university of groningen

campus fryslân

Enhanced Disease Classification in Respiratory Sounds: A Transfer Learning Approach Utilizing ICBHI and Coswara Datasets

Soogyeong Shin

university of groningen

campus fryslân

University of Groningen

Enhanced Disease Classification in Respiratory Sounds: A Transfer Learning Approach Utilizing ICBHI and Coswara Datasets

Master's Thesis

To fulfill the requirements for the degree of
Master of Science in Voice Technology
at University of Groningen under the supervision of
Prof. Dr. Vass Verkhodanova (Voice Technology, University of Groningen)

Soogyeong Shin (s5661285)

June 10, 2024

Contents

Page

List of Figures

Acknowledgments

First of all, I would like to appreciate my supervisor, Vass Verkhodanova, for providing invaluable guidance and bringing the weight of her considerable experience and knowledge to this project. Her high standards and kind advice have made me better at what I do.

Thanks also to Professor Shekhar Nayak for his technical advice on model training in this project, which has greatly contributed to the development of my research. I learned a lot of machine learning knowledge from his course.

I would also like to thank Professor Matt Coler for helping me shape the design of my thesis. His insights, particularly gained from his Thesis Design course in Term 2A, have been instrumental in refining my approach.

I extend my gratitude to the creators and maintainers of the ICBHI 2017 dataset and the Coswara dataset for providing valuable data that formed the foundation of my research.

Lastly, I acknowledge the support of the University of Groningen Voice Technology Team, where I am currently enrolled, for providing resources and a conducive environment for my research.

Abstract

The early and accurate detection of respiratory diseases, such as asthma, chronic obstructive pulmonary disease (COPD), and pneumonia, is critical for improving patient outcomes. This need has been emphasized after the COVID-19 pandemic. Traditional diagnostic methods, such as auscultation, depend on the experience of clinicians and are constrained by their subjective assessment. To address these limitations, this study seeks to increase the accuracy and efficiency of respiratory disease diagnosis using transfer learning through data augmentation. This research utilized publicly accessible respiratory sound datasets, ICBHI 2017 and Coswara, aiming to overcome the constraints of smaller datasets. A transfer learning strategy employing a Residual Networks (ResNet)-based model was implemented to enhance the accuracy of respiratory disease diagnosis. The models, initially trained on the ICBHI dataset, were fine-tuned with data from Coswara, enabling them to detect abnormal lung sounds variations associated with specific respiratory conditions. The results indicate that the AI-enhanced model achieved accuracy peaks of 95.80% and 93% with the first and second fine-tuning strategies, respectively. These findings demonstrate the successful integration of two respiratory sound datasets through transfer learning. This study not only highlights the effectiveness of advanced AI techniques in medical diagnosis, but also highlights the importance of dataset augmentation to ensure robust model performance. Integrating multiple datasets through transfer learning can be an effective strategy not only for respiratory diseases but also for developing diagnostic tools for a variety of conditions with small datasets.

Keywords: Respiratory Diseases, Auscultation, Transfer Learning, Respiratory Sound Analysis, Automated Diagnosis, ResNet, ICBHI Dataset and Coswara Dataset

1 Introduction

Respiratory diseases like asthma, chronic obstructive pulmonary disease (COPD), lower respiratory tract infection, lung cancer, and tuberculosis are the leading causes of death worldwide, constituting four of the 12 most common causes of death [1]. Identifying respiratory diseases early is essential for reducing their spread and minimizing their negative impact on life expectancy and quality. Moreover, following the COVID-19 pandemic, there has been a heightened need for rapid and precise diagnostic technologies. The conventional approach for detecting lung diseases is auscultation, which utilizes a stethoscope to listen to the respiratory sounds of patients in person[2]. Auscultation is critical in diagnosing lung-related diseases. Although it's a trusted technique, its success relies heavily on the medical professional's expertise and human hearing capabilities. Additionally, it takes a considerable amount of time for the doctor to auscultate each patient[3]. Since more people are interested in technology like telemedicine after COVID-19, using AI for auscultation, which normally requires being in person, is getting a lot of attention. AI auscultation has strong potential to revolutionize early detection and rapid diagnosis of respiratory diseases by analyzing breath sounds. It can save time for doctors during patient diagnosis and enhance consistency among varying diagnostic opinions that arise from traditional auscultation methods. It can also help to monitor patients outside a clinic by less-skilled workforce such as community health workers.[4]. However machine learning algorithms, such as those proposed by Sebers et al, often require expert knowledge and significant time to extract features suitable for classification models[5]. Recently, deep learning methodologies have been introduced to minimize human involvement in the feature extraction process from respiratory sounds, allowing the system to learn features autonomously[6].The motivation for this study is to utilize AI-enhanced auscultation systems, combining augmented data from small datasets, to improve the accuracy and accessibility of respiratory disease diagnostics.

