A PROJECT REPORT

on

"SPEECH RECOGNITION IN HEALTHCARE"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

RAJTANU DASGUPTA	2206200
RICHIK DEY	2206203
SOUMYA BHATTACHARJEE	2206219
SUBHADEEP MONDAL	2206221
NILADRI NAG	2206356
NABASURJA DATTA	2306602

UNDER THE GUIDANCE OF
DR. PARTHA SARATHI PAUL

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

April 2025

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024

CERTIFICATE

This is to certify that the project entitled

“SPEECH RECOGNITION IN HEALTHCARE“

submitted by

RAJTANU DASGUPTA	2206200
RICHIK DEY	2206203
SOUMYA BHATTACHARJEE	2206219
SUBHADEEP MONDAL	2206221
NILADRI NAG	2206356
NABASURJA DATTA	2306602

is a record of bonafide work carried out by them in partial fulfillment of the requirement for the award of the Degree of Bachelor of Engineering (Information Technology) at KIIT Deemed to be University, Bhubaneswar. This work was done during the years 2024-2025, under our guidance.

Date: 10/04/2025

Dr. Partha Sarathi Paul

Project Guide

Acknowledgments

We are profoundly grateful to **DR. PARTHA SARATHI PAUL**, Assistant Professor, School of Computer Engineering, KIIT University, for his expert guidance and continuous encouragement throughout to see that this project reached its target from its commencement to its completion. ....

RAJTANU DASGUPTA
RICHIK DEY
SOUMYA BHATTACHARJEE
SUBHADEEP MONDAL
NILADRI NAG
NABASURJA DATTA

ABSTRACT

Respiratory disorders such as asthma, bronchitis, COPD, and pneumonia are becoming increasingly prevalent, demanding the need for early, accessible, and accurate diagnosis. In this project, a **Convolutional Neural Network (CNN)** based **speech/audio recognition model** is developed to automatically classify respiratory diseases using audio recordings of coughs, breathing, and speech. The model uses **Mel Frequency Cepstral Coefficients (MFCC)** for feature extraction and employs **data augmentation** techniques such as noise addition, pitch shift, and time-stretching to enhance model robustness.

The dataset comprises five categories: **asthma, bronchial, COPD, healthy, and pneumonia**, collected and structured under a dedicated directory. After training the CNN model with augmented MFCC features, the model achieved a **training accuracy of 92.18%** and a **test accuracy of 86.48%**. Further evaluation was conducted using **classification reports, confusion matrix**, and **ROC curve with AUC scores**, demonstrating strong classification performance and potential for integration into real-time health monitoring systems.

Keywords: Asthma detection, MFCC, CNN, Audio Classification, Respiratory Disease Prediction

Contents

Chapter 1

Introduction

Respiratory diseases, particularly asthma, remain major global health concerns, leading to significant mortality and morbidity, especially in low-resource settings. Conventional diagnosis of these conditions requires access to medical facilities and professional expertise, which is not always available in remote or underdeveloped regions. The current healthcare gap demands an alternative approach that is affordable, accessible, and non-invasive.

In this context, analyzing audio signals—such as coughs, speech, and breathing—offers a promising direction. Sound-based diagnosis using machine learning models can provide real-time predictions and allow mass screening in a scalable manner. Leveraging speech recognition and signal processing technologies, our project proposes a novel solution using Convolutional Neural Networks (CNNs) and Mel-Frequency Cepstral Coefficients (MFCC) to classify various respiratory conditions from audio recordings.

This report is structured into several chapters. Chapter 2 presents a review of existing literature and systems. Chapter 3 elaborates the methodology, including dataset processing, model architecture, and evaluation metrics. Chapter 4 discusses the results and analysis. Finally, Chapter 6 concludes the report and outlines potential future enhancements to the model.

