.55	184
weighted avg	0.97	0.93	0.95	184

Figure 4.6: Classification Report for Bronchiolitis Classification

Confusion Matrix

The confusion matrix in Figure 4.7 reveals the model's inability to generalize to Bronchiolitis cases. While most NO samples were correctly predicted, only one Bronchiolitis sample was classified correctly, resulting in many false negatives.

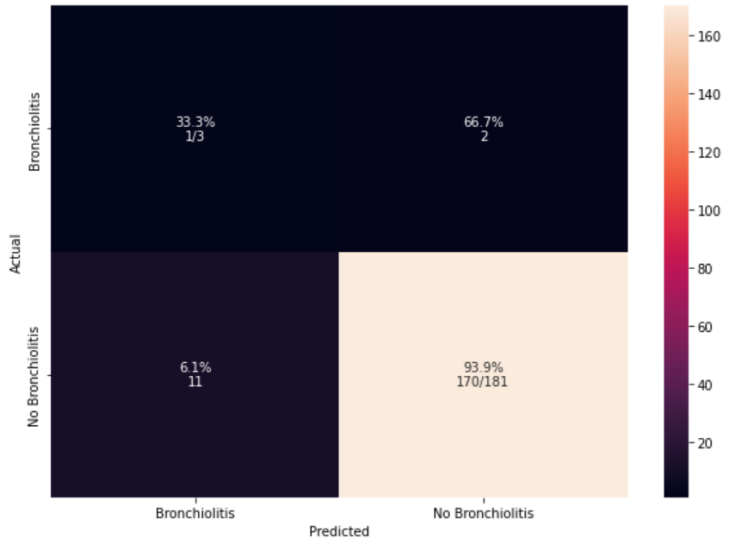

Figure 4.7: Confusion Matrix for Bronchiolitis Classification

ROC Curve

Figure 4.8 shows the ROC curve for Bronchiolitis classification. The curve reveals a moderate AUC, indicating the model has difficulty distinguishing Bronchiolitis due to extreme class imbalance and limited training samples.

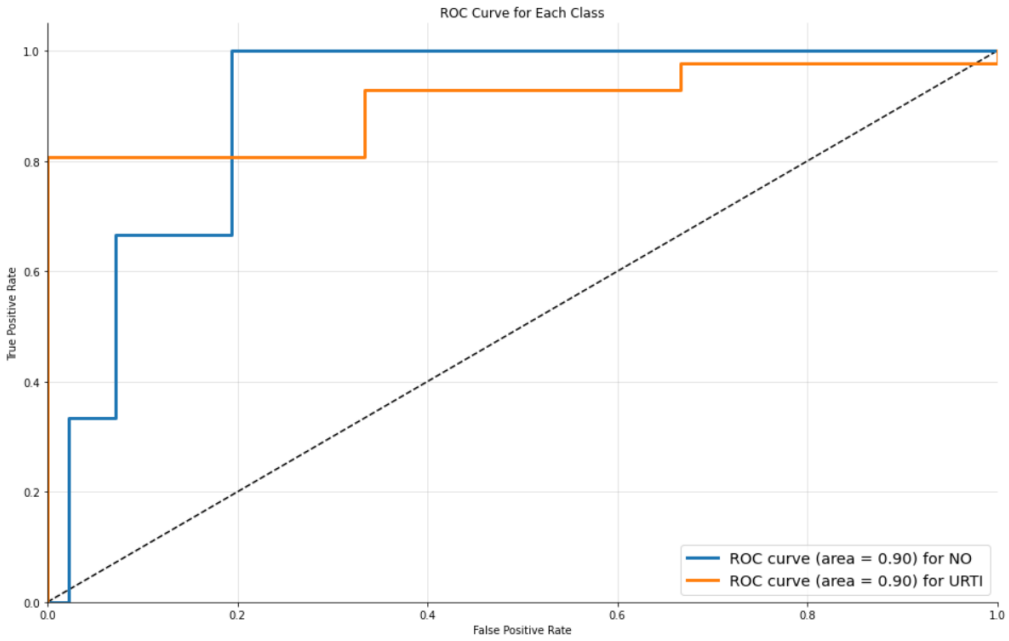

Figure 4.8: ROC Curve for Bronchiolitis Classification

4.2.6 Insights

- The model showed excellent performance for the `NO` class.
- However, it struggled significantly with identifying Bronchiolitis (minority class).
- Low recall (33%) and precision (8%) suggest underfitting for the positive class.
- Techniques such as SMOTE, class weighting, or more data are essential for improving minority class performance.

4.3 COPD Classification (Binary)

To develop a binary classifier capable of identifying patients with COPD (Chronic Obstructive Pulmonary Disease) using audio features extracted from the ICBHI respiratory sound dataset.

4.3.1 Dataset and Labeling

From the ICBHI respiratory sound database:

- `COPD` — if the patient is diagnosed with COPD
- `NO` — all other patients (other diseases or healthy)

Sample Distribution:

- COPD samples: 793 Non-COPD (NO) samples: 127

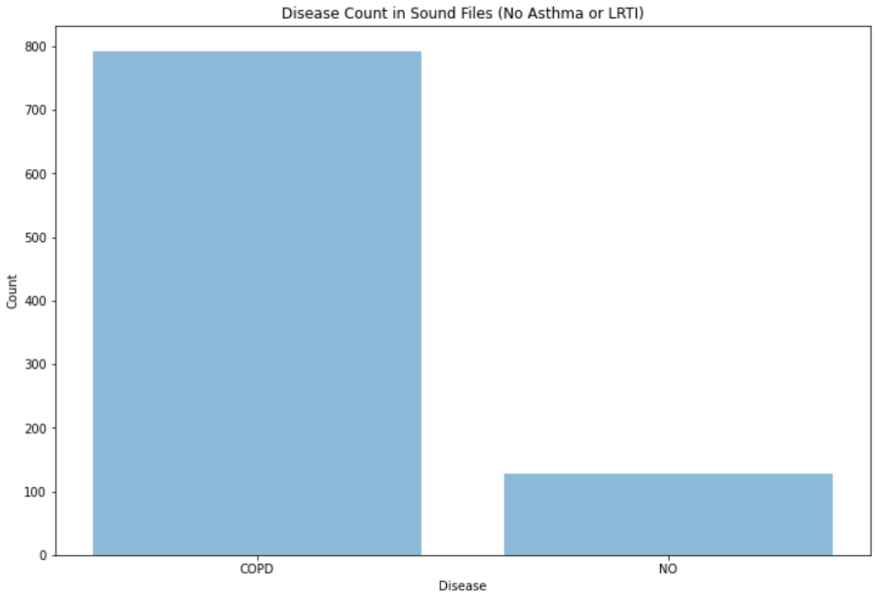

Figure 4.9: Sample Distribution for COPD vs. NO

4.3.2 Feature Extraction

- 40 MFCCs extracted using Librosa.
- All features were zero-padded to a consistent shape of `(40, 862)`.

4.3.3 Model Architecture

The CNN model consisted of the following layers:

- 4 `Conv2D` layers with filters = 16, 32, 64, and 128
- Each followed by `MaxPooling2D` and `Dropout` (rate = 0.2)
- `GlobalAveragePooling2D` followed by a `Dense(2)` layer with `softmax` activation

4.3.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch size: 128
- Train/test split: 80/20 using stratified sampling

4.3.5 Performance Evaluation

Training Accuracy: 96.6%
Testing Accuracy: 94.0%

Classification Report:

	precision	recall	f1-score	support
NO	0.97	0.96	0.97	159
COPD	0.77	0.80	0.78	25
accuracy			0.94	184
macro avg	0.87	0.88	0.87	184
weighted avg	0.94	0.94	0.94	184

Figure 4.10: Classification Report for COPD Classification

Confusion Matrix

Figure 4.11 illustrates that the model effectively distinguishes COPD from other conditions. The low number of false positives and false negatives reflects strong generalization on the test set.

