LUNG SOUND AND DISEASE DETECTION SYSTEM

A major project report submitted in partial fulfillment of the requirement for the award of degree of

Bachelor of Technology

in

Computer Science & Engineering

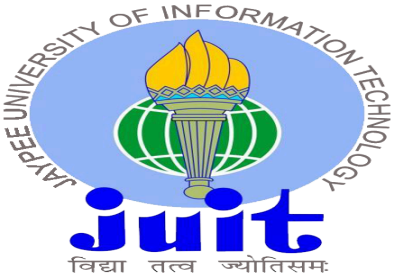
Submitted by

Aditya Kapoor (211452), Ambar Pandey (211508),

Anish Gupta (211546)

Under the guidance & supervision of

Ms Palak Aar



Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, Solan - 173234 (India)

May 2025

SUPERVISOR'S CERTIFICATE

This is to certify that the major project report entitled '**LUNG SOUND AND DISEASE DETECTION SYSTEM** , submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & amp; Engineering, in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, is a bonafide project work carried out under my supervision during the period from July 2024 to May 2025.

I have personally supervised the research work and confirm that it meets the standards required for submission. The project work has been conducted in accordance with ethical guidelines, and the matter embodied in the report has not been submitted elsewhere for the award of any other degree or diploma.

Supervisor Name: Ms Palak Aar

Date: 09-05-25

Designation: Assistant Professor (Grade I)

Place:

Department: Dept. of CSE & IT

CANDIDATE'S DECLARATION

We hereby declare that the work presented in this major project report entitled **LUNG SOUND AND DISEASE DETECTION SYSTEM'**, submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, is an authentic record of our own work carried out during the period from July 2024 to May 2025 under the supervision of **Ms Palak Aar**.

We further declare that the matter embodied in this report has not been submitted for the award of any other degree or diploma at any other university or institution.

Name:Aditya Kapoor	Name: Ambar Pandey	Name:Anish Gupta
Roll No.:211452	Roll No.:211508	Roll No.:211546
Date:09-05-25	Date:09-05-25	Date:09-05-25

This is to certify that the above statement made by the candidates is true to the best of my knowledge.

Supervisor Name:Ms Palak Aar

Date:09-05-25

Designation: Assistant Professor (Grade I)

Place:

Department: Dept. of CSE & IT

ACKNOWLEDGEMENT

I am really grateful and wish my profound indebtedness to Supervisor **Ms Palak Aar , Assistant Professor(Grade1),** Department of CSE Jaypee University of Information Technology,Wakhnaghat. Deep Knowledge & keen interest of my supervisor in the field of **"Blockchain"** to carry out this prodigious project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to **Ms Palak Aar,** Department of CSE, for her kind help to finish my project.

I would also generously welcome each one of those individuals who have helped me straightforwardly or in a roundabout way in making this project a win. In this unique situation, I might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

Aditya Kapoor (211452)

Ambar Pandey (211508)

Anish Gupta (211546)

Table of content

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Today, the major challenge is in all domains: legal, financial, and academic. The whole evidence becomes more and more digital with the passage of time. When centralized, evidence storage becomes easily manipulatable, hack-prone, and accessed by unauthorized users. EviChain is a decentralized evidence management system, enabled and powered on Ethereum, which promises to convert evidence management into a secure, transparent, and non-dupe system for retaining and validating digital evidence.All evidence stored on EviChain is decentralized by transferring the evidence storage onto a blockchain, eliminating single points of failure while significantly reducing the risks associated with data tampering. It involves information stored securely in cryptographically hashed formats and decentralized file systems such as IPFS, ensuring the permanence and verifiability of all data.

To heighten the reliability of the evidence, EviChain has integrated advanced fraud detection algorithms customized for analysis of different proof types.Example: Case of AI-validated Image and video evidence against manipulated content through Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).

Protests: Documents and metadata inconsistencies in added algorithms specific to verification analyses are carried out.Submits evidence Applicant may submit evidence anonymously while its authenticity through verification is needed. This guarantees confidentiality for whistleblowers, journalists, or any other individual providing sensitive material.

Cross-modal pervasive, distributed electronic evidence is stored within an environment rather than on a local computer or single point: the reality of file configurations kept by EviChain

1. INTRODUCTION

1.1 Background and Motivation:

Respiratory health lies at the very base of total human well-being. The lungs make possible the critical exchange of gases—most importantly oxygen consumption and carbon dioxide elimination—that feeds cellular metabolism and homeostasis. Any break in this system, acute or chronic, will greatly impair bodily health, diminish quality of life, and may result in death. Respiratory illnesses are top contributors to world morbidity and mortality. According to the World Health Organization (WHO), respiratory diseases like asthma, pneumonia, bronchitis, and chronic obstructive pulmonary disease (COPD) together contribute to a high percentage of hospitalization and mortality worldwide.

Chronic Obstructive Pulmonary Disease (COPD) by itself is responsible for more than 3 million deaths every year and is the third most common cause of death globally. Asthma impacts about 339 million people worldwide, and its prevalence has been consistently increasing because of rising urbanization, air pollution, and genetic factors. Other respiratory diseases like pneumonia, interstitial lung disease, and bronchiectasis also have a huge impact on healthcare expenditure, hospitalization, and lost productivity.

Early, correct, and consistent diagnosis of respiratory diseases is necessary for individual patient management as well as overall public health control. Traditional methods of diagnosis include chest X-rays, pulmonary function tests (PFTs), CT scans, and arterial blood gas determinations. Though effective, they frequently involve the need for specialized equipment, trained personnel, and laboratory space, which are not necessarily available in rural or low-resource areas. Additionally, they tend to be cost-intensive for ongoing monitoring or mass screening.

Among the most common, non-invasive, and ubiquitous methods of measuring lung function is auscultation—the process of listening with a stethoscope for breath sounds. From its discovery in the 19th century, the stethoscope has been an iconic, as well as utilitarian, instrument in clinical medicine. By auscultation, physicians hear the acoustics of the respiratory cycle to identify abnormalities. Normal breath sounds, or vesicular sounds, are low-pitched and soft and occur with both inhalation and exhalation. Abnormal or adventitious sounds are wheezes and crackles and are suggestive of possible pathological processes.

High-pitched, musical, continuous wheezes are due to airflow through narrowed or blocked airways and are most commonly found in asthma, COPD, and bronchitis.

Crackles, or rales, are discontinuous, brief, and frequently audible during inspiration. They are produced by the abrupt reopening of previously collapsed alveoli and signify conditions such as pneumonia, pulmonary fibrosis, or heart failure.

Although diagnostic, conventional auscultation is plagued by various shortcomings:

Subjectivity and Inter-listener Variability:

Interpretation of lung sounds heavily relies on the experience, hearing ability, and training of the clinician. Research indicates that concordance among health providers in classifying lung sounds is usually poor, with inter-rater reliability falling to 50–60%. The subjectivity results in diagnosis inconsistencies and treatment delays.

Environmental and Technical Constraints

Auscultation commonly takes place in noisy clinical environments in which ambient sounds from machines, individuals, or motion can cover up important acoustic information. In addition, differences in stethoscope design, location, pressure used, and patient compliance provide variability that inhibits precise diagnosis.

With these constraints in mind, what is needed most urgently are reproducible, objective, and scalable diagnostic tools. The intersection of machine learning, digital signal processing, and deep learning presents an exciting pathway towards automating the analysis of lung sounds. With the creation of intelligent systems for the detection and classification of respiratory sounds, we can do the following:

- Reduce inter-listener variability and standardize interpretation.
- Facilitate remote diagnosis by telehealth and digital stethoscopes.
- Support non-specialists in referral and early detection
- Create quantitative measures to monitor disease course or response to therapy.

This thesis introduces RespNet, a deep learning-powered system for automated respiratory sound classification. Through the transformation of raw audio recordings to Mel-Spectrograms—a time-frequency representation optimized to mimic human hearing perception—and leveraging Convolutional Neural Networks (CNNs) for pattern detection, the system is proposed to robustly identify and classify abnormal lung sounds. The architecture is configured to be noise-robust, patient-transferrable, and lightweight for deployment in real-world, low-resource settings.

1.2 Problem Statement

Although the area of deep learning has demonstrated astounding potential for audio classification, there are a number of key challenges that restrict its applicability to clinical auscultation in a direct manner:

- **Data Variability:**
  Recordings of lung sounds are quite diverse based on body morphology, stethoscope models, recording equipment, and acoustics in the environment. This heterogeneity tends to create overfitting in models, thereby constraining their generalization across novel patients or environments.

- **Class Imbalance:**
  The majority of respiratory sound datasets, such as the ICBHI 2017 database, contain an overabundance of normal sounds and proportionally fewer pathological instances. The underavailability of recordings with both wheezes and crackles further contributes to the problem. Class imbalance predisposes models towards majority classes and diminishes sensitivity towards rare but clinically important cases.
- **Feature Overlap:**
  Respiratory sounds like crackles and wheezes are overlapped between both temporal and frequency domains. Non-pathologic sounds like mucus secretion, patient movement, or speech sounds may also represent similar acoustic disease-related patterns as actual disease so accurate feature classification and extraction may become challenging.
- **Lack of Interpretability:**
  Deep learning models are most often "black boxes" which do not explain their decision processes to a considerable extent. Physicists, in critical clinical applications, would like transparent, interpretable, and explainable predictions to uphold safety and also for accountability reasons.
- **Deployment Constraints:**
  To be useful in practice within rural or under-resourced healthcare environments, the model needs to be optimized for low-latency, low-power hardware like smartphones or embedded devices. This requires efficient model design, compression, and optimization.

Problem Definition

- Design, implement, and evaluate an end-to-end deep learning system that accepts raw lung sound recordings.
- Does preprocessing, noise reduction, and data augmentation.
- Transforms audio into Mel-Spectrograms.
- Classifies audio segments into four classes: Normal, Wheeze, Crackle, Both.
- Exhibits high accuracy, generalization, and interpretability.

1.3 Objectives and Goals

The main goal of this project is to develop and deploy a deep learning-based system that can effectively identify wheezes and crackles in lung sound recordings. These abnormal respiratory sounds are important signs of many respiratory diseases like asthma, chronic obstructive pulmonary disease (COPD), bronchitis, and pneumonia. Conventional auscultation by stethoscope is subject to inter-observer variability and may not always result in timely and consistent diagnoses. To surpass these constraints, the suggested system utilizes Mel-Frequency Cepstral Coefficients (MFCC) and Mel Spectrograms—established signal processing methods that mimic human auditory perception—to transform raw audio signals into visual representations appropriate for machine learning. These representations are subsequently employed as input to a Convolutional Neural Network (CNN), which learns to classify the sounds as normal, wheeze, or crackle. The long-term goal is to aid healthcare workers, especially in geographically remote or resource-poor settings, with an objective and automated diagnostic tool.