This study aims to create an AI-enhanced auscultation system by augmenting small-scale respiratory sound datasets. A representative open access respiratory sound dataset is the ICBHI 2017 dataset[7]. A lot of research have been done using this dataset, and their models distinguish between a variety of breath sounds, from normal breath sounds to unusual sounds such as harsh crackles and wheezes, and even sounds that are generally difficult to interpret. However, there were limitations due to the small size of the dataset. In particular, many researchers have augmented datasets by dividing breathing sounds into multiple cycles[4][8], but there are difficulties in translating these results into the diagnosis of diseases such as asthma and pneumonia. In this study, I propose a respiratory diagnostic model trained by data augmentation combining the ICBHI dataset and the Coswara dataset[9], also known as the Covid dataset[9]. This combined dataset was used first for training and then used for fine-tuning to increase model's diagnostic accuracy. Using the trained residual networks (ResNet) model with these datasets, the diagnostic accuracy for respiratory diseases was compared.

1.1 Research Questions and Hypotheses

This study seeks to enhance the accuracy and efficiency of respiratory disease diagnosis using transfer learning through data augmentation. I use the ICBHI dataset to develop a pre-trained model sensitive to respiratory sounds. To enhance diagnostic classification, I fine-tune this model adding the Coswara dataset. In particular, to demonstrate the effectiveness of data augmentation, I examine how different strategies for training the model, such as freezing specific layers during transfer learning or modifying classifiers when incorporating additional datasets, affect the model's diagnostic capabilities. To summarize, this thesis focuses on the following research question:

Q1. Do different freezing strategies (freezing up to the 1st, 2nd, and 3rd convolutional base) impact the accuracy and loss metrics of models when conducting transfer learning with two respiratory sound datasets?

Q2. When fine-tuning with the different dataset, which strategy shows better performance: freezing strategies or changing the classifier?

H1. Freezing different convolutional base (up to the 1st, 2nd, and 3rd) during transfer learning with two respiratory sound datasets will significantly influence the model's accuracy and loss metrics. Freezing earlier convolutional base is hypothesized to preserve essential low-level features, resulting in lower loss and higher accuracy compared to less or no freezing.

H2. When fine-tuning with different datasets, the strategy of changing the classifier is expected to demonstrate better performance in terms of accuracy and loss reduction compared to freezing strategies. This is because it is assumed that the Coswara and ICBHI datasets contain sufficiently similar samples and that utilizing different dataset will address the issue of class imbalance.

1.2 Thesis Outline

Chapter 1 introduced the motivation behind employing AI in the diagnosis of respiratory diseases, focusing on the limitations of traditional methods and the benefits of AI-enhanced diagnostic systems. In Chapter 2, a comprehensive background is provided on respiratory sounds and diseases. This includes an in-depth look at normal and abnormal respiratory sounds, such as crackles, wheezes, and rhonchi, and their associations with various respiratory conditions. Chapter 3 reviews existing research utilizing the ICBHI and Coswara datasets, providing a critical review of previous methodologies and their contributions to the field. The methodologies employed in the research are outlined in Chapter 4, detailing the datasets used, data preprocessing techniques, and the specifics of the transfer learning approach implemented. This chapter describes the step-by-step process of training the ResNet model, from initial data handling to the fine-tuning stage. Chapter 5 presents the results and analysis based on the described methods in chapter 4. In chapter 6, the research questions and hypothesis are reviewed, comparing previous literature reviews. Chapter 7 provides conclusion and propose follow up research. Finally, chapter 8 discusses the ethical considerations of using anonymized datasets and deploying AI in healthcare, focusing on data security and transparency.

2 Background

Auscultation is non-invasive, real-time, inexpensive, and very informative [10]. Recent electronic stethoscopes made it possible to record lung sounds, and it facilitated the studies of automatically analyzing lung sounds[11]. In this chapter, the types of abnormal lung sounds and respiratory diseases are reviewed.

2.1 Respiratory Sounds

Following the definition of classification standards for respiratory sounds at the 10th International Conference on Lung Sound Analysis (ILSA), the stepwise classification of respiratory sounds has become a central topic in the examination of respiratory sounds. Respiratory sounds are categorized into two main types: normal and abnormal. This classification aids in the structured analysis and diagnosis of respiratory conditions.