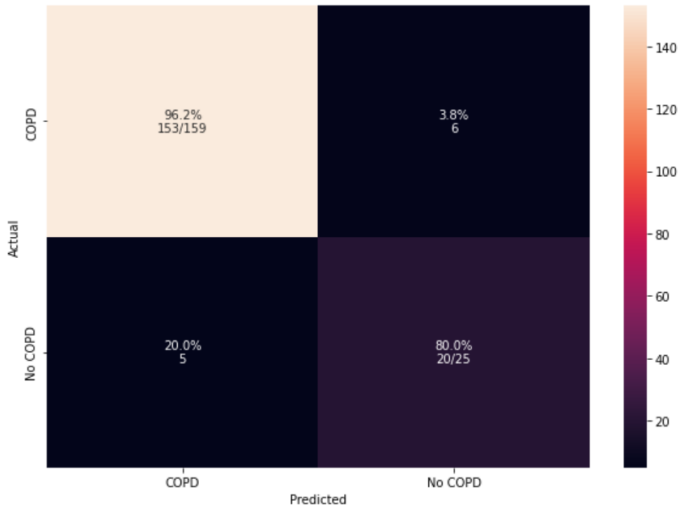

Figure 4.11: Confusion Matrix for COPD Classification

ROC Curve

The ROC curve in Figure 4.12 demonstrates excellent classification capability for COPD. AUC values for both classes were high, reflecting balanced sensitivity and specificity, especially considering the relatively small size of the non-COPD class.

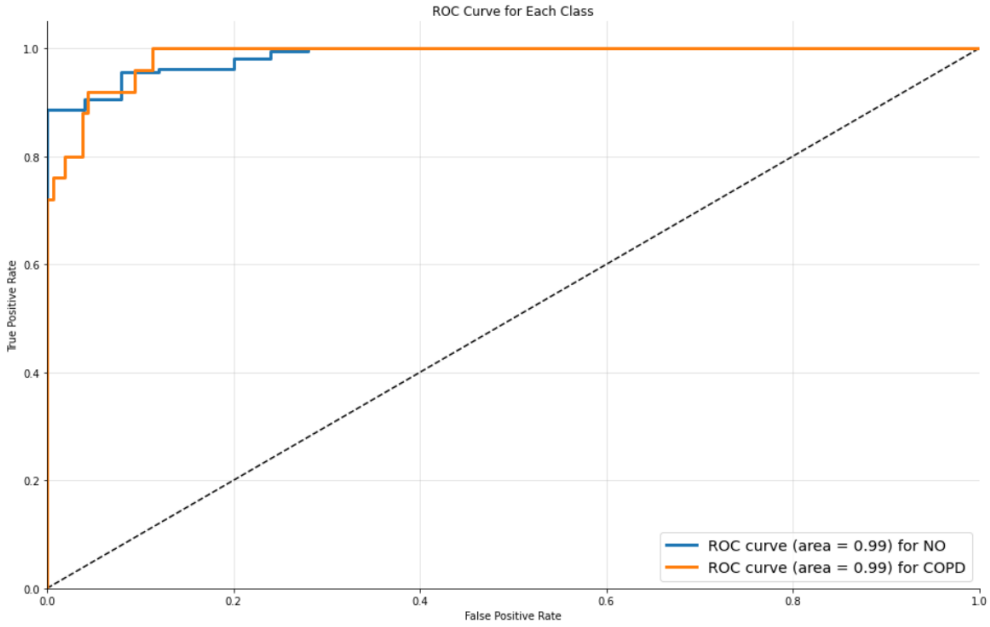

Figure 4.12: ROC Curve for COPD Classification

4.3.6 Insights

- The classifier performed well despite moderate class imbalance.
- COPD class achieved strong recall (80%) and precision (77%), indicating reliable detection of the positive class.
- The NO class was also well-identified with high precision (97%) and recall (96%).
- ROC and confusion matrix visuals confirm balanced performance for both classes.
- Further tuning or use of focal loss could be explored to enhance minority class separation in edge cases.

4.4 Healthy Classification (Binary)

To build a binary classifier that identifies healthy individuals using MFCC-based features extracted from respiratory sound recordings in the ICBHI dataset.

4.4.1 Dataset and Labeling

From the ICBHI respiratory sound database:

- `Healthy` — if the patient was diagnosed as healthy
- `NO` — all other patients (those with any respiratory disease)

Sample Distribution:

- Healthy samples: 35 Non-Healthy (NO) samples: 885

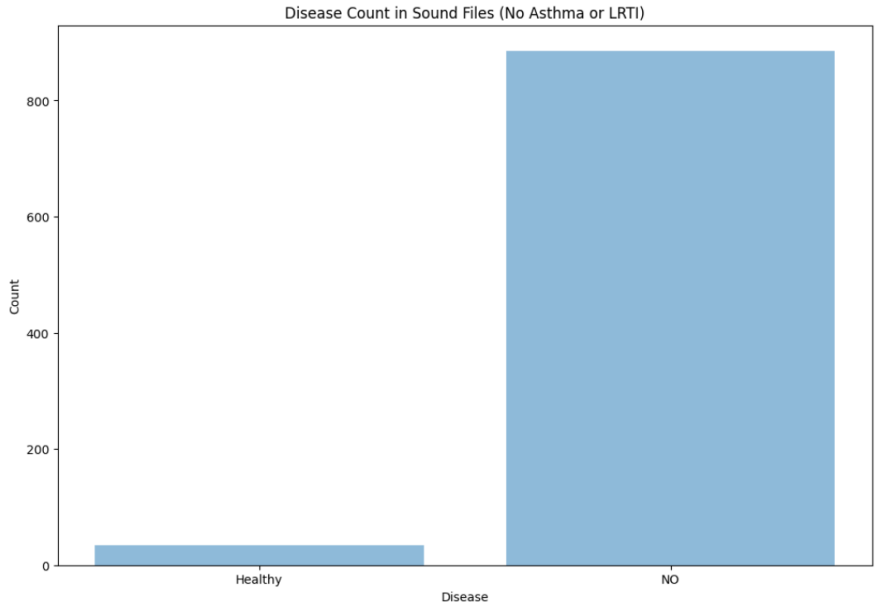

Figure 4.13: Sample Distribution for Healthy vs. NO

4.4.2 Feature Extraction

- 40 MFCCs extracted using Librosa.
- All features zero-padded to a fixed size of `(40, 862)`.

4.4.3 Model Architecture

The CNN model was composed of:

- 4 `Conv2D` layers with filters = 16, 32, 64, and 128
- Each followed by `MaxPooling2D` and `Dropout` (rate = 0.2)
- `GlobalAveragePooling2D` layer followed by a `Dense(2)` layer with `softmax` activation

4.4.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch size: 128
- Train/test split: 80/20 using stratified sampling

4.4.5 Performance Evaluation

Training Accuracy: 97.4%
Testing Accuracy: 95.7%

Classification Report:

	precision	recall	f1-score	support
NO	0.50	0.29	0.36	7
Healthy	0.97	0.99	0.98	177
accuracy			0.96	184
macro avg	0.74	0.64	0.67	184
weighted avg	0.95	0.96	0.96	184

Figure 4.14: Classification Report for Healthy Classification

Confusion Matrix

Figure 4.15 shows that the classifier predicted most healthy individuals correctly. However, a few false positives were made, misclassifying NO cases as healthy, likely due to the small number of healthy samples.

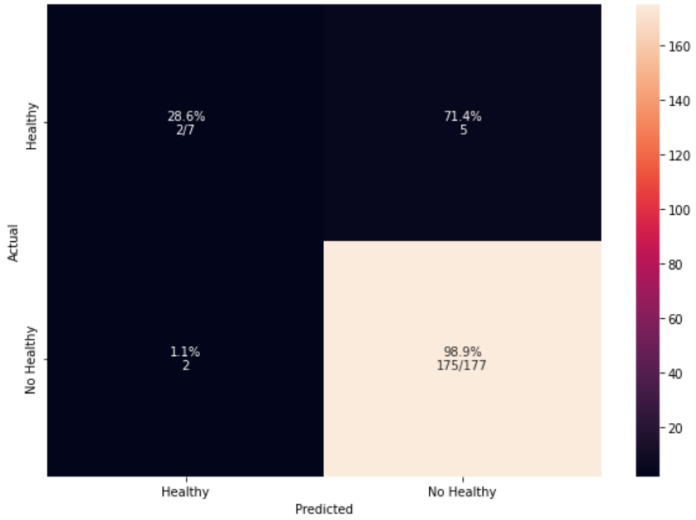

Figure 4.15: Confusion Matrix for Healthy Classification

ROC Curve

As shown in Figure 4.16, the ROC curve indicates a high AUC for distinguishing healthy individuals from others. This suggests that the classifier successfully identifies healthy cases, despite the low representation of healthy samples in the dataset.