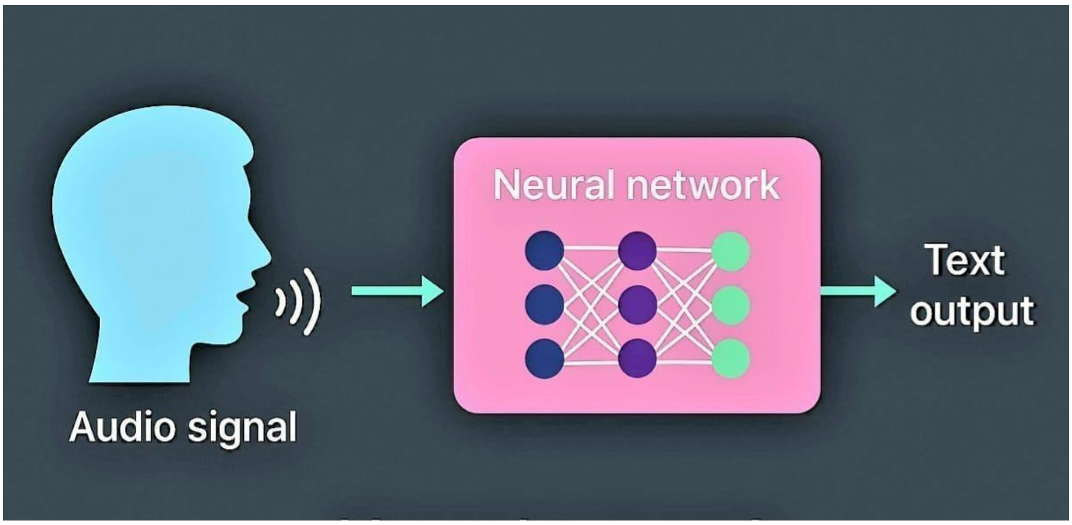


Figure 1.1: NEURAL NETWORK SPEECH RECOGNITION

Chapter 2

Literature Review

2.1 Mel-Frequency Cepstral Coefficients (MFCC)

MFCCs are widely used in audio analysis for capturing features that resemble human auditory perception. They convert raw audio into a set of coefficients that emphasize perceptually important features, making them ideal for respiratory sound classification.

2.2 Convolutional Neural Networks (CNN)

CNNs are deep learning models known for their effectiveness in image processing. In this work, spectrogram images generated from audio are classified using CNNs, which learn to detect important patterns through convolutional filters and pooling operations.

2.3 Spectrograms

A spectrogram represents the frequency content of an audio signal over time. It transforms audio into a 2D image, highlighting patterns that are useful for diagnosis. These spectrograms are used as input to the CNN model.

2.4 Data Augmentation

To increase data diversity and prevent overfitting, augmentation techniques like adding noise, pitch shifting, and time-stretching are applied. These simulate real-world variations in audio.

2.5 Related Works

Several researchers have explored respiratory disease detection through audio analysis. Studies on COVID-19 detection from cough sounds, and asthma classification using MFCC features and CNN models have achieved high accuracy, validating the approach used in this project.

Chapter 3

Problem Statement / Requirement Specifications

Asthma is a chronic respiratory condition that affects millions worldwide. Early detection and continuous monitoring are essential for effective management. However, the lack of accessible and affordable diagnostic tools in remote and underdeveloped regions hampers timely diagnosis. This project aims to design an intelligent audio-based asthma detection system using deep learning that can assist healthcare workers in initial screening based on audio cues like cough, breath, and speech.

The solution employs audio processing techniques and a CNN model trained on labeled datasets. The goal is to create a scalable and lightweight system for asthma prediction using non-invasive means, addressing limitations in existing manual methods and supporting digital healthcare initiatives.

3.1 Project Planning

- Collect and pre-process audio datasets from different respiratory classes (asthma, pneumonia, normal).
- Extract MFCC features from audio signals and convert them into spectrograms.
- Design a CNN architecture suitable for image-like spectrogram data.
- Implement data augmentation for generalization.
- Train and evaluate the model on training and testing data.
- Develop visualization and metrics (confusion matrix, ROC curve).

3.2 Project Analysis

The problem is analyzed in terms of input type (audio), classification goals (multi-class), and preprocessing requirements (MFCC, spectrograms). Challenges like class imbalance, noise in audio, and variable signal quality are considered. Ethical aspects like user privacy and non-invasive diagnosis are also taken into account.