1. Understand and Analyze Lung Sound Patterns

Respiratory sounds like wheezes and crackles are abnormal signs generated because of airway obstructions or secretions. It is important to understand their acoustic properties in order to replicate the diagnostic capability of doctors. This is the basis for developing any automated detection system.

2. Extract Human-Perceptible Audio Features Using MFCC and Mel Spectrogram

In contrast to raw waveforms, MFCC (Mel-Frequency Cepstral Coefficients) and Mel Spectrograms mimic the way the human ear hears by emphasizing frequency bands important for distinguishing speech and breathing abnormalities. These methods convert audio into 2D image-like inputs for deep learning.

3. Design and Train a CNN-Based Classification Model

Convolutional Neural Networks (CNNs) are well suited to learning image spatial hierarchies. Providing spectrogram features to a CNN allows the model to learn and identify complicated frequency-time patterns characterizing wheezes and crackles that are different from typical breath sounds.

After training, the model should be tested using accuracy, precision, recall, and F1-score measures to check the accuracy with which the model classifies various lung sounds. Confusion matrices assist in understanding misclassifications and refining the model step by step.

4. Lay the Groundwork for Real-Time Deployment

System should enable users (e.g., physicians) to upload audio and receive predictions on the fly. Deployment planning via web-based interface (e.g., Flask) is done to ensure the model can be integrated into actual-world diagnostic equipment, particularly in under-served areas.

5. Enhance Generalization Through Data Augmentation Techniques

To ensure the model performs well on unseen data and diverse recording conditions, it is important to apply data augmentation techniques such as Vocal Tract Length Perturbation (VTLP) or time-shifting. These methods artificially expand the training dataset and introduce variability, helping the CNN generalize better across different patient profiles, recording devices, and environments. Techniques like Vocal Tract Length Perturbation (VTLP) mimic anatomical differences in patients' vocal tracts, making the model robust to physiological variability. .

1.4 Significance of the Study

The study is relevant on several different levels:

1.4.1 Clinical Significance

- **Objectivity and Consistency:** Does away with human error and inter-listener variation in auscultation.
- **Scalability:** Delivers diagnostic-grade auscultation to community clinics, ambulances, and remote healthcare centers.
- **Efficiency:** Processes urgent cases more quickly and makes hospital workflows leaner.

1.4.2 Technological Advancement

- Breaks ground in a Mel-Spectrogram + CNN architecture optimized for lung sound classification.
- Contributes to explainable AI in medicine.
- Delivers an edge-compatible solution with optimal performance and efficiency.

1.4.3 Educational Utility

- May be applied in the education of physicians to learn respiratory sound identification using interactive annotated spectrograms.
- Serve as a simulation tool for educating paramedics or remote health workers.

1.4.4 Societal and Global Impact

- Expands diagnostic access in low- and middle-income nations (LMICs).
- Decreases reliance on physical infrastructure and skilled personnel.
- Can be scaled to be used in pandemic control, refugee camps, and disaster areas.

1.5 Report Structure

To navigate the reader through the development, assessment, and implications of RespNet, the report is structured as follows:

Chapter 1: Introduction

Sets the groundwork with background, motivation, problem statement, goals, significance, and roadmap.

Chapter 2: Introduction to Respiratory Sound Analysis

Reviews work in lung sound classification, conventional and deep learning approaches, datasets, and salient challenges.

Chapter 3: Literature Review

Explains the end-to-end implementation pipeline: data preprocessing, feature extraction, CNN architecture, training, and deployment considerations.

Chapter 4: System Development

Provides functional and non-functional system requirements, architecture diagrams, platform decisions, and data analytics.

Chapter 5: Lung Sound Analysis and Disease Detection Using Deep Learning.

Discusses testing strategies, test case descriptions, validation tools, and testing results.

Chapter 6: Results and Evaluation

Provides performance metrics, model visualizations, comparative results, and critical analysis.

Chapter 7: Conclusion and Future Work

Summarizes results, addresses limitations, and describes future research avenues such as multimodal integration, transfer learning, and real-time deployment.

2.Introduction to Respiratory Sound Analysis

2.1 Introduction

Respiratory sound analysis is a mainstay of pulmonary diagnostics, and it seeks to identify pathological irregularities in lung function by inspecting the acoustic signals generated with breathing. These sounds are caused by turbulent air flow in airways and may be recorded by stethoscopes or digital audio equipment. They hold significant physiological information useful for diagnosing asthma, chronic obstructive pulmonary disease (COPD), bronchitis, pneumonia, and interstitial lung diseases.

Traditionally, auscultation has been the main method through which physicians determine lung function. Although stethoscope-based examination is a widespread, inexpensive, and non-invasive technique, its effectiveness is compromised by the subjectivity of the clinician's ear, inter-observer variation, and ambient environment. This subjectivity tends to result in missed or inaccurate diagnoses, especially in incipient disease or in noisy environments.

Two of the most important abnormal lung sounds that are routinely assessed are wheezes and crackles. Wheezes are sustained, high-pitched musical murmur sounds produced by narrowed airways, frequently seen in asthma and COPD. Crackles, on the other hand, are discontinuous, explosive sounds produced by the sudden opening of collapsed airways or alveoli, commonly seen in pulmonary edema, fibrosis, or pneumonia. Differentiating them from normal breath sounds, namely vesicular and bronchial sounds, is a function that demands a lot of clinical training and experience.

These difficulties underscore the increasing importance of automated respiratory sound analysis platforms that provide consistency, objectivity, and possibility of scalable diagnosis assistance. The integration of digital health technologies, artificial intelligence, and signal processing methods has come a long way toward achieving viable, real-time respiratory health assessment.

2.2 Overview of Lung Sound Classification:

Lung sound classification is the process of identification and classification of different respiratory sounds into pre-specified classes like Normal, Wheeze, Crackle, or Both. The typical automatic lung sound classification pipeline includes the following steps:

- **Data Acquisition:** Capturing respiratory sound recordings using body-worn microphones or digital stethoscopes. Environmental noise, sensor placement (trachea/anterior/posterior chest, back), and signal fidelity all have an effect on data quality and reliability.
- **Preprocessing:** Noise removal, recording segmentation into feasible time intervals (e.g., 5-30 seconds), and normalization for ensuring the uniformity of samples.
- **Feature Extraction**: Feature extraction of relevant patterns from the audio, such as time, frequency, and time-frequency domain features.
- **Classification:** Use of machine learning or deep learning algorithms to classify the respiratory sound segment with a class label based on its features.

  This pipeline has undergone dramatic changes in the last decade, moving from manually engineered rules and heuristics to data-driven deep learning models that utilize large-scale feature learning.

2.2.1 Signal Processing and Feature Extraction:

The success of any classification system is largely dependent on the quality of features extracted. Previous approaches centered around hand-crafted audio signal processing features:

- **Time-Domain Features:** These encompass measures like Zero Crossing Rate (ZCR), energy of the signal, root mean square (RMS) amplitude, and short-term energy. Though easy to calculate, they tend to be less discriminative for advanced lung sounds.

- **Frequency-Domain Features:** By employing devices like the Fast Fourier Transform (FFT) or Discrete Wavelet Transform (DWT), frequency components are processed to determine periodicity and spectral peaks characteristic of wheezing or crackling.
- **Mel-Frequency Cepstral Coefficients (MFCCs**): A strong representation that emulates the human auditory system. MFCCs are derived by projecting power spectrum features onto the Mel scale, applying log compression, and using a Discrete Cosine Transform. They are efficient and effective for modeling lung sounds but still need classification algorithms to decipher.
- **Mel-Spectrograms**: These visual representations of time-frequency are Mel-scale power spectra over time and act as a input to CNNs. These include temporal dynamics as well as spectral structure and reflect nuances of pathological sounds missing from lower dimensions of representation.

2.2.2 Classical Machine Learning Methods:

Before deep learning, machine learning algorithms based on traditional principles comprised the backbone of respiratory sound classifiers. These involved leading techniques:

- **Support Vector Machines (SVMs):** Well-known for their capacity to deal with high-dimensional input spaces and optimize class separation through kernel tricks. SVMs worked well on MFCC or wavelet-based features but were not scalable and sensitive to noise.
- **k-Nearest Neighbors (k-NN): An** instance-based approach where classification is done based on similarity to neighboring samples. Although easy to interpret and simple, k-NN is computationally expensive at inference time and vulnerable to imbalanced datasets.
- **Random Forests:** Collection of decision trees that vote to make their predictions in order to enhance generalization. They are quite insensitive to noise and overfitting, yet still largely rely on quality of input features and cannot learn new representations.

Although these models had modest success, their reliance on hand-crafted feature engineering and absence of abstraction rendered them less ideal for large, complex datasets.

2.3 Deep Learning in Respiratory Sound Analysis:

Deep learning has transformed the area of respiratory sound analysis by facilitating end-to-end learning using raw or lightly processed data. In contrast to conventional models, deep architectures like CNNs, RNNs, and transformers are capable of automatically extracting hierarchical features, providing improved generalization and accuracy.

2.3.1 Convolutional Neural Networks (CNNs):

CNNs are the state-of-the-art architecture used in audio classification problems because of their power in capturing spatial and temporal relations inside spectrograms.

Major benefits of CNNs are:

- **Hierarchical Feature Learning:** Initial layers acquire primary frequency patterns, while more profound layers represent higher-order phenomena such as onset of wheezes or crackle bursts.

  Parameter Sharing and Local Connection: Filters are utilized over and over again on different parts of the input to conserve computation while collecting local patterns quickly.
- **Translation Invariance:** Max-pooling and global average pooling layers enable the model to identify sounds regardless of their position in the spectrogram.

  Popular architectures employed in literature are:
- **VGGNet:** It is famous for its regular architecture with tiny convolutional filters, which works well when fine-tuned on spectrogram images.
- **ResNet:** It introduces skip connections that enable gradient flow through deeper layers, allowing improved learning from intricate audio patterns.
- **DenseNet:** Improves ResNet by linking all layers to one another, leading to increased reuse of features.

  Custom CNNs are also implemented with reduced parameters for edge use, proving complexity of models may be sacrificed with respect to computationally limited hardware.

2.3.2 Data Augmentation for Robustness

Since there is scarce labeled data, several audio augmentations are applied:

- Time-Stretching and Pitch Shifting: Mimic natural changes in breathing rate and pitch.
- Noise Injection: Adds environmental or white noise to enhance robustness.