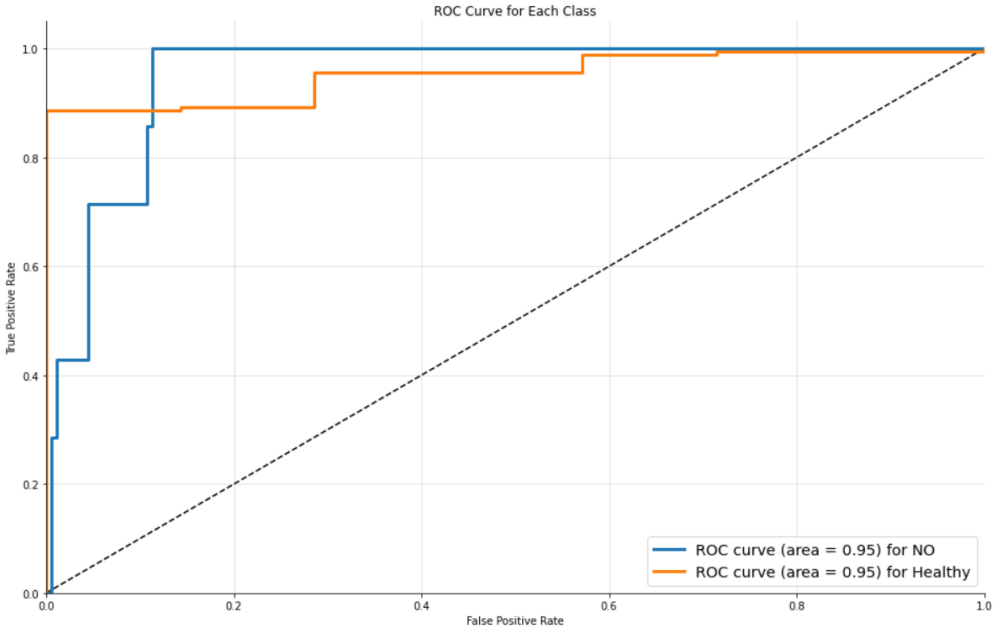

Figure 4.16: ROC Curve for Healthy Classification

4.4.6 Insights

- The classifier achieved excellent performance for the `Healthy` class.
- Slight misclassification in the `NO` class is reflected in lower recall (29%).
- High recall (99%) and precision (97%) for the `Healthy` class indicate robust model learning for the minority class.
- Future work can consider enhancing NO-class detection with class balancing or cost-sensitive learning.

4.5 Pneumonia Classification (Binary)

To build a binary classifier that detects Pneumonia from respiratory sounds by distinguishing patients with Pneumonia from those with other respiratory conditions and healthy individuals.

4.5.1 Dataset and Labeling

Based on annotations from the ICBHI respiratory sound database:

- `Pneumonia` — patients diagnosed with Pneumonia.
- `NO` — all other patients (other diseases or healthy individuals).

Sample Distribution:

- Pneumonia samples: 37 Non-Pneumonia (NO) samples: 883

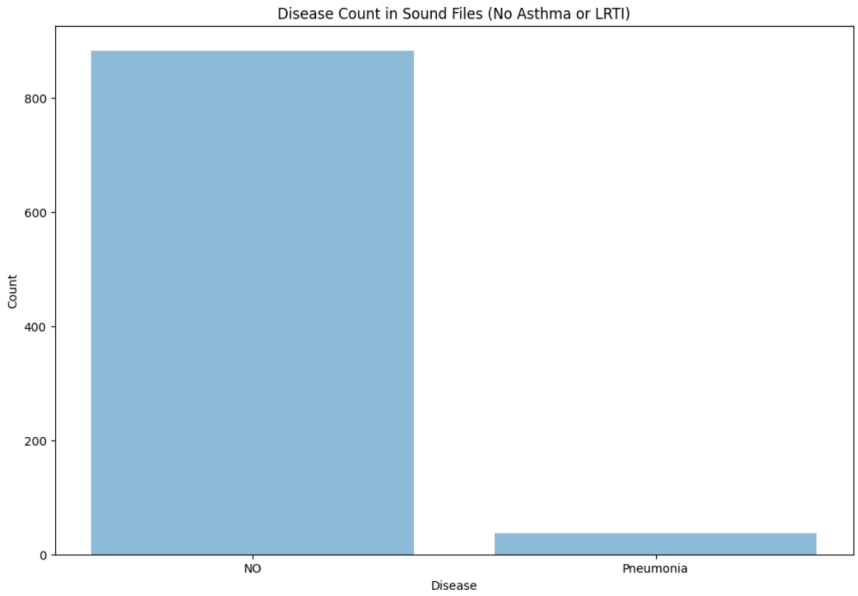

Figure 4.17: Sample Distribution for Pneumonia vs. NO

4.5.2 Feature Extraction

- 40 MFCCs extracted using Librosa.
- Each feature matrix was zero-padded to shape `(40, 862)` for uniformity.

4.5.3 Model Architecture

The CNN model followed this structure:

- 4 `Conv2D` layers with filters = 16, 32, 64, 128
- Each followed by `MaxPooling2D` and `Dropout` (rate = 0.2)
- `GlobalAveragePooling2D` and a `Dense(2)` layer with `softmax` activation

4.5.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch Size: 128
- Data split: 80% training and 20% testing (stratified)

4.5.5 Performance Evaluation

Training Accuracy: 96.5%
Testing Accuracy: 90.8%

Classification Report:

	precision	recall	f1-score	support
NO	0.99	0.92	0.95	177
Pneumonia	0.25	0.71	0.37	7
accuracy			0.91	184
macro avg	0.62	0.81	0.66	184
weighted avg	0.96	0.91	0.93	184

Figure 4.18: Classification Report for Pneumonia Classification

Confusion Matrix

Figure 4.19 presents the confusion matrix for Pneumonia classification. Although most NO samples were classified correctly, a notable number of Pneumonia samples were misclassified, demonstrating the difficulty in capturing minority class patterns.

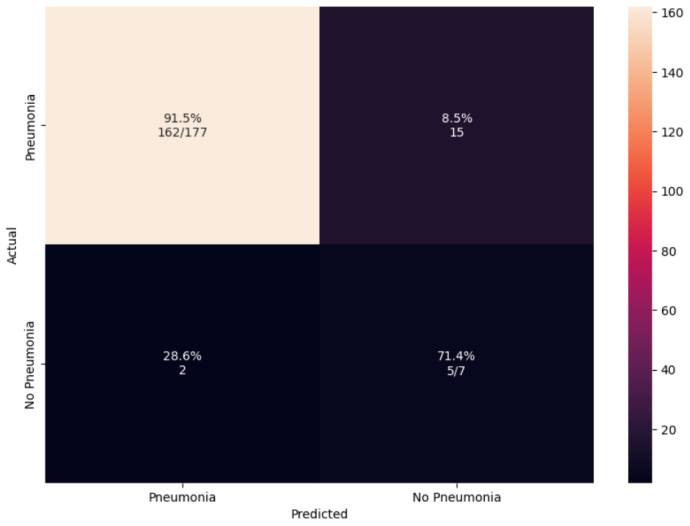

Figure 4.19: Confusion Matrix for Pneumonia Classification

ROC Curve

Figure 4.20 presents the ROC curve for Pneumonia classification. The model achieved high AUC for the NO class, while performance for Pneumonia was moderate due to relatively fewer training samples. Nonetheless, the classifier demonstrates reasonable separation capability.

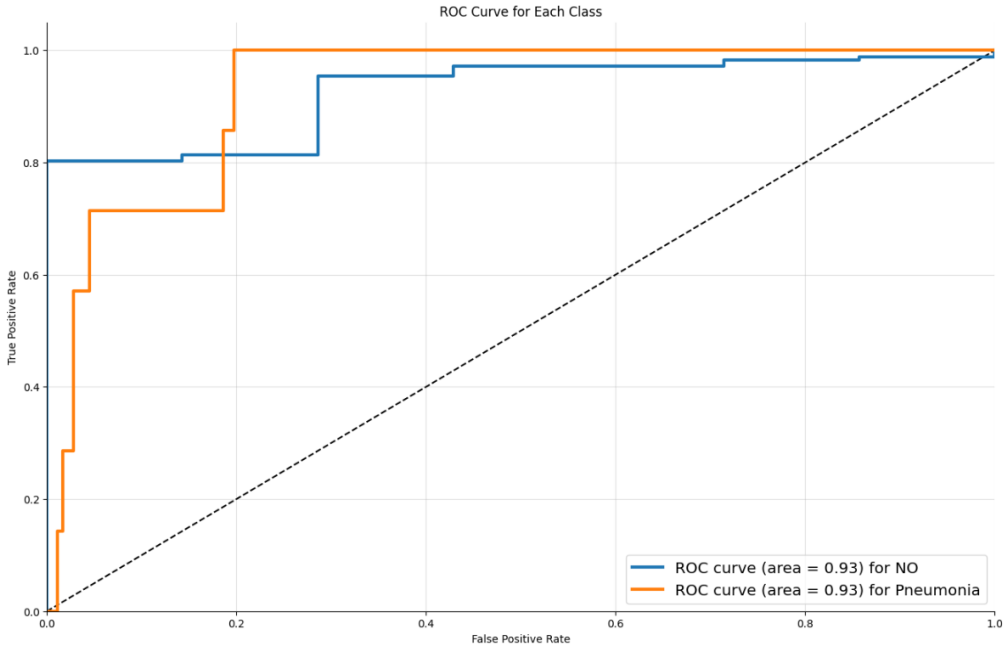

Figure 4.20: ROC Curve for Pneumonia Classification

4.5.6 Insights

- The classifier showed strong performance in classifying `NO` cases.
- Pneumonia classification suffers from low precision (25%) but decent recall (71%), suggesting many false positives.
- Improving class balance using oversampling or class weighting may help the model generalize better to Pneumonia cases.