Normal respiratory sounds are the sounds heard when a patient has no respiratory issues. These sounds are typically described by capturing the movement of air through the respiratory tract. Common respiratory sounds, such as those from the trachea, are characterized by a broad range of noises, including high-frequency components that can be heard during both the inhalation and exhalation phases. Subtypes like ”tracheal,” ”bronchial,” and ”vesicular” sounds refer to the specific areas of the respiratory tract where these sounds are most prominent[12].

On the other hand, the second category of respiratory sounds is abnormal sounds. These differ from the first category based on their natural and distinctive patterns of action, appearing when a patient has respiratory issues or disorders. These sounds indicate abnormalities in the respiratory function, manifesting as distinct auditory cues during the respiratory cycle[13].

2.2 Abnormal Respiratory Sounds

Abnormal sounds are unwanted respiratory noises that are superimposed on the normal breathing sounds. These sounds are generally lower in intensity and force. They are classified based on several factors that aid in detecting each class individually. Adventitious sounds are abnormal sounds that are heard over a patient’s lungs and airways. They are broadly categorized into continuous and discontinuous sounds, distinguished by the duration of occurrence during respiration [14]. Continuous adventitious sounds (CAS), belonging to the category of abnormal sounds, typically last around 250ms, although this is not universally applicable to all CAS types. Abnormal lung sounds include crackles, wheezes, rhonchi, and pleural friction rubs. Based on pitch[1], these sounds are further classified into high-pitched (such as wheezes and stridor) and low-pitched (such as rhonchi and squawks). Discontinuous adventitious sounds (DAS) lasts less than 25ms and is further categorized into fine crackles, coarse crackles, and pleural rubs. Among these, the most commonly detected sounds are crackles, wheezes, and rhonchi, which are crucial for diagnosing pulmonary diseases[14][3].

2.2.1 Crackle

Crackles, characterized as brief, abrupt, and non-tonal sounds, are typically associated with diseases affecting the lung parenchyma such as pneumonia, interstitial pulmonary fibrosis (IPF), and pulmonary edema. These sounds are generated by air passing through fluid or mucus in the larger

[1] In this context, pitch refers to the categorization of sound frequencies, distinguishing higher from lower frequency respiratory sounds.

bronchi, creating a coarse crackling noise. These coarse crackles are usually discernible during the early stages of inhalation and noticeable upon exhalation. They exhibit a low pitch, around 350 Hz, and are relatively brief, lasting about 15 ms. Conditions such as chronic bronchitis, bronchiectasis, and chronic obstructive pulmonary disease (COPD) often present with these coarse crackle sounds[15].

2.2.2 Wheeze

Wheezes are described as high-pitched sounds often related to conditions like asthma and chronic obstructive pulmonary disease (COPD)[10]. These sounds are characterized by their sharp, sustained, and rhythmic nature, typically exhibiting a frequency starting at 400 Hz. Wheezes result from the constriction of the airways, leading to restricted airflow. The duration of wheeze sounds can vary, with some instances reported as short as 80 to 100 ms[16]. Common conditions linked with wheezing include asthma and COPD. When wheezing is localized, it might indicate an obstruction due to a foreign object such as a tumor in the airway[17].

2.2.3 Rhonchi

Rhonchi are characterized as deep, resonant sounds that are similar in quality to snoring, typically signaling the presence of mucus in the larger airways, which may be alleviated through coughing. These sounds are identified as continuous adventitious sounds, audible during inhalation but predominantly during exhalation, or throughout the entire respiratory cycle[13]. They often manifest with a dominant frequency around 200 Hz and durations approximately between 80 and 100 ms, persisting continuously. The presence of rhonchi in both inhalation and exhalation phases is typically due to the narrowing of airways in larger bronchial sections[18].