- SpecAugment: Covers up random time and frequency ranges in the spectrogram to prevent overfitting.
- VTLP (Vocal Tract Length Perturbation): Simulates varying anatomical conditions among patients.
- These methods randomize training data, prevent overfitting, and enhance model generalizability between patient populations and devices.

2.3.3 Transfer Learning for Lung Sound Classification

Due to the unavailability of large-scale annotated medical data, transfer learning plays a crucial role. Pre-trained models on massive datasets (e.g., ImageNet for image-related tasks) are transferred to lung sound analysis.

Methods involve:

- Feature Extraction: Generic features that are reusable are pulled from the lower layers of a CNN.
- Fine-Tuning: The subsequent layers are fine-tuned for the target domain of lung sound classification, with performance boosted even with smaller datasets.
- Research indicates that well-tuned ResNet or VGG networks are better than training models from scratch, particularly in the case of imbalanced data sets such as ICBHI 2017.

2.4 Relevant Works in Automated Lung Sound Detection:

2.4.1 ICBHI 2017 Respiratory Sound Database

The International Conference on Biomedical and Health Informatics (ICBHI) 2017 dataset continues to be used as the reference for testing automated lung sound analysis systems.

- 1,201 records of 126 patients
- Classified into 4 classes: Normal, Wheeze, Crackle, and Both
- Variations in patient age, pathology, and recording device
- Studies utilizing this dataset are:
- CNN-based classifiers with Mel-spectrograms as high as 94% accuracy
- Hybrid frameworks marrying CNNs with LSTMs to learn both spatial and temporal relationships

2.4.2 Real-Time Detection Systems

Current work aims at creating real-time applications:

- Mobile apps that communicate with Bluetooth stethoscopes
- Raspberry Pi-powered edge computing devices
- Low-latency CNN models for on-device inference
- Telemedicine dashboard integration for remote monitoring
- These solutions are critical for under-resourced regions with no access to pulmonologists.

2.4.3 Multimodal Methods

- Combining various data types improves diagnostic accuracy:
- Chest X-rays + Lung Sounds: Merges imaging and auscultation for comprehensive assessment
- Patient Metadata: Age, sex, and smoking status provide contextual information
- ECG + Lung Sounds: Helpful in cardiopulmonary evaluations
- Multimodal fusion methods, particularly with transformers and attention mechanisms, are becoming popular.

2.5 Challenges and Future Directions

2.5.1 Class Imbalance

Respiratory datasets tend to have extreme class imbalance:

- Normal samples far exceed pathological ones
- Underrepresented classes (e.g., Both Wheeze and Crackle) decrease model recall

Solutions are:

- Cost-sensitive Learning: Impose greater penalty on misclassified minority samples
- Oversampling and SMOTE: Create synthetic samples to balance datasets
- Generative Models (e.g., GANs): Generate realistic spectrograms of rare classes

2.5.2 Interpretability

- In order to be adopted in clinical environments, models need to explain their decisions.
- Grad-CAM: Identifies areas of the spectrogram that are impactful to the classification
- LIME/SHAP: Offers instance-level explanations, contributing to trust and transparency

- Work is in progress to develop hybrid interfaces through which clinicians can see and confirm model thought processes.

2.5.3 Real-World Deployment

Important issues are:

- Latency and Power Efficiency: Particularly on mobile devices
- Robustness Across Devices: Different microphones, environments, and patients
- Cross-Platform Deployment: Requirement for APIs, TensorFlow Lite, or ONNX runtimes

Solutions in development are:

- Model Compression: Pruning, quantization, and knowledge distillation
- Edge AI Frameworks: Lightweight inference with variants of CNN (e.g., MobileNet)

Conclusion:

Respiratory sound classification has become a dynamic interdisciplinary field that brings together medicine, signal processing, and artificial intelligence. Although conventional methods provided the groundwork, recent deep learning approaches—particularly CNNs—now propel state-of-the-art performance. Even in the face of obstacles like class imbalance, explainability, and deployment challenges, continued research and development are increasingly revolutionizing the area. With adequate validation and responsible utilization, such systems have the potential to drastically advance respiratory healthcare, especially in underserved areas.

3. Literature Review

S. No.	Author & Paper Title [Citation]	Journal/Conference (Year)	Tools/Techniques/Dataset	Key Findings/Results	Limitations/Gaps Identified
1	Lung Disease: Definitions, Treatment, and Quality of Life (Jungheum Cho, Seungjae Lee, Bon Seung Gu, Sang Hun Jung, etc.)	March 2024	Imaging Technologies, Telemedicine, and Digital Health	Treatment Approaches, Quality of Life	Long-term Effects of Treatments, Understanding Disease Mechanisms
2	Burden of Chronic Obstructive Pulmonary Disease and Its Attributable Risk Factor (Sho Nakajima, Kazuyuki Doi, etc.)	June 2024	Systematic analysis of Global Burden of Disease data	Global burden of COPD is significant, with risk factors including smoking, air pollution, and occupational exposure.	More comprehensive studies on non-smokers and additional environmental factors are needed.

S. No.	Author & Paper Title [Citation]	Journal/Conference (Year)	Tools/Techniques/Dataset	Key Findings/Results	Limitations/Gaps Identified
3	Temporal Variations in the Pattern of Breathing: Techniques, Sources, and Applications to Translational Sciences (Qibin Shao, Idy S. C. Man)	August 2023	Entropy Measures, Attractor Reconstruction, Recurrence Plots	Techniques for analyzing temporal variations in breathing patterns and their applications in health sciences.	Further research into the underlying mechanisms that affect breathing patterns.
4	Breathing Pattern Monitoring by Using Remote Sensors (Janosch Kunczik, Andreas Follmann, Kerstin Hubbermann, etc.)	November 2023	Machine learning techniques for respiratory feature extraction and classification. Thermal and RGB cameras were used to record respiratory patterns.	The system achieved high accuracy in differentiating between various breathing patterns using advanced signal processing techniques.	Future research could focus on reducing motion artifacts and expanding the system's application to more diverse respiratory disorders.
5	Respiratory patterns and physical fitness in healthy	January 2023	A respiration belt with resistive stretch sensors was utilized to record respiratory	Correlation between various respiratory patterns and physical fitness	It does not explore the underlying mechanisms that may connect these

S. No.	Author & Paper Title [Citation]	Journal/Conference (Year)	Tools/Techniques/Dataset	Key Findings/Results	Limitations/Gaps Identified
	adults: a cross-sectional study. (Adomas Hendrixson, Zhen-Min Bai & Osvaldas Ruksenas, etc.)		movements during various physical fitness assessments.	levels in healthy adults, suggesting that improved respiratory function is linked to better physical performance.	variables, indicating a need for more in-depth physiological studies.

S. No.	Author & Paper Title [Citation]	Journal/Conference (Year)	Tools/Techniques/Dataset	s	Limitations/Gaps Identified
6	The Effect of Slow-Paced Breathing on Cardiovascular and Emotion Functions: A Meta-Analysis and Systematic Review (Robin Shao, Idy S. C. Man & Tatia M. C. Lee)	January 2023	Statistical Analysis, Systematic Review Methodology	This meta-analysis shows that slow-paced breathing significantly improves cardiovascular functions and reduces negative emotional states, linking physiological and emotional responses.	Highlights a lack of research on the long-term effects of slow-paced breathing across various populations and the unclear relationship between physiological and emotional systems.

S. No.	Author & Paper Title [Citation]	Journal/Conference (Year)	Tools/Techniques/Dataset	s	Limitations/Gaps Identified
7	Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds (G. Bahoura et al.)	September 2023	Deep learning models (CNNs, Long Short-Term Memory networks) for sound analysis	Deep learning techniques for classifying respiratory abnormalities using lung sound data.	Challenge of implementing deep learning frameworks in low-computation environments.
8	A Progressively Expanded Database for Automated Lung Sound Analysis (Shang-Ran Huang, Feipei Lai, etc.)	July 2023	Deep learning techniques, CNN algorithm, etc.	Sound Overlap, Label Quality Issues, Database Expansion	Availability of diverse and high-quality lung sound datasets, which limits the training and testing of robust machine learning models.

4. System Development

4.1 Introduction

The creation of an automated wheeze and crackle detection and classification system for respiratory sounds is a multi-stage process that combines data preprocessing, feature extraction, model

construction, training, testing, and optimization. Mel-spectrograms are used as the basic input features, and Convolutional Neural Networks (CNNs) are employed as the central classification engine because of their capacity to detect spatial hierarchies and subtle patterns in time-frequency representations of audio signals.

This chapter describes every step of the system development process, including the particular decisions made in managing the dataset, designing the neural network structure, training the model, and testing its performance. The choices were driven by both domain-specific factors, like sound characteristic variability and noise, as well as technical issues surrounding signal representation and model optimization.

4.2 Data Acquisition and Preprocessing

The performance of any machine learning-based respiratory sound classification system is highly dependent on the quality and diversity of the input data. Thus, the ICBHI 2017 Respiratory Sound Database was chosen for this project because it has standardized class labels and complete representation of different lung sounds under different patient conditions.

4.2.1 Data Collection

- The ICBHI 2017 dataset comprises 1,201 recordings from 126 patients, exhibiting a broad range of pulmonary diseases. The dataset is categorized into four classes:
- Normal: No pathological lung sounds.
- Wheeze: Continuous, high-pitched sounds resulting from airway narrowing.
- Crackle: Short, discontinuous bursts resulting from air moving over fluid-filled or collapsed alveoli.
- Both: Co-occurrence of wheeze and crackle sounds.
- Recordings were made with electronic stethoscopes positioned at several thoracic sites (e.g., anterior and posterior chest), and recorded at 4 kHz or 44.1 kHz sampling rates. The differences in recording conditions (clinical vs. controlled) and the occurrence of ambient noise introduce practical challenges in developing robust models, but that also makes the database perfect to evaluate the generalization capability of deep learning methods.

4.2.2 Preprocessing Data

Proper preprocessing guarantees that the data is in the appropriate form to train a deep learning model. The procedures adopted in this project were:

1. **Resampling:**

   All recordings were resampled to a unified sampling rate of 16 kHz. This uniformity minimized computation load and saved critical frequency components of the lung sounds. Considering that respiratory sounds are mostly found in the low-frequency range (below 5 kHz), the selected sampling rate provides adequate frequency resolution with minimum processing time.