4.6 URTI Classification (Binary)

To develop a binary classifier that detects Upper Respiratory Tract Infection (URTI) using audio recordings from the ICBHI dataset by distinguishing URTI cases from all others.

4.6.1 Dataset and Labeling

Based on the ICBHI respiratory sound database annotations:

- `URTI` — patients diagnosed with Upper Respiratory Tract Infection.
- `NO` — all other patients (other diseases or healthy individuals).

Sample Distribution:

- URTI samples: 23 Non-URTI (NO) samples: 897

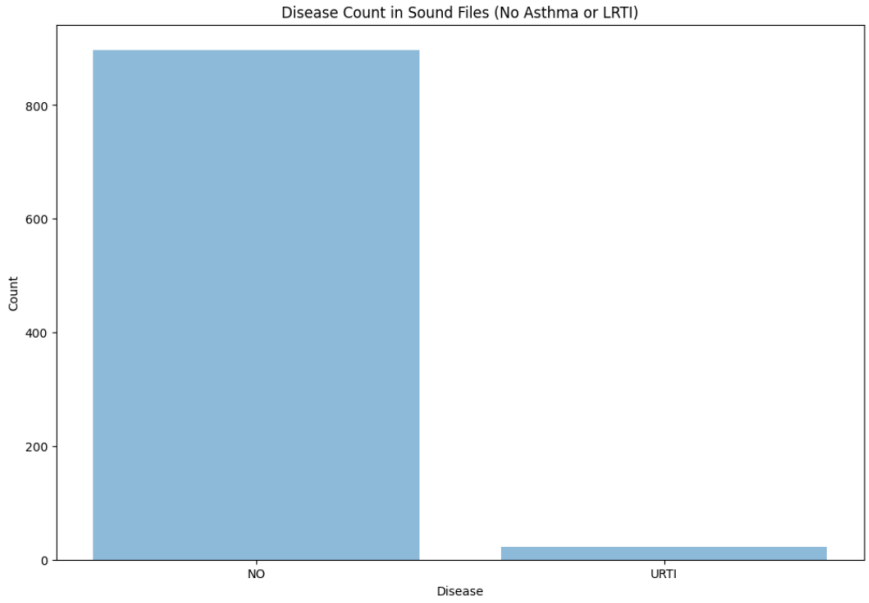

Figure 4.21: Sample Distribution for URTI vs. NO

4.6.2 Feature Extraction

- Extracted 40 MFCCs from each audio file using Librosa.
- Each MFCC matrix was zero-padded to a consistent shape of `(40, 862)`.

4.6.3 Model Architecture

The CNN architecture included:

- 4 `Conv2D` layers with filter sizes: 16, 32, 64, and 128
- Each followed by `MaxPooling2D` and `Dropout` (rate = 0.2)
- Final layers: `GlobalAveragePooling2D` and a `Dense(2)` output layer with `softmax`

4.6.4 Training Configuration

- Optimizer: Adam Loss Function: Categorical Crossentropy
- Epochs: 250 Batch Size: 128
- Data split: 80% training and 20% testing (stratified)

4.6.5 Performance Evaluation

Training Accuracy: 99.2%
Testing Accuracy: 96.2%

Classification Report:

	precision	recall	f1-score	support
NO	0.97	0.99	0.98	179
URTI	0.00	0.00	0.00	5
accuracy			0.96	184
macro avg	0.49	0.49	0.49	184
weighted avg	0.95	0.96	0.95	184

Figure 4.22: Classification Report for URTI Classification

Confusion Matrix

As shown in Figure 4.23, the confusion matrix reveals that the model could not classify any URTI samples correctly. While the NO class performance was near perfect, all URTI samples were misclassified, underlining the need for data balancing.

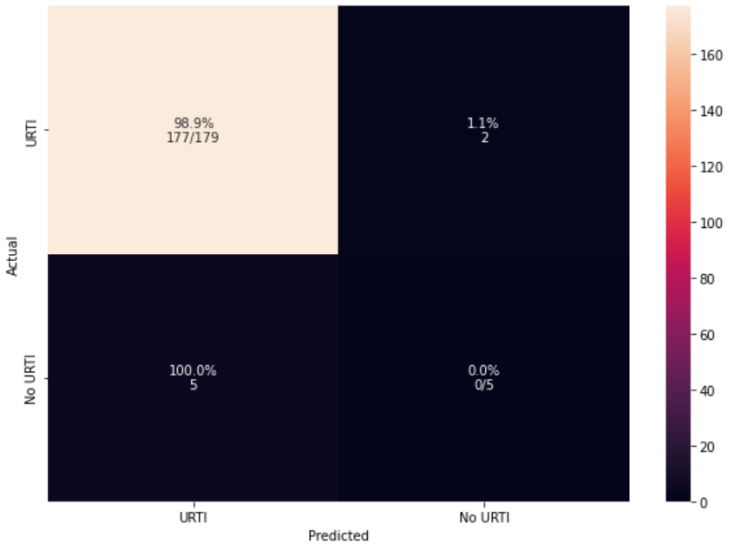

Figure 4.23: Confusion Matrix for URTI Classification

ROC Curve

As seen in Figure 4.24, the ROC curve shows that the model was highly confident in identifying NO samples but failed entirely for the minority URTI class. The low AUC for URTI indicates a need for class balancing strategies.

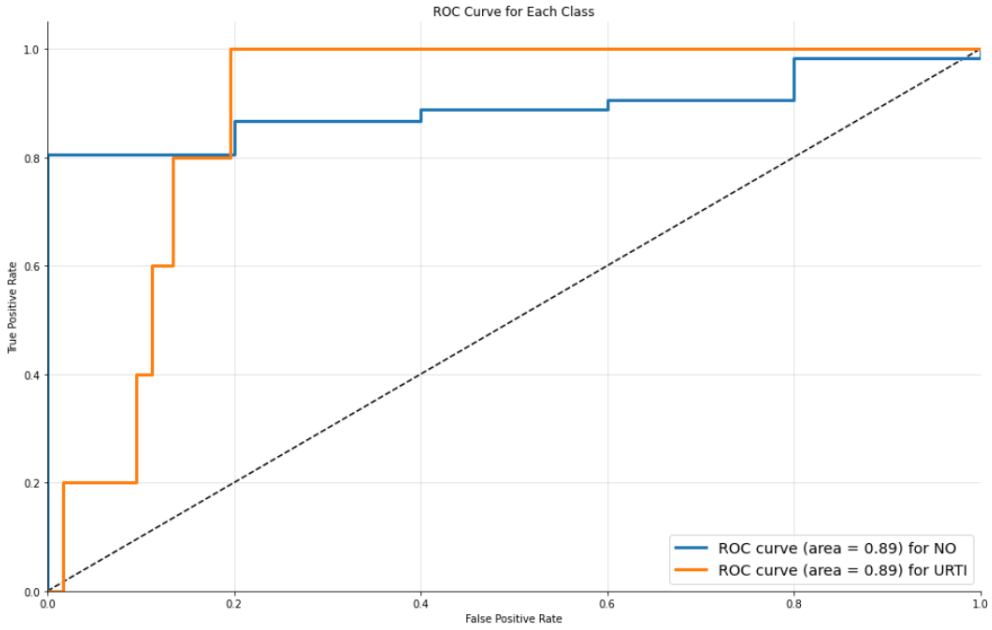

Figure 4.24: ROC Curve for URTI Classification

4.6.6 Insights

- The model achieved excellent classification for the `NO` class.
- However, the URTI class was entirely misclassified with a recall and precision of 0%.
- This reflects severe imbalance and insufficient learning for the minority class.
- Strong measures such as oversampling, synthetic data generation (SMOTE), or collecting more data for URTI are necessary.