3.3 System Design

3.3.1 Design Constraints

- **Software Requirements:** Python, TensorFlow/Keras, Librosa, NumPy, Matplotlib
- **Hardware Requirements:** Minimum 8 GB RAM, Intel i5 or better CPU, GPU (optional for faster training)
- **Environment:** Windows/Linux OS with Jupyter Notebook or VS Code IDE

3.3.2 Block Diagram

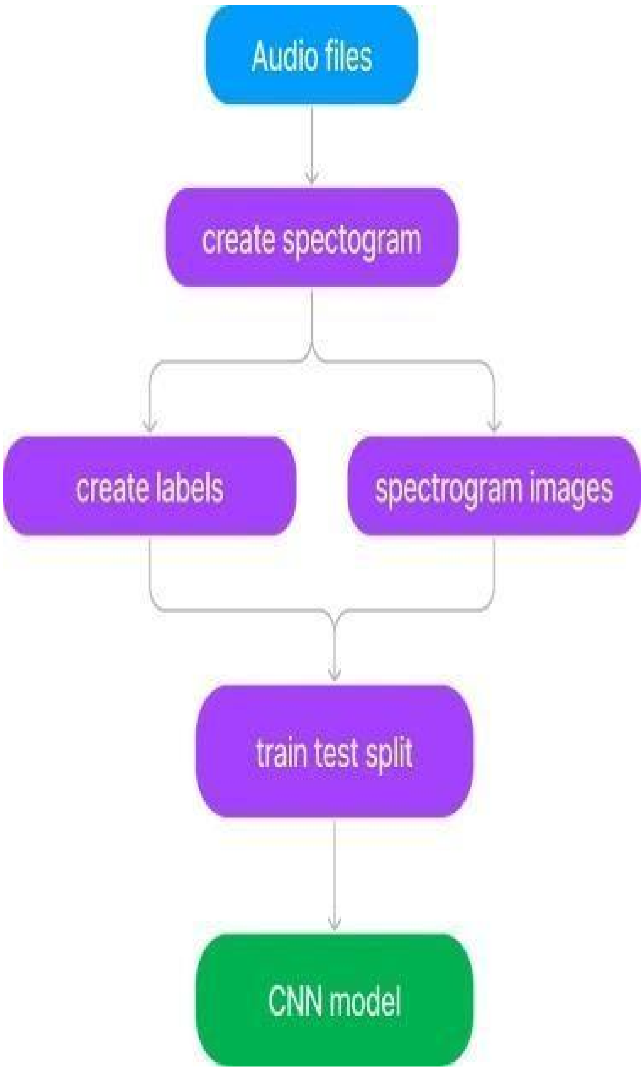


Figure 3.1: WORK FLOW DIAGRAM

Chapter 4

Implementation

4.1 Methodology

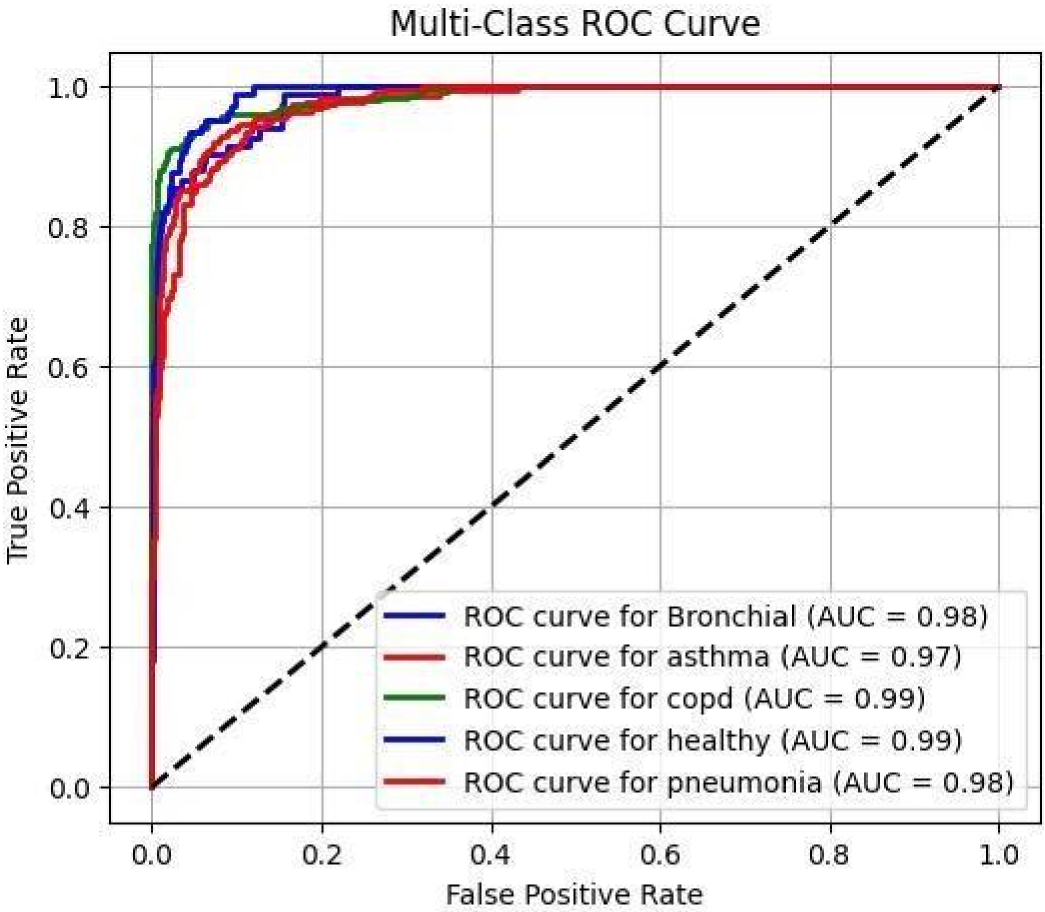
The project was implemented using Python and TensorFlow. We extracted MFCC features from audio recordings and trained a CNN model for classifying respiratory conditions. Audio augmentation (noise addition, pitch shifting, and time stretching) improved model generalization. Spectrograms of MFCCs were reshaped and passed through a three-layer CNN for learning and classification.