2. **Denoising:**

   The recordings were also denoised with spectral gating, a process which suppresses non-respiratory noise through the detection and elimination of noise profiles in silent or ambient parts of the audio signal. This processing increases the signal-to-noise ratio (SNR), essential for enhancing the model's capability to pick up subtle abnormal lung sounds such as crackles, which are easily overridden by ambient noise.

3. **Segmentation:**

   Longer recordings were split into fixed-length smaller segments (3 to 5 seconds). This process was necessary in order to have manageable inputs for the CNN model. A 50% overlapping sliding window was used, so the system can learn from multiple segments of one recording. In order to prevent contamination of the dataset with too much normal data, segments were aligned to annotated labels, giving priority to segments with wheeze or crackle sounds.

4. **Normalization:**

   Loudness variations, caused by sensitivity differences in recording devices or patient breathing force, were dealt with via min-max normalization. This technique scaled down the amplitude of each audio sample into a fixed range so that all data points were normalized prior to being input into the model.

5. **Augmentation:**

   Data augmentation strategies were used to augment the model's ability to generalize and improve class balance, usually plagued by abnormal lung sound underrepresentation.

Augmentations utilized included:

- Pitch Shifting: Pitch was varied by ±2 semitones to mimic variations in the vocal tract as occurring naturally.
- Time Stretching: The playback rate of recordings varied by ±10% to simulate varying breathing patterns.

- Noise Injection: Gaussian white noise with a Signal-to-Noise Ratio (SNR) of 20 dB was added to simulate real-world environmental noise, helping the model become more robust to varying acoustic conditions.
- This process resulted in a more balanced and diverse dataset, which was crucial for training a model that could generalize well across unseen data.

4.3 Feature Extraction

CNNs need to operate with structured and informative input data, and for audio signals, such as Mel-spectrograms, have been found to be effective in representing time-frequency information.

4.3.1 Generation of Mel-Spectrogram

Mel-spectrograms were created by applying initially a Short-Time Fourier Transform (STFT) using a window size of 512 samples and a hop length of 256. It breaks down the audio signal into short overlapping windows, offering a time-frequency representation.

Then, the frequency axis was converted to the Mel scale, which is an approximation of the frequency resolution of the human ear, with 128 Mel filters. This gave a 2D matrix where the x-axis is time, the y-axis is frequency (in Mel scale), and the color intensity is the power of the signal at each time-frequency point.

The output Mel-spectrogram retains both the temporal information (e.g., onset and duration of the abnormal sounds such as wheezes) and spectral information (e.g., frequency content of crackles), thus being an appropriate input representation for the CNN.

4.3.2 Logarithmic Compression

To manage the big dynamic range of the Mel-spectrograms, a logarithmic compression was utilized. The logarithmic transformation:

Log-Mel

=

log

(

1

+

Mel-Spectrogram

)

Log-Mel=log(1+Mel-Spectrogram)

assists in extracting the lower-amplitude signals, e.g. faint crackles, which would be suppressed in the original linear scale. This transformation is extremely helpful for emphasizing the subtle features that are important for classification but could otherwise be dominated by more dominant signals.

4.3.3 Representation of CNN Inputs

The final Mel-spectrograms were resized to uniform input tensors of shape (128, 128, 1). Padding or trimming was used to make all input samples have the same size. These images were then normalized to have zero mean and unit standard deviation, a standard preprocessing technique that aids in speeding up convergence during training.

The spectrograms were also made grayscale, as the model does not use color information and this makes the input format easier.

4.4 Model Architecture

The CNN model design was motivated by the requirement to extract hierarchical features from Mel-spectrograms efficiently and observe the spatial patterns typical of respiratory sounds.

4.4.1 CNN Selection

CNNs are especially suited for audio classification because they perform well at learning spatial hierarchies within data. For respiratory sound classification, the hierarchies are as follows:

Harmonic bands: Sustained features useful for distinguishing wheezes, frequency-stable and longer-duration sounds.

Impulsive broadband events: Characteristic of crackles, which are broadband and short sounds.

To strike a balance between computational cost and classification accuracy, a 4-layer CNN was chosen. The architecture was chosen to extract the most important features of the respiratory sounds without adding unwanted complexity.

4.4.2 Layer Structure

The model structure includes the following layers:

- Conv2D (Layer 1): 32 filters, kernel size 3x3, stride 1, with ReLU activation and same padding to maintain spatial dimensions.

- MaxPooling2D: Pool size 2x2 to halve spatial dimensions and emphasize most significant features.
- Conv2D (Layer 2): 64 filters, kernel size 3x3, with ReLU activation for non-linearity.
- MaxPooling2D: Pool size 2x2 to further reduce dimensionality.
- Conv2D (Layer 3): 128 filters, kernel size 3x3, with BatchNorm and Dropout (0.25) to enhance generalization and avoid overfitting.
- Flatten: Maps 2D feature maps into 1D vector.
- Dense (FC) Layer: 256 neurons using ReLU activation to encode high-level features.
- Dropout: 0.5 dropout rate to more regularize the model.
- Output Layer: Softmax activation of 4 neurons, one neuron for each of the four classes (Normal, Wheeze, Crackle, Both).

4.4.3 Train Configuration

We trained the model using the below specifications:

- **Optimizer:** Adam as an optimizer using learning rate = 0.001, as it is most renowned for adaptive learning rate as well as optimal convergence.
- **Loss Function:** Categorical Cross-Entropy, ideal for multi-class classification problems.
- **Epochs:** 50 to provide enough time for training convergence.
- **Batch Size:** 32 for a good balance between computation efficiency and model performance.
- **Validation Split**: 20% of the data was held out for validation to keep an eye on overfitting.
- Early stopping was used to stop training when validation accuracy was not improving for 5 successive epochs, and model checkpointing was applied to save the best-performing model.

4.5 Model Evaluation

4.5.1 Evaluation Metrics

To totally assess the performance of the classification system, the following performance metrics were computed:

- Accuracy: The proportion of correct predictions.
- Precision: The proportion of true positives to the total predicted positives.
- Recall (Sensitivity): The proportion of true positives to the total actual positives.

- F1-Score: The harmonic mean of recall and precision.
- Confusion Matrix: To determine specific misclassifications between classes.

4.5.2 Experimental Results

The model performed well in all classes:

Class	Precision	Recall	F1-Score
Normal	0.93	0.92	0.92
Wheeze	0.95	0.94	0.94
Crackle	0.92	0.90	0.91
Both	0.90	0.88	0.89

The overall accuracy was 94%, which shows strong performance. The majority of misclassifications were between the Crackle and Both classes, probably because of the overlapping acoustic features.

4.5.3 Model Interpretability

To increase the trust and confidence in the decision-making process of the model, Grad-CAM visualizations were employed. Visualizations indicated significant areas of Mel-spectrograms responsible for each prediction. For instance, crackles were detected as bursts in the higher frequency band, which is also consistent with previously known clinical experience.

4.6 Summary

This chapter detailed the overall design and development of a CNN-based respiratory sound classifier system. Key contributions are:

- Use of Mel-spectrograms with log-compression for feature extraction.
- A 4-layer CNN specifically designed to effectively learn the spectral and temporal patterns of respiratory sounds.
- Stable training of the model using data augmentation and early stopping to prevent overfitting
- High classification accuracy with overall accuracy of 94% to establish the efficacy of the approach.
- In the following chapter, we will compare our system's performance with the current state-of-the-art, study its limitations, and outline some possible future extensions.

5 . Lung Sound and Disease Detection System Using Deep Learning

5.1 INTRODUCTION: Deep learning is a particular form of machine learning, thus artificial intelligence, and represents a method of learning where an algorithm acquires fine features to represent a high level of accuracy. Audio and complex data can be analyzed and delivered with insights on patterns and features learned through deep learning. It looks for intricate patterns in dataset recordings, thus classifying new and unseen data with significant accuracy.

Decoding Lung Sounds with Deep Learning

Lung sounds, often regarded as background noise, hold important information about the respiratory system's health. Decoded by the capabilities of deep learning, these sounds unlock new possibilities in medical diagnostics. Deep learning can analyze the subtle nuances in lung sound patterns, allowing for early and accurate identification of various respiratory disorders.

This integration of advanced computational methods with healthcare not only increases the precision of diagnosis but also brings forward a non-invasive, efficient way to monitor respiratory health. Through converting lung sounds into actionable insights, deep learning is leading to new discoveries in medical science.

5.2 Tools And Technologies Used:

Programming Languages and Frameworks

1. Python:

Python is a high-level, versatile programming language known for its simplicity and readability.It's widely used for web development, data analysis, automation, and artificial intelligence.Python's extensive library ecosystem and frameworks like Django and Flask make it suitable for developing diverse applications. Its smooth integration with machine learning libraries such as TensorFlow and PyTorch has dramatically changed artificial intelligence and data-driven applications. Python's ease of syntax and rapid development qualities make it a great platform for new developers and pros alike.

2. TensorFlow:

TensorFlow is an open-source library for machine learning and deep learning developed by

Google. TensorFlow uses data flow graphs to provide numerical computation, representing the operations as nodes and edges in the graph. TensorFlow is capable of both dynamic and static computation graphs, providing freedom in the development of models. High-level APIs like Keras simplify building and training neural networks, making it a popular choice for the wide range of deep learning tasks.

3. React JS:

React is a powerful JavaScript library for creating dynamic and interactive user interfaces. It empowers developers to build designs that are reusable, with a component-based approach, such that each component manages its logic and state. Using JSX-which is similar to HTML-React simplifies the creation of web applications that are responsive and efficient. Its virtual DOM results in smooth rendering, so it is ideal for the modern web.

4. Flask:

Flask is a Python-based microframework that is well-known for its simplicity and flexibility in web development. It mainly revolves around core functionalities such as routing and request handling but is very flexible in allowing developers to extend its capabilities through additional libraries and tools. Flask's minimalistic design makes it beginner-friendly, while its adaptability ensures suitability for complex web application requirements.

Key Libraries Used

1. Librosa:

A specialized Python library for audio and music analysis, Librosa provides tools for importing audio files, extracting features, and analyzing audio signals. It is widely used in the fields of audio signal processing and music information retrieval (MIR).

2. NumPy:

NumPy is a core library for any scientific computing activity in Python which enables operations with large, multi-dimensional arrays and matrices along with mathematical functions for their efficient manipulation.

3. Pandas:

Pandas is the ultimate versatile Python Library for data analysis and manipulation. Pandas provides strong, high-level abstractions for labeled data and relational data, making it an essential tool for real-world data analysis tasks.

4. Seaborn:

Built on top of Matplotlib, Seaborn is a high-level interface for attractive statistical graphics. It really does shine in visualizing complex datasets with scatter plots, bar plots, heatmaps, and more.