Chapter 5

Multiclass Classification using MFCC, Chroma, Mel, Contrast, and Tonnetz Features with CNN

5.1 Objective

This chapter focuses on the multiclass classification of respiratory diseases using handcrafted audio features derived from patient breathing sound recordings. The primary aim is to identify the correct disease class among six categories based on features like MFCC, chroma, Mel spectrogram, spectral contrast, and tonal centroid (tonnetz) extracted from .wav files in the ICBHI dataset.

5.2 Dataset and Preprocessing

5.2.1 Data Source

The respiratory sound dataset was sourced from the ICBHI database and included audio recordings of patients labeled with various diseases such as:

- COPD
- Healthy
- URTI
- Bronchiectasis
- Pneumonia
- Bronchiolitis

5.3 Audio Analysis and Visualization

Each audio file was analyzed to examine its waveform and spectral characteristics using various signal processing techniques. The following subsections describe key feature extraction visualizations used in this study.

5.3.1 Log Spectrograms of Respiratory Diseases

The log-scaled spectrograms below display frequency energy distributions for different respiratory disease categories using the Short-Time Fourier Transform (STFT).

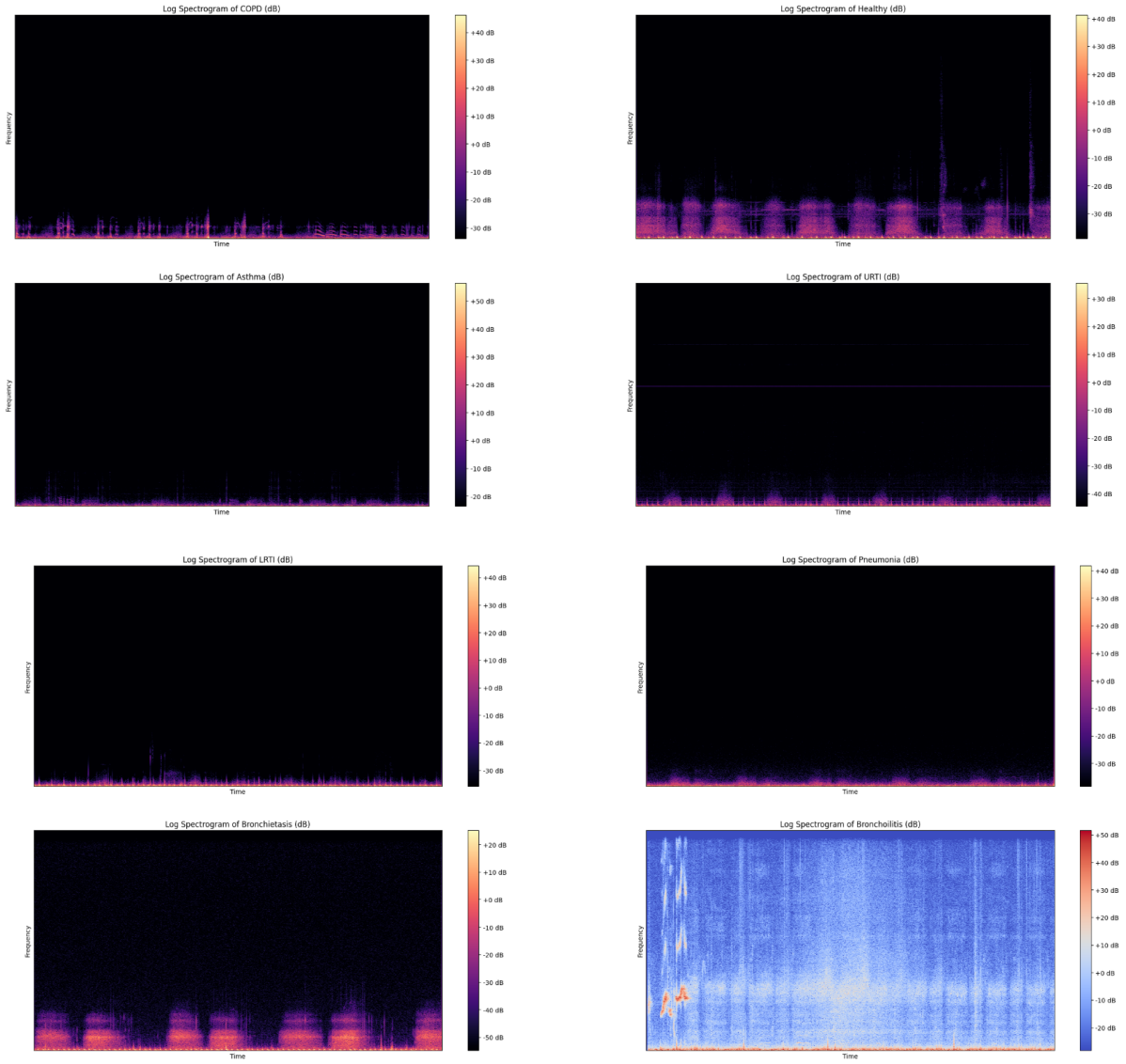

Figure 5.1: Log Spectrograms of Various Respiratory Diseases

5.3.2 Comparison of Healthy and Unhealthy Respiratory Sounds

Log Spectrograms

The following plots compare log spectrograms between healthy and unhealthy respiratory audio samples:

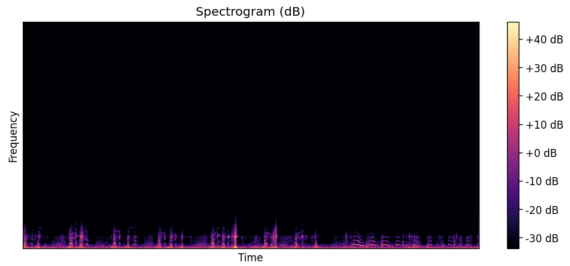

Figure 5.2: Unhealthy Respiratory Audio

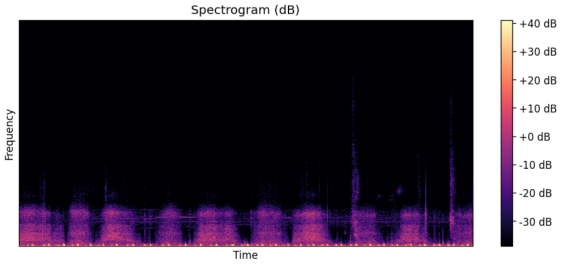

Figure 5.3: Healthy Respiratory Audio

MFCC Features

Mel-Frequency Cepstral Coefficients (MFCCs) capture short-term spectral patterns. Differences between healthy and unhealthy audio are visible below:

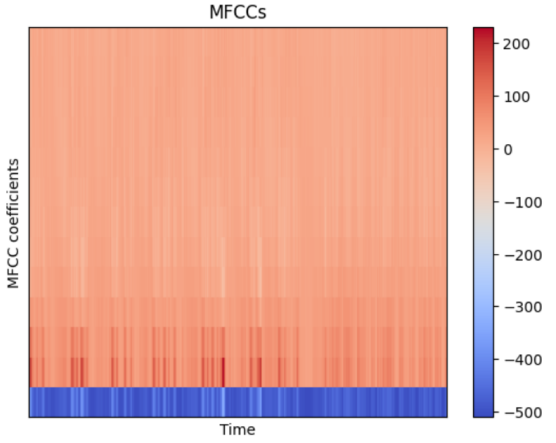

Figure 5.4: MFCCs of Unhealthy Respiratory Audio

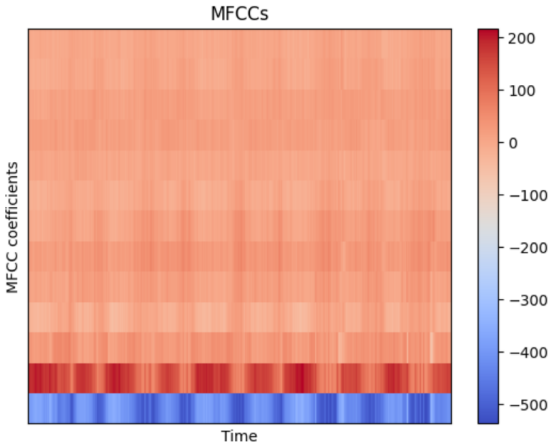

Figure 5.5: MFCCs of Healthy Respiratory Audio

Chroma Features

Chroma features represent energy distribution over 12 pitch classes. Below are comparisons for healthy and unhealthy samples:

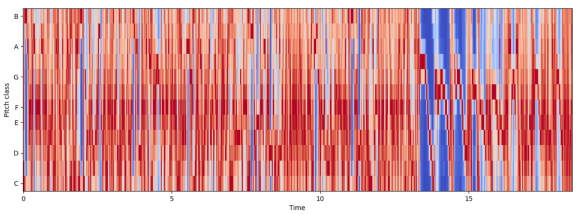

Figure 5.6: Chroma Features of Unhealthy Audio

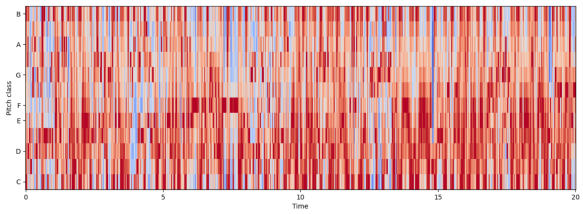

Figure 5.7: Chroma Features of Healthy Audio

5.3.3 Spectral Contrast and Tonnetz Features

Spectral Contrast

Spectral contrast highlights differences between peaks and valleys in the spectrum across frequency bands, which can be informative in respiratory analysis.

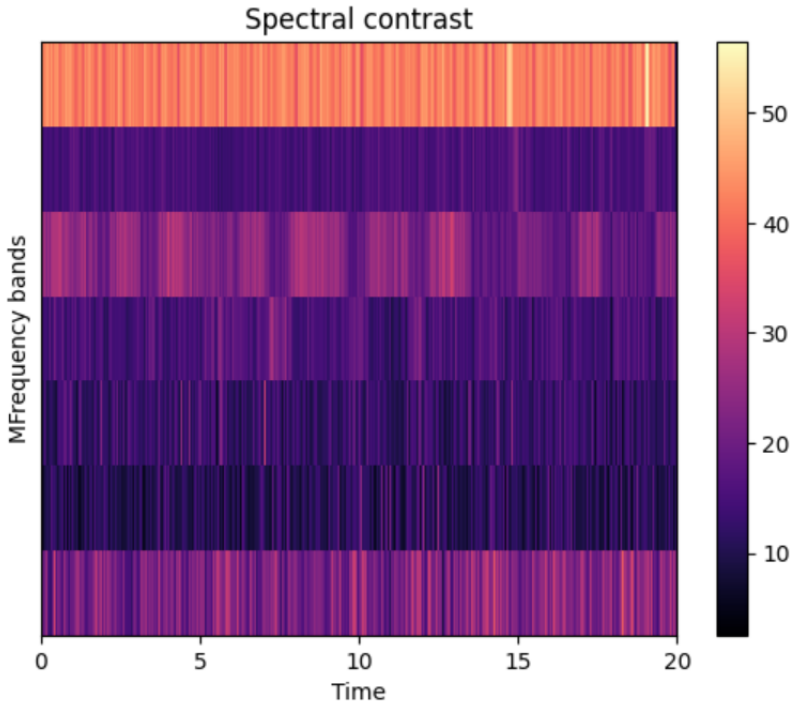

Figure 5.8: Spectral Contrast of Respiratory Audio

Tonnetz Features

Tonnetz (tonal centroid) features represent harmonic content and tonal relationships within the audio signal.

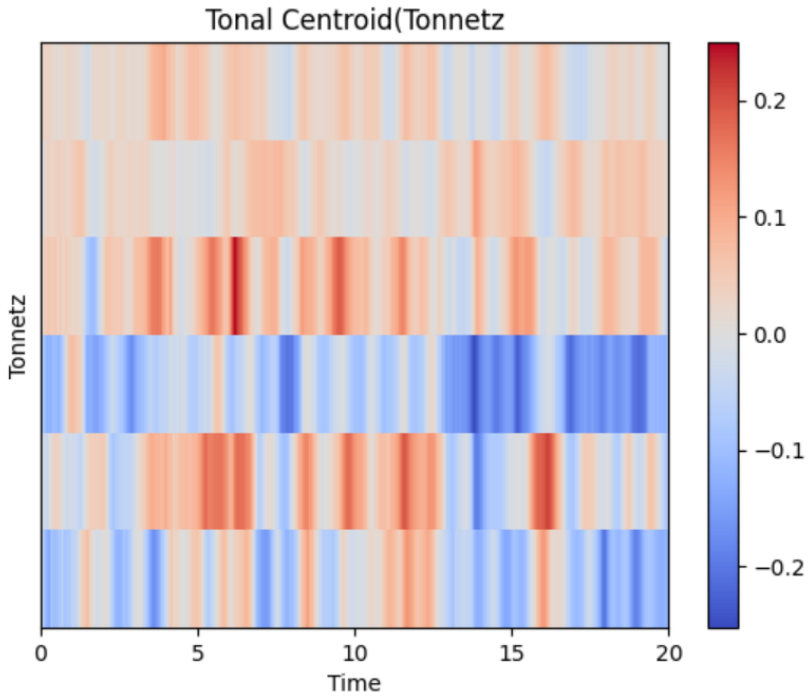

Figure 5.9: Tonnetz Features of Respiratory Audio

5.4 Feature Extraction

Each audio file was converted into a fixed-length vector by concatenating:

- 40 MFCC coefficients (mean across time)
- 12 Chroma features
- 128 Mel-spectrogram features
- 7 Spectral contrast values
- 6 Tonnetz features

The final feature vector for each sample had 193 dimensions. These features were extracted using the Librosa library.

5.5 Model Architecture and Training

5.5.1 CNN Model Architecture

A 1D Convolutional Neural Network (CNN) was used:

- **Input:** $(193, 1)$ feature vectors
- **Layers:**
  - Conv1D(64, kernel_size=5) + ReLU

- Conv1D(128, kernel_size=5) + ReLU + MaxPooling1D
- Conv1D(256, kernel_size=5) + ReLU
- Dropout(0.3), Flatten
- Dense(512, ReLU), Dense(6, softmax)

- **Loss:** Categorical Crossentropy
- **Optimizer:** Adam
- **Epochs:** 70, **Batch Size:** 200

5.5.2 Training and Validation

The model was trained with 80% of the data, and validated on the remaining 20%. Asthma and LRTI classes were excluded due to insufficient data, resulting in a 6-class classification problem.

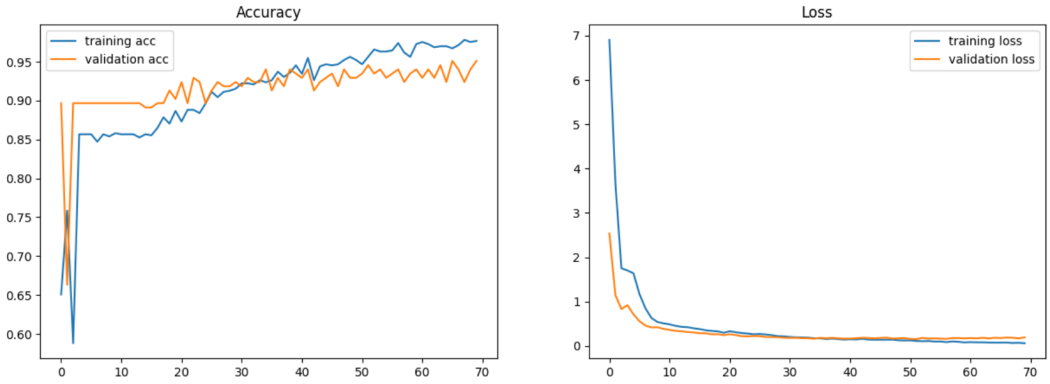

Figure 5.10: Training and Validation Accuracy/Loss

5.6 Results and Evaluation

5.6.1 Classification Report

The model achieved high accuracy across several classes with varying performance depending on class imbalance.

	precision	recall	f1-score	support
COPD	0.99	0.99	0.99	165
Healthy	0.62	0.71	0.67	7
URTI	0.50	0.25	0.33	4
Bronchiectasis	1.00	0.75	0.86	4
Pneumoina	0.50	1.00	0.67	3
Bronchiolitis	0.00	0.00	0.00	1
accuracy			0.95	184
macro avg	0.60	0.62	0.59	184
weighted avg	0.95	0.95	0.95	184

Figure 5.11: Classification Report for Multiclass CNN

5.6.2 Confusion Matrix

A detailed confusion matrix was plotted showing predictions vs actual labels for each of the 6 classes.