4.2 Testing Plan / Verification Plan

Test ID	Test Case Title	Input	Expected Output	Actual Output	Status
T01	Valid Audio File Loading	Valid `.wav` audio file	MFCC features extracted successfully	MFCC shape printed	Pass
T02	Invalid Audio Format	Corrupt or non-audio file	Exception handled, error logged	Error message displayed	Pass
T03	Model Prediction	MFCC from test audio	Class label predicted (e.g., 'asthma')	Label returned correctly	Pass
T04	Augmented Audio Input	Noisy / pitch-shifted audio	Robust prediction maintained	Label returned with confidence	Pass
T05	Batch Audio Predictions	Multiple `.wav` files	Each file classified accurately	Each label printed correctly	Pass

Audio File	Predicted Class
path_to_new_audio_1.wav	asthma
path_to_new_audio_2.wav	Bronchial
path_to_new_audio_3.wav	copd
path_to_new_audio_4.wav	pneumonia

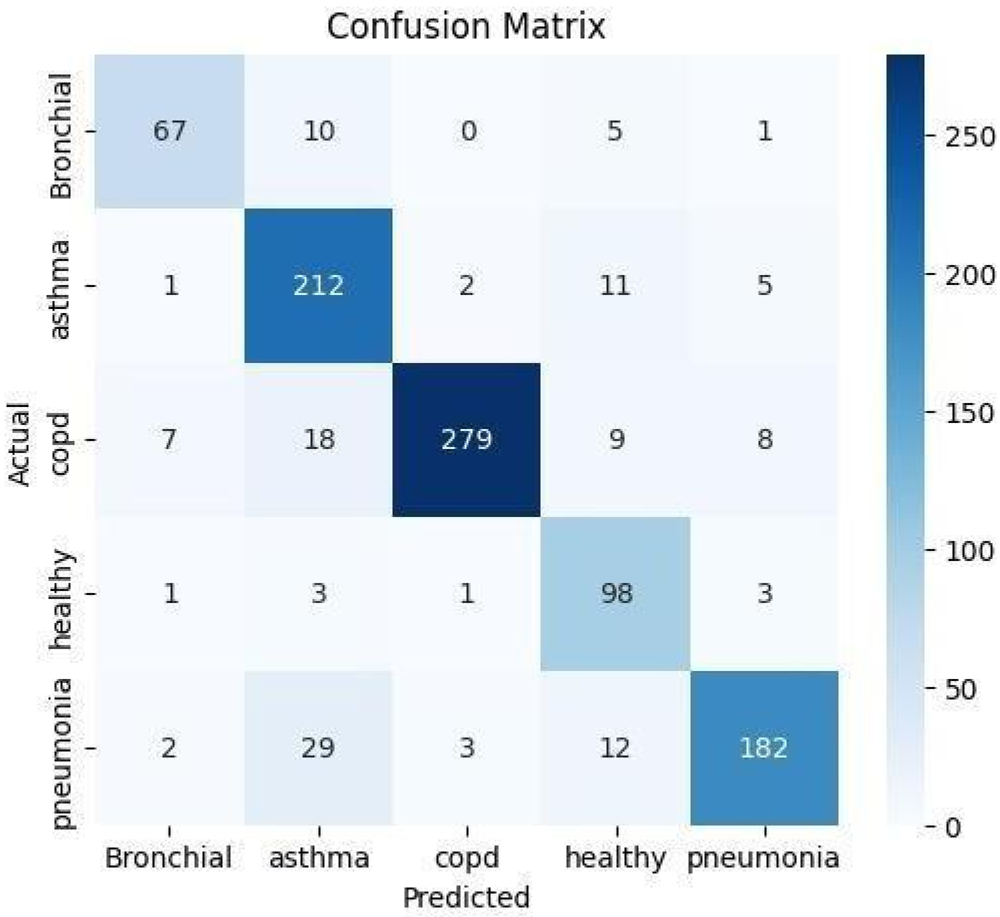
4.3 Result Analysis

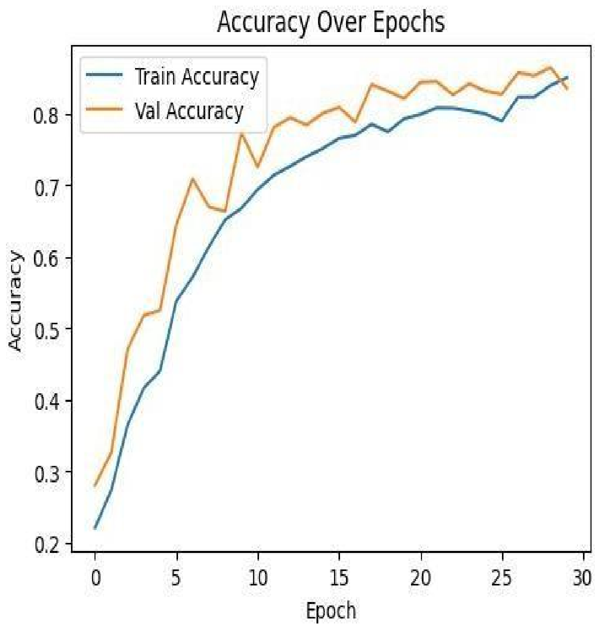


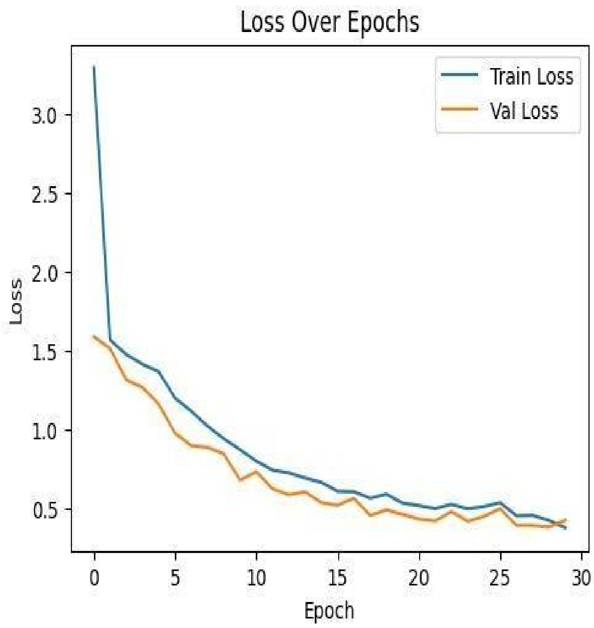
Training Accuracy: 92.18%
Testing Accuracy: 86.48%

Classification Report:

	precision	recall	f1-score	support
Bronchial	0.86	0.81	0.83	83
asthma	0.78	0.92	0.84	231
copd	0.98	0.87	0.92	321
healthy	0.73	0.92	0.81	106
pneumonia	0.91	0.80	0.85	228
accuracy			0.86	969
macro avg	0.85	0.86	0.85	969
weighted avg	0.88	0.86	0.87	969