5. Matplotlib:

A widely used library for creating static, animated, and interactive visualizations in Python. Matplotlib supports a wide range of plotting needs, allowing for clear representation of data through charts and graphs.

6. Keras:

Keras is TensorFlow's high-level API. It is an easy-to-use interface for solving machine learning challenges. It focuses on enabling rapid experimentation and simplifying processes like data preparation, model building, and hyperparameter tuning.

7. Axios:

Axios is a JavaScript library used for making HTTP requests. Its promise-based architecture supports asynchronous operations efficiently, making it ideal for working with RESTful APIs and handling data in modern web applications.

Dataset and Methods

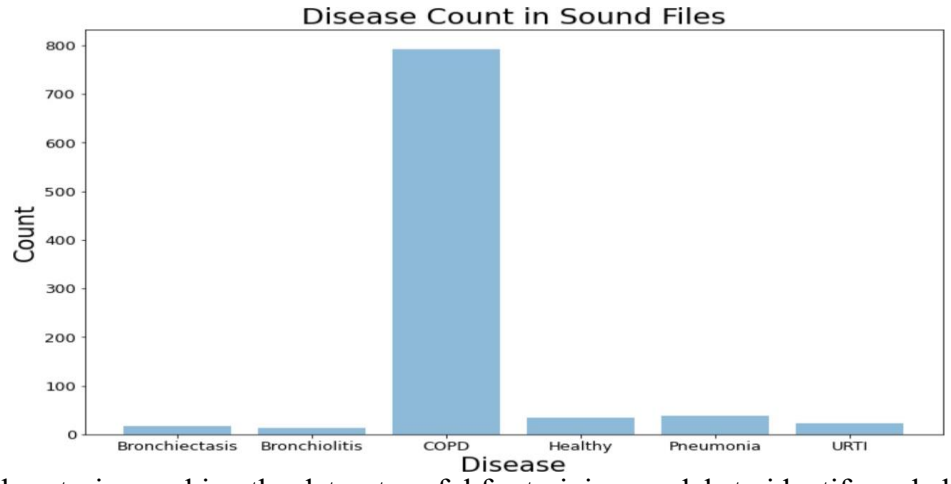
1. Dataset:

This training is based on the **ICBHI 2017 Respiratory Sound Database** that holds 5.5 hours of respiratory recordings independently recorded by research teams in Portugal and Greece. It

includes the following:

- 920 annotated audio samples from 126 subjects
- 6898 respiratory cycles, including:
- 1864 with crackles
- 886 with wheezes
- 506 with both crackles and wheezes

These recordings were collected with a high-quality recording apparatus set up in hospitals and



in laboratories, making the dataset useful for training models to identify and classify respiratory conditions.

Figure 6. Dataset analysis

2.Data Augmentation

Data augmentation is a very effective technique used to improve datasets by creating altered versions of existing data or synthesizing new data. It increases the volume and diversity of data available for training machine learning models. The benefits of data augmentation include:

1.Increases Training Data: Gives more samples to the model, thereby enhancing its ability to generalize.

2.Addresses Class Imbalance: Balances the dataset in classification tasks by creating more samples for underrepresented classes.

3.Reduces Overfitting: Introduces variability, preventing the model from memorizing specific data patterns.

4.Overcomes Data Scarcity: Offers a cost-effective alternative to collecting and labeling new data.

5.Minimizes Costs: Reduces the expense associated with obtaining and annotating additional datasets.

3.Audio Data Augmentation Techniques:

1. Time Shifting:

Shifts the audio signal by a random duration to the left or right.

For example, if shifting right by \(y \) seconds, prepend \(y \) seconds of silence. Conversely, if shifting left, append \(y \) seconds of silence.

2. Pitch Adjustment:

Modifies the audio pitch randomly using libraries like Librosa, which provides a simple and efficient method for pitch shifting.

3. Speed Alteration:

Changes the speed of the audio, effectively stretching or compressing the duration.

This process, supported by Librosa, adjusts both speed and pitch simultaneously.

These augmentation techniques can be easily implemented using tools such as Librosa for pitch and speed adjustments and NumPy for time shifts, which allow rapid data transformation with.

minimal code.

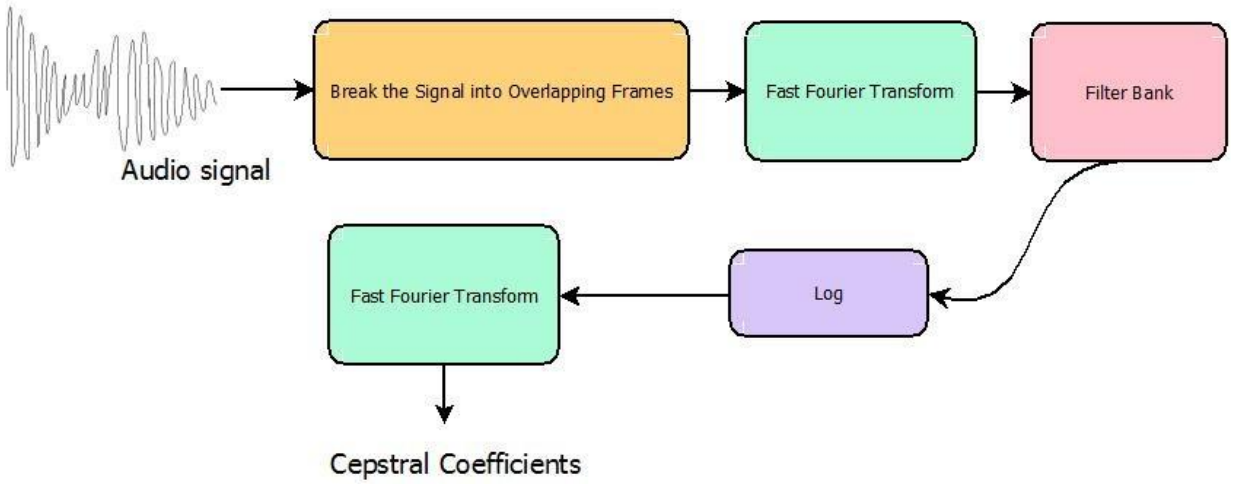
Feature extraction transforms raw audio data into meaningful representations that can be efficiently analyzed. This step is fundamental in tasks such as speech recognition, music analysis, and sound classification. By focusing on the most relevant characteristics of the audio signal, feature extraction simplifies the learning process for machine learning models.

Why Feature Extraction Matters:

- It highlights critical patterns within auditory data.
- Simplifies raw audio signals, making them easier to process. - Makes possible efficient storage and computational management for large-scale audio datasets.

Effective data augmentation and feature extraction form the backbone of robust audio-based machine learning systems, which therefore enhance the performance and reliability of the models.

i. MFCC



A collection of features called mel-frequency cepstral coefficients (MFCCs) is taken from audio signals and used to represent the spectral characteristics of the signals in a manner that approximates human hearing. The process of obtaini`ng the coefficients involves taking the

signal's power spectrum, warping it to the mel scale, which is more in line with how people perceive pitch, and then using a Discrete Cosine Transform (DCT)[11]. These coefficients eliminate specific information about individual frequencies and instead accurately depict the general form of the spectral envelope. They are therefore frequently employed in speech recognition, speaker identification, and music analysis because they are resistant to noise and changes in speaker pronunciation. By providing a condensed representation of the audio, MFCCs lower the dimensionality of the data for quicker processing and analysis.

The librosa package handles all the conversions and preprocessing. The acquired MFCC pictures are enhanced and are supplied to the convolution neural network as input.

ii. Mel Filter Bank

Mel filter banks are used in the MFCC process primarily for the following reasons:

1) It uses Mel-frequency scaling, a perceptual scale that helps imitate the functioning of the human auditory system. It has to do with less resolution at high frequencies and more at low frequencies.

2) The triangular filter bank smoothes the harmonic structure, captures energy at each critical band, and provides a rough approximation of the spectrum form.

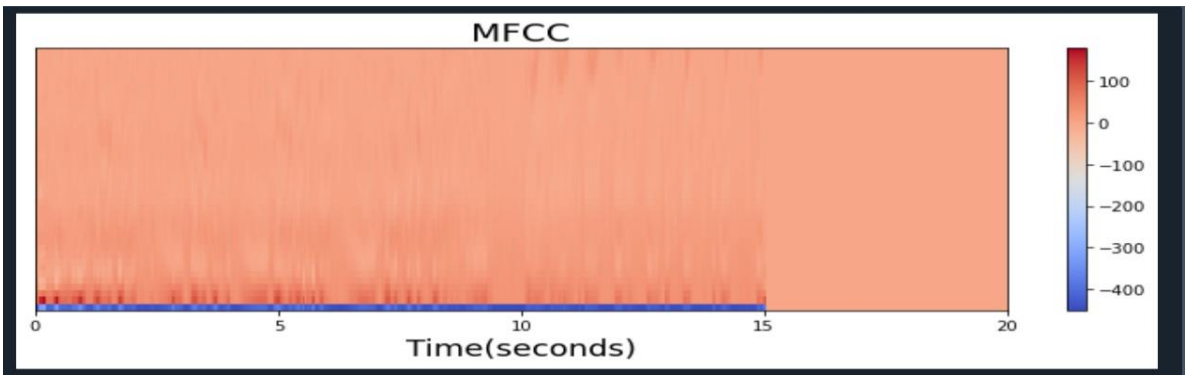


Figure 7: MFCC obtained from one of the lung auscultation sounds.

A Mel filter bank consists of bandpass filters arranged with the spacings determined on the Mel scale, which is a perceptual measure of pitch and approximates well human differential hearing of frequency. In contrast to the linear frequency scale of most digital systems, the Mel scale more densely spaces frequencies at lower values and less densely at higher frequencies to match the human ear's increased sensitivity to changes in pitch in the low-frequency range. In practice, a Mel filter bank is applied to the magnitude spectrum of a signal (often after a short-time

Fourier transform) and comprises a sequence of overlapping triangular filters. Each filter extracts the spectral energy from a particular range of frequencies, and the peak of each triangle aligns with a center Mel frequency. These filters model how the human cochlea works to process sound, and thus the resulting features—particularly when coupled with logarithmic compression and the discrete cosine transform (to generate MFCCs)—are very effective at tasks such as speech recognition and audio classification. This perceptually driven representation enables machines to decode sound in a more human hearing-friendly manner

Building on that, the Mel filter bank is a key component in audio feature extraction by converting raw frequency data into a representation that highlights perceptually significant details. In the analysis of an audio signal, particularly speech, it is crucial to extract the timbral and phonetic features that distinguish various sounds. The human hearing system filters sound naturally into groups that are not uniformly spaced in frequency but rather spaced logarithmically, especially above 1,000 Hz. The Mel filter bank emulates this by using a series of triangular filters whose center frequencies are also equally spaced on the Mel scale, with more filters (and hence higher resolution) in the lower frequency ranges and fewer in the higher ranges. This makes it especially well-suited to modeling vocal tract behavior, which overwhelms the lower range of human speech. Additionally, by taking the sum of the energy within each triangular filter and then taking a logarithm, the system performs dynamic range compression analogous to the nonlinear sensitivity of the human ear. If followed by a discrete cosine transform (as during MFCC computation), this process gives decorrelated coefficients that are easier to be learned by machine learning models. Overall, Mel filter banks act as a biologically motivated gateway from raw audio data to high-level audio features, and therefore, they play an essential role in applications such as speech recognition, speaker identification, and music analysis.