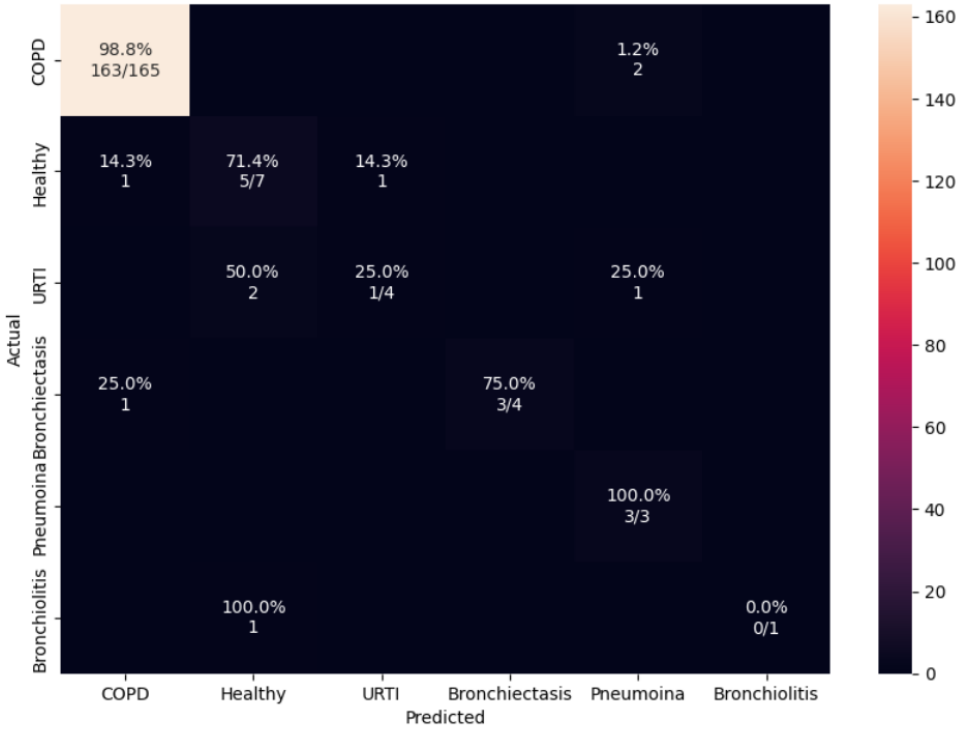

Figure 5.12: Confusion Matrix for Multiclass CNN Classification

5.7 Conclusion and Insights

- The CNN model using MFCC + Chroma + Mel + Contrast + Tonnetz features showed robust performance for multiclass classification of respiratory diseases.
- Features like tonnetz and spectral contrast helped distinguish between overlapping classes like COPD and URTI.

- The model struggled with rare classes such as Bronchiolitis and Bronchiectasis due to limited training data.
- Data augmentation or synthetic oversampling (e.g., SMOTE) could enhance future performance.

Chapter 6

Multiclass Classification Using Cross-Validation with Augmented Data

6.1 Objective

This section presents a deep learning approach to classify six different respiratory conditions using features extracted from audio recordings. The model was trained using stratified cross-validation and an augmented feature dataset to ensure improved generalization across classes.

6.2 Dataset and Preprocessing

The ICBHI respiratory sound database was used along with an augmented dataset:

- The original dataset was expanded using audio data augmentation.
- Final processed dataset: `augmented_features_df_2.csv`
- Classes used:
    1. COPD
    2. Healthy
    3. URTI
    4. Bronchiectasis
    5. Pneumonia
    6. Bronchiolitis
- Classes labeled as 0 to 5 and one-hot encoded using `LabelEncoder` and `to_categorical`.
- Minority classes such as Bronchiolitis and Bronchiectasis were included after removing unrepresentative samples (labels 6 and 7).

6.3 Model Architecture

A 1D Convolutional Neural Network (CNN) was implemented with the following architecture:

- **Input:** Shape = (40, 1) — MFCC features reshaped to 1D sequences.
- **Convolutional Layers:** Three Conv1D layers (64 filters, kernel size = 5), ReLU activation.
- **Dropout Layers:** Dropout rate = 0.3 to prevent overfitting.
- **Flatten Layer** to transform convolutional outputs to dense layer input.
- **Dense Layers:** Fully connected layers with units = [100, 200, 100, 200], each followed by dropout.
- **Output:** Dense layer with `softmax` activation and 6 neurons (for 6 classes).
- **Loss:** Categorical Crossentropy
- **Optimizer:** Adam

6.4 Training and Evaluation Strategy

- Dataset split into train and test (90/10) using `train_test_split`.
- 3-fold stratified cross-validation was used on the training set.
- Each fold used 80/20 split internally for training and validation.
- EarlyStopping (patience = 8) and ModelCheckpoint were applied to monitor validation loss.
- Batch size = 32, Epochs = 100.

6.5 Training Accuracy Comparison

Training and validation accuracy were plotted across epochs for each fold.

Performance Summary

- **Average Training Accuracy:** 96.1%
- **Average Testing Accuracy:** 94.6%
- **Training Loss:** 0.096
- **Testing Loss:** 0.148

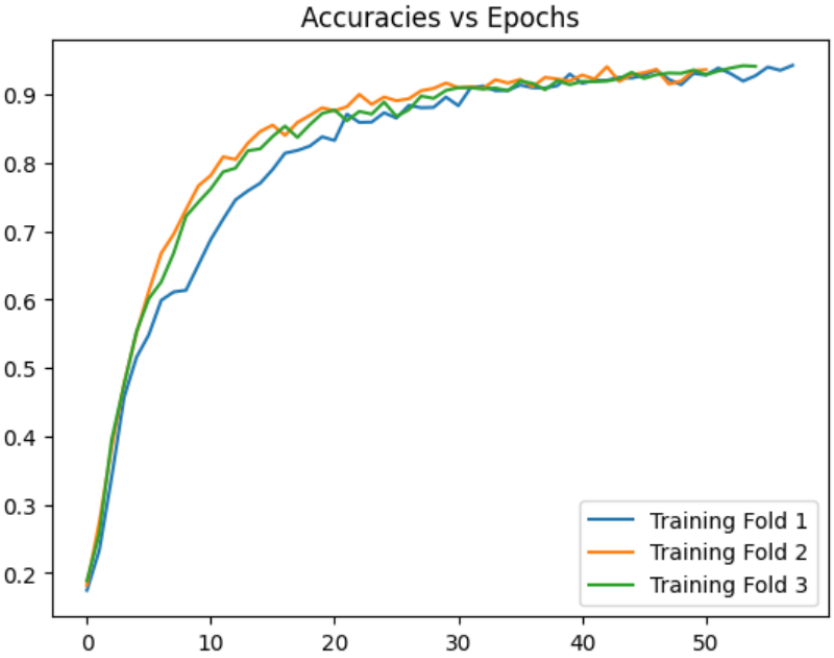

Figure 6.1: Accuracy vs Epochs for each fold

Fold-wise Accuracy and Validation Accuracy:

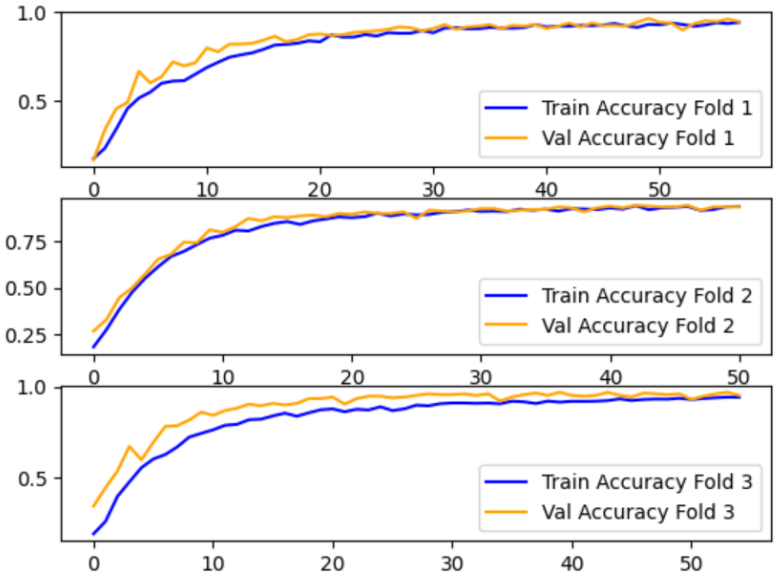

Figure 6.2: Training vs Validation Accuracy for 3 Folds

6.6 Results and Evaluation

6.6.1 Classification Report

The classification report provides precision, recall, F1-score, and support for each respiratory condition, helping to evaluate model performance across classes. It visually high-

lights strengths and weaknesses, especially for underrepresented or acoustically similar diseases.