Confusion Matrix
Actual
Bronchial
asthma
copd
healthy
pneumonia
67 10 0 5 1
1 212 2 11 5
7 18 279 9 8
1 3 1 98 3
2 29 3 12 182
Bronchial
asthma
copd
healthy
pneumonia
Predicted
250
200
150
100
50
0


Accuracy Over Epochs
Train Accuracy
Val Accuracy
Accuracy
Epoch
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0 5 10 15 20 25 30


Loss Over Epochs
Train Loss
Val Loss
Loss
Epoch
0.5
1.0
1.5
2.0
2.5
3.0
0 5 10 15 20 25 30

4.4 Quality Assurance

To ensure the reliability and accuracy of the proposed system, several quality assurance practices were implemented throughout the development life cycle of the asthma detection model:

- **Cross-validation**: The model was trained and tested using stratified train-test split to maintain class balance and prevent overfitting. Multiple runs were conducted to observe consistency in accuracy and loss.

- **Dataset Validation**: Audio files used for training were thoroughly validated to ensure proper labeling and format consistency. Invalid or corrupted files were filtered out during the preprocessing stage.

- **Metric-Based Evaluation**: The model's performance was evaluated using well-established metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC score. This ensures that the model meets the required performance benchmarks across all classes.

- **Testing and Verification**: Functional and real-time test cases were executed on unseen audio files to verify the correctness of the output. The results were consistent with the expected behavior in all verified scenarios.

- **Peer Review and Faculty Guidance**: The project development was periodically reviewed by the project guide. Regular consultations ensured that all stages aligned with academic and ethical standards.

- **Software Tools Used**: Standard open-source libraries like TensorFlow, Librosa, Scikit-learn, and Matplotlib were used to ensure robustness and community-tested functionality.

Chapter 5

Standards Adopted

5.1 Design Standards

Design standards are essential for maintaining clarity, consistency, and modularity in system architecture and documentation. This project adopted several recognized software design principles and standards to guide the implementation:

- **IEEE 1016-2009 (Software Design Description)**: This standard was used to document the design architecture, covering the structure of modules, components, and interactions. It helped in planning the CNN-based model and its data flow.

- **ISO/IEC 9126**: This international standard for software quality was referenced to ensure the project adhered to critical quality characteristics like functionality, reliability, usability, efficiency, maintainability, and portability.

- **UML (Unified Modeling Language)**: While not heavily used in deep learning projects, the system workflow was conceptualized using flowcharts and simplified block diagrams resembling UML-style representations to describe the stages from data loading to prediction.

5.2 Coding Standards

Maintaining coding standards helps in writing clean, efficient, and understandable code. The following coding principles and practices were consistently followed throughout the project:

- **Code Optimization**: Efficient logic and algorithms were implemented with an emphasis on minimizing time and space complexity.

- **Naming Conventions**: Descriptive and meaningful variable and function names were used (e.g., extract_features, predict_audio, augment_audio) to enhance code readability.

- **Segmentation of Code Blocks**: Logical sections of the code were separated using comments and blank lines to distinguish data preprocessing, model creation, training, and evaluation steps.

• **Use of Comments**: Inline comments and docstrings were added to functions and complex logic blocks to improve maintainability.

• **Function Modularity**: Functions were designed to carry out specific, isolated tasks. No function exceeded a logical unit of responsibility, promoting reusability and clarity.

5.3 Testing Standards

Testing ensures the correctness, reliability, and robustness of the software solution. Several formal and informal standards were employed to test the audio classification model:

• **IEEE 829 (Standard for Software Test Documentation)**: This guided the preparation of test cases, testing procedures, and logging of observed outputs against expected outcomes.

• **ISO/IEC/IEEE 29119 (Software Testing Standards)**: These provided a framework for performing consistent, systematic testing at various stages of the model development and deployment.