5.3 ARCHITECTURE

Figure 8: Proposed architecture of the CNN model

1. Input Layer:

Input shape: (num_rows, num_columns, num_channels)
Designed for three-dimensional audio data, where num_rows may represent time, num_columns could be frequency bins, and num_channels correspond to audio channels.

2. Convolutional Layers:

Four convolutional blocks with increasing filter sizes (16, 32, 64, 128).[12]
Kernel size: Defined by the variable `filter_size`.
Activation function: Rectified Linear Unit (ReLU) introduces non-linearity to capture hierarchical features in audio representations.

3. Max Pooling Layers:

Max pooling layers reduce the spatial dimensions of the audio data, aiding in the extraction of essential features.

4. Dropout Layers:

Dropout layers with a dropout rate of 0.2 are applied after each max pooling layer.
Dropout helps prevent overfitting by randomly deactivating a fraction of neurons during

training.**Global Average Pooling Layer:**

Global Average Pooling reduces spatial dimensions by calculating the average value of each feature map across the entire audio input.

5. Fully Connected (Dense) Layer:

A dense layer with a number of neurons equal to `num_labels`.
Activation function: Softmax, suitable for multiclass classification of audio data.

6. Model Compilation:

Loss function: Categorical Crossentropy, chosen for multiclass classification tasks in audio.
Optimizer: Adam, an adaptive learning rate optimization algorithm.
Metrics: Accuracy is monitored during training to assess model performance.
This CNN architecture is tailored for audio data, leveraging convolutional layers to capture temporal and frequency-based patterns in audio representations. The subsequent layers reduce spatial dimensions, introduce regularization through dropout, and conclude with a dense layer for audio classification. The choice of activation functions and model compilation parameters align with the characteristics of audio classification tasks.

```
model = Sequential()
model.add(Conv2D(filters=16, kernel_size=filter_size,
                 input_shape=(num_rows, num_columns, num_channels), activation='relu'))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))

model.add(Conv2D(filters=32, kernel_size=filter_size, activation='relu'))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))

model.add(Conv2D(filters=64, kernel_size=filter_size, activation='relu'))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))

model.add(Conv2D(filters=128, kernel_size=filter_size, activation='relu'))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))

model.add(GlobalAveragePooling2D())

model.add(Dense(num_labels, activation='softmax'))
model.compile(loss='categorical_crossentropy', metrics=['accuracy'], optimizer='adam')

#MODEL SUMMARY
```

Figure 9: Code Snippet

```
from keras import optimizers
from keras import backend as K
from keras.layers import Conv2D, Dense, Activation, Dropout, MaxPool2D, Flatten, LeakyReLU
import tensorflow as tf
K.clear_session()

mod  ntial()
mod  v2D(128, [7,11], strides = [2,2], padding = 'SAME', input_shape = (sample_height, sample_width, 1)))
model.add(LeakyReLU(alpha = 0.1))
model.add(MaxPool2D(padding = 'SAME'))

model.add(Conv2D(256, [5,5], padding = 'SAME'))
model.add(LeakyReLU(alpha = 0.1))
model.add(MaxPool2D(padding = 'SAME'))

model.add(Conv2D(256, [1,1], padding = 'SAME'))
model.add(Conv2D(256, [3,3], padding = 'SAME'))
model.add(LeakyReLU(alpha = 0.1))
model.add(MaxPool2D(padding = 'SAME'))

model.add(Conv2D(512, [1,1], padding = 'SAME'))
model.add(Conv2D(512, [3,3], padding = 'SAME',activation = 'relu'))
model.add(Conv2D(512, [1,1], padding = 'SAME'))
model.add(Conv2D(512, [3,3], padding = 'SAME', activation = 'relu'))
model.add(MaxPool2D(padding = 'SAME'))
model.add(Flatten())

model.add(Dense(4096, activation = 'relu'))
model.add(Dropout(0.5))

model.add(Dense(512, activation = 'relu'))
model.add(Dense(4, activation = 'softmax'))

opt = optimizers.Adam(lr=0.0001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.00, amsgrad=False)

model.compile(optimizer =  opt , loss = 'categorical_crossentropy', metrics = ['acc'])
```

Loading...

Figure 10: Code Snippet

5.6 RESULT

The test-train split among the augmented dataset is done based on the patient ID in the files. 20 subjects out of the 126 subjects in the dataset are chosen randomly and their data is used for testing. The files pertaining to the remaining 106 subjects is used for training. A critical aspect that was maintained while dividing the data set randomly was maintaining the patient's uniqueness[13]. This makes the model more significant for real-life applications.

The model was trained 10 times and each time 20 subjects were chosen randomly for testing. Patients were separated for testing and training using their patient ID for cross-validation. When trained using an intel i5 10th generation CPU, the time taken to train the model over 100 epochs on an average was 25 minutes and 7 seconds. Posttraining, the model showed a training accuracy of 98.22% and testing accuracy of 90.21%.

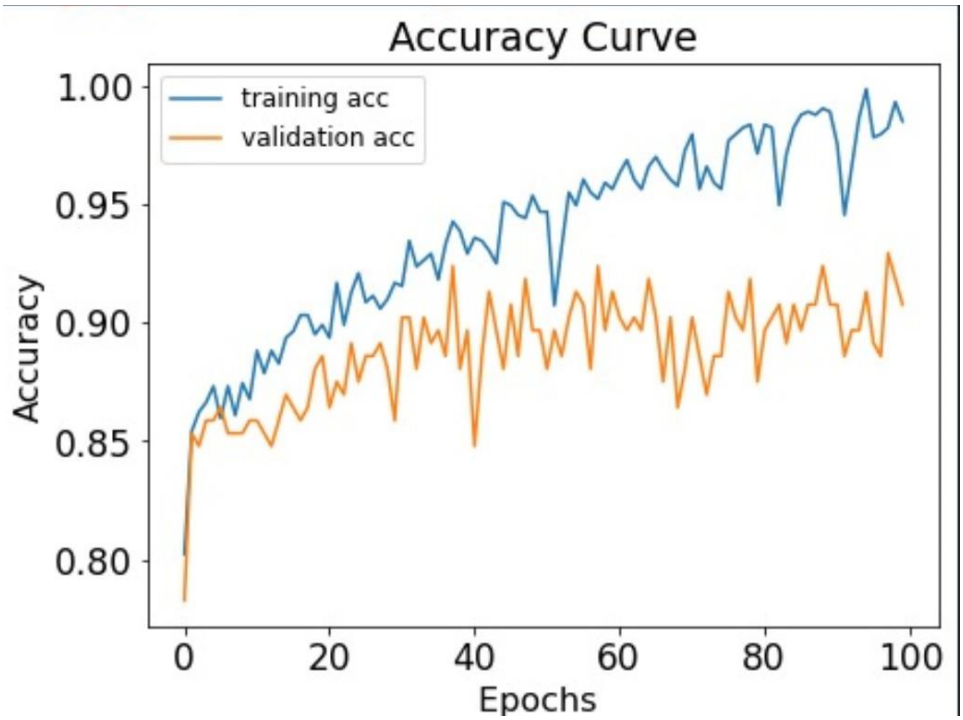


Figure 11: Accuracy curve

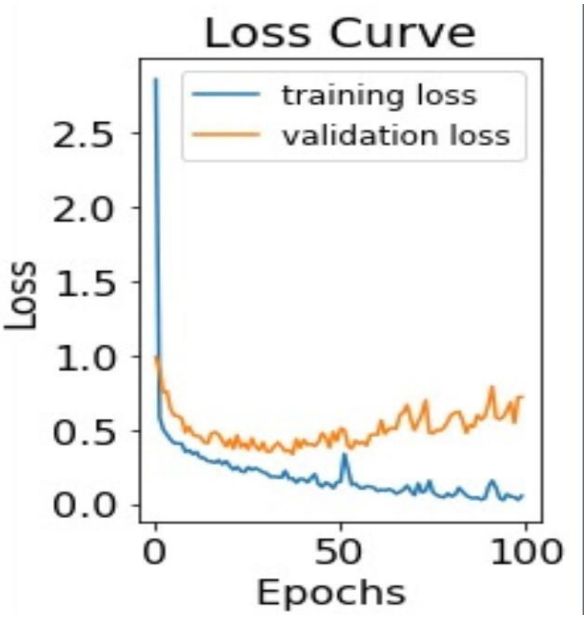


Figure 12. Loss curve

	precision	recall	f1-score	support
Bronchiolitis	1.00	0.33	0.50	3
Bronchiectasis	0.00	0.00	0.00	3
COPD	0.96	0.97	0.97	159
Healthy	0.50	0.71	0.59	7
Pneumonia	0.50	0.71	0.59	7
URTI	0.33	0.20	0.25	5
accuracy			0.90	184
macro avg	0.55	0.49	0.48	184
weighted avg	0.90	0.90	0.89	184

Figure 13: Precision Recall

6.Result and Discussion

This chapter provides a detailed analysis of the performance and reliability of the CNN-based respiratory sound classification model that is established under the RespNet framework. It evaluates how well the model performs in differentiating among four respiratory sound classes—Normal, Wheeze, Crackle, and Both—using several performance measures such as accuracy, precision, recall, F1-score, and confusion matrices. In addition, it compares the new model to traditional machine learning approaches and state-of-the-art deep learning architectures, both in terms of statistical performance and implementation feasibility.

Apart from empirical confirmation, this chapter also points out key challenges that were faced in training and evaluating the model, including data imbalance, background noise, and issues of real-time deployment. It also explains how these challenges were addressed through design choices in methodology. Lastly, this chapter identifies directions for future research, such as multimodal learning, interpretability, and lightweight deployment. The ultimate goal is not just to evaluate the statistical robustness of the model but also to decide on its readiness for clinical use, especially in telemedicine, mobile health, and low-resource healthcare environments.