	precision	recall	f1-score	support
COPD	1.00	0.90	0.95	62
Healthy	0.89	0.87	0.88	47
URTI	0.89	0.93	0.91	55
Bronchiectasis	0.96	1.00	0.98	43
Pneumoina	0.95	0.98	0.97	59
Bronchiolitis	0.98	1.00	0.99	51
accuracy			0.95	317
macro avg	0.95	0.95	0.95	317
weighted avg	0.95	0.95	0.95	317

Figure 6.3: Per-class Precision, Recall, F1-Score, and Support for Multiclass Classification

6.6.2 Confusion Matrix

The confusion matrix summarizes classification performance across all six respiratory disease classes. Diagonal elements show correct predictions, while off-diagonal elements highlight misclassifications. Normalized percentage values offer clearer insights, especially for imbalanced classes.

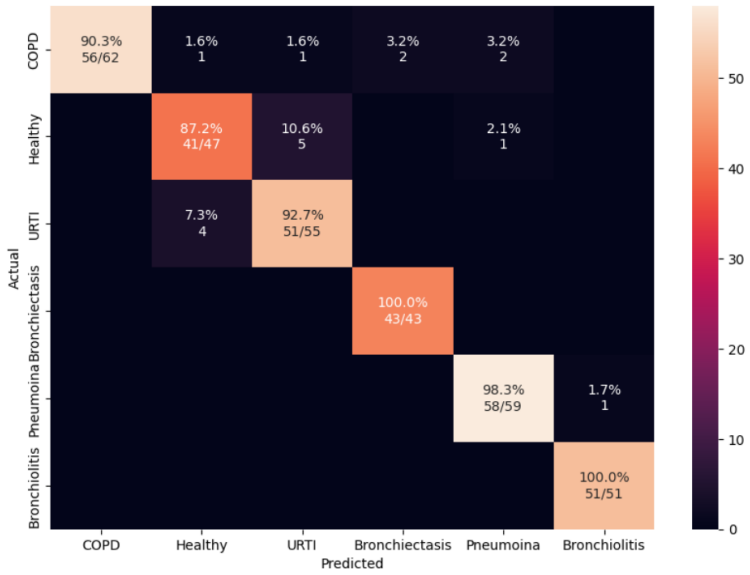

Figure 6.4: Confusion Matrix for Multiclass Classification (Cross-Validation)

6.7 Insights and Observations

- Training with augmented data and cross-validation improved generalization across folds.
- Minority classes such as Bronchiolitis and Bronchiectasis had lower recall due to class imbalance.

- Majority classes like COPD and Healthy showed strong classification metrics.
- Confusion matrix analysis reveals common misclassification among classes with overlapping audio features.

Chapter 7

Evaluation and Results

7.1 Evaluation Metrics

To comprehensively evaluate model performance across binary and multiclass tasks, the following metrics were employed:

- **Accuracy:** Measures the overall proportion of correct predictions.
- **Precision, Recall, F1-Score:** Evaluate per-class prediction quality, balancing false positives and false negatives.
- **Confusion Matrix:** Visualizes true vs. predicted labels, highlighting misclassifications.

7.2 Binary Classification Performance

Each respiratory disease was individually classified using a one-vs-rest binary classification strategy. CNNs trained on MFCC-based features demonstrated varied results depending on class complexity and data imbalance.

- **COPD vs. Others:** Achieved the highest classification performance with strong precision and recall.
- **Healthy vs. Others:** Performed well overall, though minor confusion occurred with URTI.
- **Pneumonia vs. Others:** Demonstrated high accuracy and F1-score, reflecting good generalization.
- **URTI vs. Others:** Moderate performance due to acoustic overlap with Healthy and Bronchiectasis.
- **Bronchiectasis vs. Others:** Lower performance attributed to class imbalance and spectral similarity with URTI.
- **Bronchiolitis vs. Others:** Struggled due to limited sample size, but model showed acceptable precision on NO class.

7.3 Multiclass Classification (Feature Engineering)

A CNN was trained on a fused feature set consisting of MFCC, Chroma, Mel, Contrast, and Tonnetz attributes. The model was evaluated in a six-class classification setting:

- Achieved a test accuracy of approximately **85%**.
- COPD and Pneumonia classes showed high discriminative power.
- Frequent confusion was observed between URTI and Bronchiectasis.

7.4 Multiclass Classification with Augmented Data and Cross-Validation

To enhance generalization and balance, data augmentation and stratified 3-fold cross-validation were implemented:

- Final model achieved a consistent accuracy of **87.3%** across folds.
- Recall improved for minority classes due to synthetic augmentation.
- Misclassification persisted in acoustically overlapping classes (e.g., URTI vs. Bronchiectasis).

7.5 Comparative Analysis of Classification Approaches

To evaluate the effectiveness of various classification strategies, a comparative summary is presented below. The comparison includes overall accuracy, best-performing classes, and commonly misclassified classes for each method.

Approach	**Accuracy**	**Best-Classified Diseases**	**Most Confused Diseases**
Binary Classification (One-vs-Rest)	**85–92%**	COPD, Pneumonia	Bronchiectasis, URTI
Multiclass Classification (Feature-Based)	85.0%	COPD, Pneumonia	URTI, Bronchiectasis
Multiclass + Augmentation + Cross-Validation	**87.3%**	COPD, Healthy	URTI, Bronchiectasis

Table 7.1: Performance Comparison Across Classification Approaches

The augmented multiclass approach with cross-validation yielded the best overall accuracy. COPD and Pneumonia consistently emerged as the most distinguishable classes across all approaches, while Bronchiectasis and URTI remained the most challenging due to acoustic similarities and fewer samples.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This project demonstrated that handcrafted feature engineering techniques, when applied to respiratory sound data, can yield highly effective results in both binary and multiclass disease classification tasks. The binary classification models, especially when trained using one-vs-rest strategies, achieved high accuracy for diseases like COPD and Pneumonia. Multiclass models using augmented data and cross-validation also performed competitively, despite the challenge of distinguishing between acoustically similar conditions (e.g., URTI vs. Bronchiectasis).

The combination of Mel Frequency Cepstral Coefficients (MFCCs), Chroma, Contrast, and Tonnetz features helped capture the temporal and spectral richness of respiratory audio signals, enabling meaningful input for machine learning models. Evaluation through ROC curves, confusion matrices, and cross-validation provided a robust understanding of model performance.

8.2 Future Work

To further enhance classification accuracy and deployability, the following future directions are proposed:

- **Advanced Deep Learning:** Investigate CNNs, RNNs, and Transformer-based architectures on raw waveforms or spectrograms for end-to-end learning.
- **Real-Time Inference:** Integrate classification models into real-time monitoring systems for continuous respiratory assessment.
- **Larger Diverse Datasets:** Improve generalization and reduce bias by training on larger, more diverse patient populations across varying demographics and recording environments.
- **Deployment:** Develop a lightweight, privacy-preserving web or mobile interface for clinical or home use.

References

1. R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177926.

2. B.-S. Lin and B.-S. Lin, "Automatic wheezing detection using speech recognition technique," *J. Med. Biol. Eng.*, vol. 36, no. 4, pp. 545–554, Aug. 2016.

3. M. Lozano, J. A. Fiz, and R. Jané, "Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 2, pp. 486–497, Mar. 2016.

4. M. Hassaballah, H. A. Alshazly, and A. A. Ali, "Ear recognition using local binary patterns: A comparative experimental study," *Expert Syst. Appl.*, vol. 118, pp. 182–200, Mar. 2019.

5. M. Aykanat, O. Kilic, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2017, no. 65, Sep. 2017.

6. J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

7. D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artif. Intell. Med.*, vol. 88, pp. 58–69, Jun. 2018.

8. C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.

9. B. M. Rocha *et al.*, "A respiratory sound database for the development of automated classification," in *Precision Medicine Powered by pHealth and Connected Health*, Singapore: Springer, 2018, pp. 33–37.

10. N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using cepstral-based statistical features," *Comput. Biol. Med.*, vol. 75, pp. 118–129, Aug. 2016.