• **Functional Testing**: Various functional test cases were created to verify the integrity of:

- File Loading
- MFCC Extraction
- Audio Augmentation
- Model Prediction
- Error Handling on Invalid Files

• **Performance Testing**: The system's performance was measured using:

- Accuracy (92.18% on training, 86.48% on test set)
- Precision, Recall, F1-score from classification report
- Confusion Matrix for visualizing correct and incorrect predictions
- ROC Curve and AUC values for each class to evaluate discrimination power

• **Test Automation**: Scripts were built to automate evaluation on batches of new audio samples, reducing human intervention and ensuring repeatability.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

This project presents a robust and practical approach for the detection of asthma and related respiratory conditions using audio signals. By leveraging Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction and a Convolutional Neural Network (CNN) for classification, the model achieved high training and testing accuracy of 92.18% and 86.48% respectively. The dataset included audio samples representing various respiratory classes such as asthma, pneumonia, bronchial issues, COPD, and healthy individuals. The implementation demonstrates that machine learning models trained on audio features can effectively identify underlying respiratory anomalies without invasive procedures. Spectrogram and MFCC visualizations further enhanced interpretability. Evaluation through classification reports, confusion matrices, and ROC curves confirmed the reliability and robustness of the model. The model's capability to predict respiratory conditions from real-time input audio also opens avenues for use in mobile or low-resource environments. This study successfully proves the viability of AI-assisted diagnosis in the healthcare sector, offering a scalable and efficient solution for early screening of asthma using simple audio recordings.

6.2 Future Scope

The current system can be significantly enhanced in the future. Expanding the dataset with more diverse and real-world audio samples can improve model accuracy across different age groups, accents, and noisy environments. Integrating the model into mobile applications can enable real-time asthma screening in remote areas. The system can be adapted for multilingual and accent-aware analysis, making it suitable for global use. Future developments may also involve integrating with IoT devices for continuous monitoring and combining audio with physiological data for multi-modal diagnosis. Lastly, clinical validation through collaboration with medical professionals can help establish its real-world applicability.

References

1) Librosa Documentation – McFee, B., et al. (2015). *librosa: Audio and music signal analysis in Python*. [Online] Available: https://librosa.org

2) TensorFlow and Keras API – *TensorFlow: Large-scale machine learning on heterogeneous systems*, Google Inc. Available: https://www.tensorflow.org

3) *IEEE Standards Association*, Software Engineering Standards. [Online] Available: https://standards.ieee.org

4) Hinton, G., et al. (2012). *Deep Neural Networks for Acoustic Modeling in Speech Recognition*. IEEE Signal Processing Magazine, 29(6), pp. 82–97.

5) Sharma, P., & Bansal, R. (2020). *Detection of COVID-19 using Cough, Breath and Speech Sounds: A Review*. International Journal of Advanced Computer Science and Applications (IJACSA), 11(10).

6) Tiwari, V. (2010). *MFCC and its applications in speaker recognition*. International Journal on Emerging Technologies, 1(1), pp. 19-22.

7) Xu, J., et al. (2021). *Multi-Class Respiratory Sound Classification with CNN-LSTM Models*. Computers in Biology and Medicine, Volume.

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

RAJTANU DASGUPTA
2206200

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: My key responsibility was to generate and capture the output of the implemented CNN model. I trained and evaluated the model, captured performance metrics, plotted accuracy/loss graphs, confusion matrix, ROC curves, and tested the prediction function. This gave me deeper insights into model evaluation strategies and performance tuning. I also learned how to interpret classification metrics like precision, recall, F1-score, and how to validate the model using real-time testing.

Individual contribution to project report preparation: I contributed to Chapter 4 – Result Analysis, where I compiled all screenshots and plots with captions and added corresponding interpretations.

Individual contribution for project presentation and demonstration: I presented the testing results and model performance during the demonstration, showing how the predictions aligned with real test data.

Full Signature of Supervisor: Full signature of the student:

…………………………….. ……………………………..

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

RICHIK DEY
2206203

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: I actively contributed to the technical development of the project, specifically in identifying issues within the codebase, debugging errors, and modifying implementations to ensure successful training and evaluation. During the implementation phase, I assisted in resolving compatibility issues related to MFCC input shapes, model architecture mismatches, and prediction processing. I also participated in tuning model parameters and validating outputs during testing to ensure accurate predictions. My hands-on work with error diagnosis helped streamline the model training and improve its performance. In addition to technical support, I took full responsibility for the project report writing. I drafted all chapters of the report, including the Abstract, Introduction, Literature Review, Methodology, Results, and Conclusion. I also formatted the document according to academic guidelines and ensured that visual assets such as graphs, ROC curves, and confusion matrices were well-integrated and explained. These tasks deepened my understanding of machine learning workflows, especially in speech/audio-based classification. I enhanced my ability to translate complex technical work into structured academic documentation and gained valuable experience in collaborative project development.