Table 1: Characteristics and Acoustic Features of Abnormal Respiratory Sounds

	Location	**Characteristics**	**Acoustic features**	**Related diseases**
Crackle	Peripheral lung	Discontinuous Low-pitched	Rapidly dampened wave deflection Frequency about 350Hz Longer duration (about 15ms)	Interstitial lung fibrosis Pneumonia Congestive heart failure
Wheeze	Bronchi	Continuous High-pitched	Sinusoid Frequency > 100-5000 Hz Duration >80ms	Asthma COPD Tumor
Rhonchi	Bronchi	Continuous Low-pitched	Sinusoid Frequency about 150 Hz Duration >80ms	Bronchitis Pneumonia

There are many other abnormal lung sounds, but only the three most common are discussed in this paper. Auscultation is a useful diagnostic method for identifying these sounds. However, the results can vary significantly based on the clinician's experience[3]. According to Melbye et al., there is notable variation among observers when distinguishing between expiratory wheezes and low-pitched wheezes, which can affect diagnosis and treatment[19]. These limitations have highlighted the need for a standardized system to accurately classify breath sounds using artificial intelligence. AI-assisted auscultation can aid in the correct diagnosis of respiratory diseases and in identifying patients who require urgent care. It is also useful for screening and monitoring patients with various lung conditions, such as asthma, COPD, and pneumonia[20].

2.3 Respiratory Diseases

Respiratory diseases significantly impact human health and are often diagnosed with the aid of abnormal respiratory sounds, as discussed in previous sections[14]. This study focuses on common respiratory diseases such as asthma and pneumonia, detailing their characteristics, associated sounds, and characteristics of the respiratory sounds associated with each condition.

2.3.1 Asthma

Asthma is a chronic inflammatory disease of the airways characterized by variable and recurring symptoms, reversible airflow obstruction, and bronchospasm[21]. It is characterized by a variety of respiratory sounds, primarily wheezes. These are high-pitched, continuous sounds that are most prominent during exhalation but can also be heard during inhalation in severe cases. Wheezes in asthma are caused by the narrowing of the bronchial tubes due to inflammation, leading to a whistling sound[22]. The pitch and duration of wheezing can vary depending on the degree of airway obstruction. During an asthma attack, these sounds become more pronounced and are a key indicator in diagnosing the severity of the episode.

2.3.2 Pneumonia

Pneumonia is an infection that inflames the air sacs in one or both lungs, which may fill with fluid or pus[23]. In pneumonia, the respiratory sounds can vary widely, but one of the most distinctive sounds associated with this condition is coarse crackles(described in section 2.2.1). These sounds are short, explosive, and non-musical[24]. They occur when air opens closed air spaces, typically during the latter part of inhalation and are a result of fluid or pus in the air sacs of the lungs. The crackles are generally low-pitched and can be heard in both phases of the respiratory cycle but are more noticeable during deep breaths. Identifying these crackles is crucial for diagnosing pneumonia and assessing its progression.

2.3.3 Chronic Obstructive Pulmonary Disease(COPD)

Chronic obstructive pulmonary disease(COPD) is a chronic inflammatory lung disease that causes obstructed airflow from the lungs[25]. It is commonly associated with a combination of wheezes and rhonchi. Rhonchi are low-pitched, snoring-like sounds that suggest the presence of thick mucus in the larger airways. They are continuous sounds and can be heard throughout the respiratory cycle, although they may be more prominent during expiration. The presence of both wheezes and rhonchi in a COPD patient indicates significant airway obstruction[26]. These sounds provide valuable information about the extent of airflow restriction and the effectiveness of ongoing treatment.

2.3.4 Upper Respiratory Tract Infection (URTI)

Upper Respiratory Tract Infections (URTI) commonly known as the common cold, typically affect the nose, throat, and upper airways[27]. These infections produce diverse respiratory sounds that can aid in diagnosis. Symptoms like harsh breathing and stertor — a snoring-like sound — often occur due to the narrowing or blockage of the upper airway passages by mucus or inflammation. Alongside these auditory signs, URTIs manifest clinical symptoms such as a runny nose, sore throat, and coughing, which are characteristic of the common cold. While mild crackles may be heard at the infection's onset, these sounds are usually transient[28]. Identifying these specific respiratory and clinical symptoms is crucial for distinguishing URTIs from more severe respiratory conditions, thereby guiding appropriate treatment and preventing the unnecessary use of antibiotics.

Focusing on the specific characteristics of respiratory sounds provides a crucial evidence in diagnosing and monitoring respiratory diseases. The ability to accurately identify and describe these sounds can guide healthcare providers in early detection and management, ultimately improving outcomes for patients with respiratory conditions. The detailed analysis of these sounds is essential in the structured examination and understanding of respiratory diseases as well.