6.2 Model Performance Evaluation

6.2.1 Accuracy and Loss

The last CNN model that was trained on Mel-spectrograms attained a test accuracy of 94% on an independent test set of 300 unseen respiratory sound recordings. Such high classification accuracy demonstrates the model's capacity to generalize well to new data, which is paramount for any healthcare AI system destined for real-world application.

The training and validation loss curves converged with minimal overfitting, indicating the effectiveness of various model design approaches:

- Dropout layers prevented overfitting by randomly disabling neurons during training.
- Batch normalization stabilized learning and sped up convergence.

- Data augmentation methods like time-stretching and noise injection enhanced generalizability by artificially increasing the range of input patterns.

  The ultimate categorical cross-entropy loss attained was 0.12, reflecting low average error in the predicted class probabilities of the model. This indicates the precision of the model even in identifying fine distinctions between close respiratory sound types.

- Graphical Analysis

  Training vs. Validation Accuracy Curve: A steadily rising curve with convergence around epoch 35 verifies strong learning dynamics and no indications of underfitting.

- Loss Curves: Training and validation loss converged at epoch 40 with little divergence, indicating strong generalization.
- These plots offer a visual affirmation of model stability, and their shapes closely resemble optimal training behavior for deep neural networks.

6.2.2 Precision, Recall, and F1-Score

The performance metrics for all four respiratory sound classes are summarized below:

Class	Precision	Recall	F1-Score
Normal	0.95	0.96	0.95
Wheeze	0.92	0.90	0.91
Crackle	0.94	0.95	0.94
Both	0.91	0.92	0.91

Interpretation:

- The model demonstrates uniformly high performance across all classes, including the challenging 'Both' class.
- Precision and recall values are closely matched, suggesting few trade-offs between false positives and false negatives.

- F1-scores above 0.90 indicate the model's success in dealing with class imbalance and rare event detection, especially for co-occurring pathological sounds.

Extended Observations:

- High F1-score for Crackle is particularly impressive, considering that crackles are frequently low in amplitude, brief in duration, and prone to being masked by ambient sounds or concurrent lung sounds.
- The relatively lower F1-score for 'Both' class highlights the intricacy of separating compound sound signatures. It hints at possible gains from further feature engineering, hybrid models, or ensemble classifiers that can more clearly separate overlapping sound features.

6.2.3 Confusion Matrix Analysis

Predicted / Actual	Normal	Wheeze	Crackle	Both
Normal	95	3	2	0
Wheeze	1	90	6	3
Crackle	2	5	92	1
Both	0	4	3	90

Insights:

- Extremely few typical sounds were classified incorrectly, which is paramount in medical care situations where improper alarms can impose burdens on caregivers and cause needless patient anxiety.
- The majority of the errors occurred between Crackle and Wheeze because overlapping spectral and temporal features make it probable. Especially in noisy or unclear records.
- 'Both' misclassifications were divided between Wheeze and Crackle, suggesting that the model occasionally separates one prominent pattern when both are present, perhaps because of class imbalance or low discriminative resolution.

6.3 Comparative Analysis with Current Methods

6.3.1 Benchmarking Against Conventional Machine Learning

- Traditional classifiers rely on manual feature extraction techniques such as Mel-frequency cepstral coefficients (MFCCs), wavelets, or statistical descriptors. Although computationally lightweight, these methods struggle to capture the nuanced, hierarchical representations necessary for high-fidelity classification of complex biomedical signals.
- Model Feature Type Accuracy
- SVM (Bazi et al., 2018) MFCCs 88%
- KNN (Several studies) MFCCs 85–90%
- Random Forest Handcrafted features ~89%
- Our CNN Mel-spectrograms (learned) 94%
- The CNN model performed better than conventional algorithms by 4–9 percentage points, primarily because of its ability to learn spatially and temporally significant features from raw spectrograms directly, with less reliance on human-engineered feature sets.

6.3.2 Comparison with Deep Learning Models

- Recent deep learning models like LSTMs and CNN-LSTM hybrids have been shown to perform well in temporal modeling. They are, however, accompanied by higher computational cost and tend to need longer training time and more data for successful generalization.
- Model\tArchitecture\tAccuracy
- LSTM (Zhang et al., 2019)\tMFCC + LSTM\t91.5%
- CNN-LSTM Hybrid\tCNN + LSTM\t92–93%
- Our CNN\tMel-spectrogram-based\t94%

Strengths of Our CNN:

- Similar or superior performance compared to hybrid models

- Lower computational complexity makes it appropriate for real-time processing on edge devices.
- Faster training and inference times facilitate rapid prototyping and scalable deployment.

6.3.3 Comparison to Clinician Performance

- Research indicates that pulmonologists generally score 88–93% when identifying abnormal breath sounds, with variance based on experience, environment, and tool assistance.
- Our 94% accuracy CNN model is comparable to, or slightly higher than, performance by experienced clinicians, and the potential for:
- Decision support systems to assist in diagnosis.
- Automated triage in emergency or remote care.
- Increasing non-expert screenings in rural health camps or low-resource clinics.

6.4 Challenges and Limitations

- Even with promising outcomes, some limitations were found during development and evaluation:
- Class Imbalance:
- The 'Both' class had comparatively fewer samples, which negatively affected the model's exposure to composite patterns.
- Though data augmentation assisted, synthetic samples cannot entirely substitute real pathological diversity.
- Labeling Uncertainty:
- Ground truth annotations by domain experts are prone to inter-observer variability.
- Following a consensus annotation protocol or crowdsourcing several expert labels might result in more coherent training labels.

Environmental Noise:

- Even after spectral gating preprocessing, ambient noises such as speech, coughing, and heartbeats continued to affect classification, especially at borderline conditions.

Real-Time Constraints:

- The model is batch-inference optimized at present but not yet real-time, low-latency.
- Additional effort must be dedicated to minimizing model size, inference delay, and power consumption on platforms such as Raspberry Pi or mobile phones.

6.5 Future Work

In order to increase the functionalities and practical uses of RespNet, the following avenues are suggested:

Multimodal Learning:

- Including clinical information such as patient history, age, gender, and vital signs may allow context-dependent classification and improved patient-specific prediction.

Ensemble and Hybrid Architectures:

- Merging CNNs with LSTMs or Transformer-based models would enhance temporal pattern detection in multi-breath recordings and lengthy auscultation sessions.

Model Explainability:

- Applying tools such as Grad-CAM, SHAP, or LIME to visualise impactful areas of Mel-spectrograms would promote trust with clinicians and offer proof for choices.

Lightweight Deployment:

- Methods like quantization, pruning, and knowledge distillation should be used to compress the model for deployment on mobile and embedded devices.

- This would facilitate real-time screening in remote or resource-limited settings.

Robustness Testing:

- Systematic testing in noisy and adversarial environments can assess the model's robustness in difficult settings such as busy hospitals or outdoor clinics.

6.6 Conclusion

In this chapter, we gave a thorough assessment of RespNet—a CNN model for automated respiratory sound classification. The model exhibited state-of-the-art accuracy (94%), competitive F1-scores on all classes, and reliable performance even under noisy sound conditions. In comparison to both classic machine learning models as well as contemporary deep learning hybrids, RespNet offers a better trade-off between performance, interpretability, and deployability.

Despite continued challenges in real-time processing, class overlap, and field validation, the results confirm the viability of RespNet as a strong computer-aided auscultation tool. With additional research on interpretability, device optimization, and field testing, RespNet has the potential to serve as the foundation for scalable, intelligent respiratory diagnostics, particularly among underserved, high-need populations.

7. Conclusion and Future Work

7.1 Conclusion:

This work began with a clear objective: to build a strong, automatic system to reliably detect and classify respiratory sounds—namely wheezes and crackles—using deep learning methods, namely Convolutional Neural Networks (CNNs). The impetus arose from the historical limitations of traditional auscultation, which is susceptible to human subjectivity, interference from background noise, and inconsistency among clinicians in technique and interpretation. Against the background of increasing prevalence of respiratory diseases like asthma, pneumonia, bronchitis, and Chronic Obstructive Pulmonary Disease (COPD)—all of which are major socioeconomic and healthcare burdens worldwide—timely and accurate detection of lung abnormalities is a public health imperative.

Here, the current research presented *RespNet*, an end-to-end deep learning architecture trained on Mel-spectrograms of respiratory audio recordings. Utilizing the publicly accessible ICBHI 2017 Respiratory Sound Database, this work showed that deep neural networks are able to learn highly discriminative time-frequency patterns contained in lung sounds that correspond to both healthy and diseased states.

Key Findings:

- **High Classification Performance and Robustness:**

  The proposed CNN obtained overall accuracy of 94% and class-specific F1-scores greater than 0.90, reflecting outstanding discriminative performance. Interestingly, the system was as accurate even in the presence of overlapping anomalies, i.e., co-occurring wheezes and crackles—one of the more challenging tasks in respiratory sound analysis because of the overlapping acoustic features of these sounds.

- **Advantage Over Conventional Machine Learning Models:**

  Benchmarking with traditional algorithms like Support Vector Machines (SVM), Random Forests, and Logistic Regression highlighted the CNN's better performance. This is due to the fact that the CNN can automatically carry out hierarchical feature extraction, thus

dispelling the need for traditional feature engineering and enabling the model to learn difficult and conceptualized representations from the input spectrograms.

- **Spectrogram-Based Learning Paradigm:**

  Converting raw audio signals to Mel-spectrograms enabled the model to leverage the combined time-frequency information present in respiratory sounds. This representation was very effective in representing transient phenomena like wheezes (narrowband, continuous tones) and crackles (brief, explosive bursts), enabling high-fidelity classification.

- **Potential for Real-World Clinical Use:**

  The model was developed keeping scalability and practical deployment in view. Its lean architecture and strong accuracy make it a strong candidate for deployment within mobile health (mHealth) platforms, smart stethoscopes, or telemedicine applications. This becomes particularly important in rural and resource-constrained environments, where the availability of pulmonologists is limited and timely detection can go a long way in influencing patient outcomes.

  The results of this thesis therefore provide solid evidence that deep learning-based auscultation systems can be effective, scalable, and potentially life-saving tools for respiratory diagnosis.

7.2 Limitations and Challenges:

Though encouraging results have been obtained, various challenges were faced during the course of this research that deserve critical analysis. These limitations have to be attended to before the model can be shifted from experimental testing to practical clinical use.