Individual contribution to project report preparation: I prepared the complete project report, ensuring consistency, clarity, and alignment with project objectives. I also handled the incorporation of all figures, visual outputs, and formatting.

Individual contribution for project presentation and demonstration: I supported the presentation by preparing speaking points for the documentation and quality analysis section. I also guided the team in preparing structured responses based on the report.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………….

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

SOUMYA BHATTACHARJEE
2206219

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: As a core member of the project team, I was primarily responsible for the code implementation of the entire machine learning pipeline. My contribution involved writing Python code for audio data loading, preprocessing using Librosa, extracting MFCC features, augmenting data with noise, pitch shifts, and stretching, and preparing the dataset for training. I implemented the CNN model using TensorFlow/Keras, defined the layers, compiled the model, and trained it using training and validation datasets. I also developed functions for visualization, such as spectrogram plots, accuracy-loss curves, and prediction evaluation using confusion matrix and classification reports. The model was fine-tuned for better performance using class weights and regularization techniques like dropout. During implementation, I encountered several challenges, such as padding MFCCs to ensure consistent input shapes and resolving environment dependency issues related to Librosa and Resampy. Overcoming these helped me strengthen my debugging skills and deepen my understanding of audio classification.

Individual contribution to project report preparation: I handled the complete preparation of the report, including all chapters and formatting. I ensured coherence and accuracy throughout the document.

Individual contribution for project presentation and demonstration: I supported the presentation by preparing speaking points for the documentation and quality analysis section. I also guided the team in preparing structured responses based on the report.

Full Signature of Supervisor: ……………………………..

Full signature of the student: ……………………………..

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

SUBHADEEP MONDAL
2206221

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: I was tasked with preparing the plagiarism report and ensuring originality in all content included in the final submission. I scanned the report using plagiarism detection tools, revised flagged sections, and ensured all figures and texts were cited properly. This taught me about the importance of academic integrity and how to maintain originality while documenting a technical project.

Individual contribution to project report preparation: I proofread the entire document for plagiarism and formatting issues. I also helped in the final review of the content before printing.

Individual contribution for project presentation and demonstration: I ensured the presentation content was plagiarism-free and original. I supported the group in demo coordination and technical Q&A preparation.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………….

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

NILADRI NAG
2206356

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: My role was to collect and create diagrams and flowcharts representing system architecture, data flow, and CNN model structure. I used tools like draw.io and diagram.net to illustrate how the dataset flows from input audio to MFCC extraction, CNN classification, and prediction output. This task helped improve my ability to visualize deep learning workflows and convert them into comprehensible diagrams, which were crucial for explaining the model during presentations and in the report.

Individual contribution to project report preparation: I contributed to Chapter 3 – System Design, especially the block diagram and design constraint sections, and ensured diagrams were properly labeled and inserted.

Individual contribution for project presentation and demonstration: I explained the block diagram and system workflow during the presentation and assisted with organizing content in the slides visually.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………….

INDIVIDUAL CONTRIBUTION REPORT:

SPEECH RECOGNITION IN HEALTHCARE

NABASURJA DATTA
2306602

Abstract: This project focuses on the development of a machine learning model using Convolutional Neural Networks (CNN) for the detection of asthma through audio signals like cough, breath, and speech. The main goal is to design a non-invasive, low-cost, and accessible method for respiratory disease diagnosis using signal processing techniques such as MFCC.

Individual contribution and findings: I was responsible for preparing the PowerPoint (PPT) presentation used during the internal and final evaluations. I structured the slides to cover all phases of the project – introduction, methodology, system design, model architecture, results, and conclusions. I used consistent formatting and visuals for clarity. Creating the presentation improved my technical communication skills and understanding of how to summarize a complete ML project for a diverse audience.

Individual contribution to project report preparation: I contributed to the layout and visual content used in the report, ensuring that figures and text aligned with what was being shown in the presentation.

Individual contribution for project presentation and demonstration: I presented the overview and background slides during the final project demo and introduced the flow of the presentation.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………….

SPEECH RECOGNITION IN HEALTHCARE
PLAGIARISM REPORT

Paperpal Plagiarism Check Standard

Overview

Similarity range

Similarity detected in your file!

17% - 23%

Sources for similarity

1	www.coursehero.com	INTERNET	9%
2	www.coursehero.com	INTERNET	8%
3	www.coursehero.com	INTERNET	7%