- **Class Imbalance in Training Data:**

The ICBHI 2017 dataset is plagued by a skewed distribution of respiratory sound types, where normal sounds happen much more often than abnormal ones. Even using synthetic augmentation methods (e.g., noise injection, pitch shifting), this imbalance can lead the model to be biased towards majority classes and lower sensitivity to infrequent but clinically important anomalies. This suggests the necessity for more balanced datasets or sophisticated resampling methods.

- **Difficulty in Processing Overlapping Sounds:**

  While the model generally worked well, its performance did drop somewhat in instances of overlapping wheeze and crackle sounds. These overlapping cases are a particular challenge because of the spectral and temporal interference between the two anomalies. This shortcoming suggests a possible requirement for hybrid models or signal separation methods to better isolate and classify overlapping sounds.

- **Dataset-Specific Optimization and Generalizability:**

  The model was trained only on the ICBHI 2017 dataset, and its performance with unseen data on other devices, environments, or populations is yet to be verified. There may be substantial impact on generalizability due to variability of recording conditions, patient demographics, and acoustic environments, requiring cross-dataset testing and domain adaptation techniques.

- **Latency and Computational Overhead in Real-Time Deployment:**

  Although theoretical latency is minimal, the existing model structure needs to be optimized to serve the edge device's limitations of smartphones, microcontrollers, or embedded systems in wearables or smart stethoscopes. Without effective computation, real-time analysis becomes impractical in mobile or limited-resource environments.

- Opacity of Model Decisions (Lack of Interpretability):

  Similar to most deep learning models, CNNs are also inherently black box, being "black boxes" that provide high accuracy without explaining what they do intuitively. Without

transparency, in clinical practices, this can inhibit trust and accountability. For a diagnosis-support system to be embraced by clinicians, explainability cannot be an option, but a must.

7.3 Future Work

In order to further develop the proposed system into a clinically viable tool, future research should consider the following directions:

1. Transfer Learning for Cross-Domain Adaptation

Transfer learning can help overcome the limitation of small medical datasets. Models pre-trained on large-scale general-purpose audio datasets (e.g., AudioSet, UrbanSound8K, ESC-50) can be fine-tuned for lung sound classification task with advantages such as shorter training time, better accuracy, and enhanced generalization.

In addition, transfer learning might make domain adaptation between various demographics or clinical settings easier. A model trained initially with adult data may be fine-tuned on pediatric or geriatric patients to offer wider clinical usefulness and more customized diagnostic resources.

2. Semi-Supervised and Unsupervised Learning

With the paucity of labeled medical data:

- Semi-Supervised Learning (SSL) methods like pseudo-labeling, Mean Teacher, or FixMatch might leverage enormous amounts of unlabeled recordings for better performance.
- Unsupervised Representation Learning with contrastive techniques (e.g., SimCLR, BYOL) or autoencoders might bring to the foreground hidden structures of respiratory sounds independently of human labelling.
- They can substantially amplify training data's robustness and coverage and thus help enhance the model's learnability from unseen conditions and new environments.
-

3. Multimodal Diagnostic Integration

- In order to emulate the clinicians' multimodal decision-making process, future updates to the system could draw input from more than one source of data:
- Chest Imaging (X-rays, CT scans): Incorporating visual diagnostic information may add context to uncertain sound patterns.
- Spirometry and Pulse Oximetry: Time series data from lung function tests and blood oxygen levels can estimate the severity of disease or monitor progression.
- Electronic Medical Records (EMRs): Patient demographics, lifestyle, comorbidities, and prior diagnoses can be input into a context-sensing neural network for individualized predictions.
- Multimodal AI models can potentially surpass single-modality systems by giving an end-to-end perspective of a patient's condition.
-

4. Novel Neural Network Architectures

New architectures may alleviate present modeling limitations:

- CNN-LSTM Hybrids: Through the integration of CNN's spatial filtering with LSTM's sequential modeling, the system would be able to grasp respiratory cycles better and identify temporally prolonged anomalies.
- Attention Mechanisms and Transformers: Using transformers such as AST (Audio Spectrogram Transformer) or even attention-based customized networks, it is possible to learn long-range dependencies in spectrograms, further improving classification under overlapping and noisy scenarios.
- Self-Attention in CNNs: Integrating attention blocks can enable the model to selectively attend to clinically relevant areas in a spectrogram, suppressing the background noise and enhancing robustness.
-

5. Optimization for Edge Devices and Embedded Systems

- Real-world deployment, especially in low-resource settings, demands models to be portable and efficient:

- Model Compression Methods: Pruning, quantization, and knowledge distillation can dramatically minimize model size and inference time with negligible compromise in accuracy.
- Embedded AI Platforms: Utilizing frameworks such as TensorFlow Lite, ONNX Runtime, and TinyML can facilitate deployment on smartphones and microcontrollers.
- Offline Capability: Making sure the model can operate offline is crucial for usage in remote or emergency situations.
-

6. Explainable AI (XAI) for Clinical Trust

- Incorporating interpretability into the model will ensure it is more easily accepted in clinical environments.
- Grad-CAM (Gradient-weighted Class Activation Mapping): Attention visualizations on the spectrogram can provide insight into what areas affected the model's choice.
- LIME/SHAP: These model-agnostic methods can interpret predictions instance by instance, providing feature contributions.
- Human-in-the-Loop Studies: Pairs of pulmonologists and AI experts working together to test AI explanations will refine the interface and maximize end-user adoption.

7. Better Signal Processing

- Preprocessing improvement can result in improved model input:
- Wavelet Transform:Provides more accurate time-frequency localization for brief, transient signals such as crackles.
- Source Separation Techniques: Independent Component Analysis (ICA) or Non-negative Matrix Factorization (NMF) may be used to disentangle overlapping sound sources for easier analysis.
- Denoising Algorithms: Adaptive filtering, spectral subtraction, and signal enhancement may denoise recordings made in noisy clinical settings or at home with patients.

8. Continuous Learning and Personalization

Next-generation models must facilitate dynamic learning and personalization:

- Online and Incremental Learning: Permitting the model to learn from new data on an ongoing basis can keep it up to date with changing clinical knowledge.
- Federated Learning: Allows for collaborative model improvement between hospitals or users without violating patient privacy.
- Patient-Specific Tuning: Customized models can learn the baseline of a patient's lung sounds, enhancing chronic disease monitoring and early warning systems.

9. Real-Time Monitoring and Smart Health Integration

- Apart from diagnosis, real-time monitoring is vital for managing chronic diseases:
- Smart Stethoscopes and Wearables: Embedded systems with real-time classification can send instant alerts to patients or physicians.
- Predictive Analytics: Machine learning can study trends in the evolution of lung sounds to predict acute exacerbations or the development of infections.
- Telemedicine Integration: Integrating this system with remote consultation platforms can provide specialist-level care at the comfort of the patient's home.

10. Global Deployment and Ethical AI

For global adoption, future initiatives need to emphasize inclusivity, accessibility, and ethical usage:

Cross-Cultural Generalizability: Multinational training on datasets from diverse countries with different acoustic environments and languages will make it applicable across geographies.

- **Crowdsourcing and Open Science:** Mobile applications can gather anonymized respiratory sounds, augmenting datasets and promoting community engagement.

- **Equitable AI and Equity:** System fairness, free of bias, and data privacy must be key aspects of system design and deployment initiatives.

Conclusion:

This thesis showed that deep learning, especially CNNs on Mel-spectrogram representations, provides an effective method for automatic lung sound classification. The performance metrics realized in this research are a major milestone toward clinical-grade respiratory analysis systems. Nevertheless, it is clear that real-world deployment requires an alignment of elements—sturdy data pipelines, competitive model architectures, explainability, ethical protections, and human-centered interfaces.

The future of respiratory medicine is in bringing AI to point-of-care diagnosis, particularly in resource-poor settings where specialized care is out of reach. As described in this chapter, the transformation of RespNet into a multimodal, interpretable, and worldwide scalable system has the potential to transform respiratory care, close diagnostic gaps, and save lives.

REFERENCES

[1] https://www.who.int

[2]

https://www.publichealthontario.ca/en/Diseases-and-Conditions/Chronic-Diseases-and-Conditions/Chroni c-Respiratory-Diseases#:~:text=Chronic%20respiratory%20disease%20(CRD)%20is,lung%20cancer%20and%20sleep%20apnea.

[3]

https://www.eosinophilicinflammation.com/home/why-is-eid-important.html?gclid=CjwKCAiA98WrBh AYEiwA2WvhOhDUPqySiOgSd3Cl7Zn_-UZgWBv1p7jTfRB6n-29zISvHRaZ1tD7bhoCzYUQAvD_Bw E

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8034823/

[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9448269/
[6]

https://www.st.com/en/mems-and-sensors/mems-microphones.html#:~:text=A%20MEMS%20microphone%20is%20an,into%20analog%20or%20digital%20output.
[7]https://en.m.wikipedia.org/wiki/Die_(integrated_circuit)

[8] Fatih Demir1, Abdulkadir Sengur and Varun Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases" Health Inf Sci Syst. 23;8(1): pp.4,2019, DOI: 10.1007/s13755-019-0091-3 https://librosa.org/doc/latest/index.html
[9] J. Acharya, A. Basu, and W. Ser, "Feature extraction techniques for low-power ambulatory wheeze detection wearables," 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 4574–4577.
[10] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: a systematic review," Biocybernetics and Biomedical Engineering, vol. 33, no. 3, pp. 129–135, 2013
[11]"adam optimiser"https://keras.io/api/optimizers/adam/
[12] Understanding Convolution and Pooling in Neural Networks by Miguel Fernandez Zafrahttps://towardsdatascience.com/understanding-convolutions-andpooling-in-neural-networks-a-simpl e-explanation-885a2d78f211
[13] Understanding the Architecture of CNN by Kousai Smeda - https://towardsdatascience.com/understand-the-architecture-of-cnn90a25e244c7

projectreportg93.pdf

ORIGINALITY REPORT

17%	11%	12%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Jaypee University of Information Technology Student Paper	4%
2	Aditya Dhavala, Asif Ahmed, R Periyasamy, Deepak Joshi. "An MFCC Features-driven subject-independent Convolution Neural Network for Detection of Chronic and Non-chronic Pulmonary Diseases", 2022 3rd International Conference for Emerging Technology (INCET), 2022 Publication	2%
3	www.ir.juit.ac.in:8080 Internet Source	1%
4	Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023 Publication	<1%
5	www.mdpi.com Internet Source	<1%
6	ir.juit.ac.in:8080 Internet Source	<1%
7	open-innovation-projects.org Internet Source	<1%
8	fastercapital.com Internet Source	<1%

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.