3 Literature Review

Analyzing respiratory sounds is crucial for diagnosing respiratory-related diseases. Particularly, distinctive abnormal respiratory sounds such as wheezing and crackling are associated with specific respiratory diseases and serve as important clues for identifying these conditions. As a result, over the past few decades, various studies have been undertaken to automate the analysis and classification of respiratory sounds using AI. Securing a large dataset is crucial when building AI models. If the data is not sufficient, the model might work well on training data but fail to generalize to new, unseen data, a problem known as overfitting. Therefore, the quality and size of the dataset directly impact the predictive power and reliability of AI systems. However, the absence of publicly available large-scale databases has made it difficult to objectively compare the performance of different methodologies. In this chapter, studies using private datasets are reviewed, along with the introduction of the ICBHI dataset, research conducted using this dataset, and investigations utilizing the Coswara dataset, a significant open dataset prompted by the respiratory disease COVID-19.

3.1 Limitations of Using Private Respiratory Datasets

Private respiratory datasets have frequently underpinned pivotal research in the field of medical diagnostics. However, the use of such private datasets often presents the limitation of future work, as they typically restrict broader access and independent verification of research findings. One interesting study quantified and characterized lung sounds from pneumonia patients to generate auscultatory pneumonia scores[29]. The sound analyzer assisted in detecting pneumonia with a sensitivity of 78% and a specificity of 88%. This study is meaningful in that it developed a model specialized for pneumonia diagnosis for the first time, but it has a limitation that follow-up research could not be conducted because the public was not allowed access to the dataset. In another study, Tomasz et al. employed neural network(NN)-based analysis to distinguish four types of abnormal sounds (wheezing, rhonchi, coarse, and fine crackles)[30]. Interestingly, the results showed that NN F1 scores were significantly better than those of physicians. While this research raised expectations for the introduction of AI models into the medical field, the lack of further studies meant that these expectations ended with anticipation alone. Beside this, Gorkem et al. utilized Support Vector Machines (SVM), k-nearest neighbor approach, and multilayer perceptrons to detect lung crackles[31]. Additionally, Gokhan et al. proposed the automatic detection of respiratory cycles and collected synchronous auscultation sounds from chronic obstructive pulmonary disease(COPD) patients[32][33]. Intriguingly, they found that deep learning could diagnose COPD and enhance the understanding of its auditory characteristics more effectively than traditional methods[32]. These studies have advanced the field of respiratory research, which previously lacked detailed investigation. However, some disappointment remains due to concerns about data bias due to small datasets and lack of follow-up studies.

3.2 Introduction of ICBHI 2017 as a Public Dataset

To overcome the challenges of limited data access and enhance reproducibility in respiratory sounds research, the 2017 International Conference on Biomedical and Health Informatics (ICBHI) Challenge released a large dataset of respiratory sounds, facilitating the development of various respiratory sound analysis algorithms. Researchers have emphasized the need to consider various anatomical factors when recording respiratory data. Given that the lungs are significantly larger than the heart, accurate analysis of diagnosis requires recordings from multiple areas of both lungs[7]. Additionally, the quality of lung sounds is easily influenced by the patient's breathing effort. As a result, the dataset provided by the competition was collected from various institutions, resulting in variability in the body areas where the respiratory sounds were measured. However, the four different recording equipment used to build the dataset and the presence of ambient noise in several samples made it difficult for researchers to build a classification model for breath sounds.

3.3 Research Using ICBHI Dataset

Sabers et al demonstrated the highest classification performance compared to the baseline methodology released by ICBHI, winning the competition[5]. The baseline methodology released by ICBHI involved setting the sampling rate for all respiratory sound samples to 4000Hz and training a decision tree model using features extracted through the Mel-frequency cepstral coefficients(MFCC) algorithm. The classification accuracy of the decision tree showed performances of 75% for normal respiratory sounds and 12% for abnormal sounds. Using the same 4000Hz sampling rate for all respiratory sound data as the baseline, they underwent three main steps to extract appropriate features for the classification model.

Firstly, to eliminate noise such as coughing sounds and heartbeats, they applied a 12th-order bandpass filter to each sample. Secondly, since wheezes are found in high-frequency areas and crackles in low-frequency areas, they used a tunable Q-factor wavelet transform to decompose the noise-filtered respiratory sounds into high, low, and remaining frequency bands. In the final step, they used the Short-Time Fourier Transform algorithm to extract spectrograms from each frequency area and the tunable Q-factor wavelet algorithm to obtain wavelet coefficients. These extracted spectrograms and wavelet coefficients, being in high-dimensional spaces, were summarized into six statistical measures (mean, standard deviation, minimum, maximum, skewness, and kurtosis) to use as final features for the respiratory sounds. The classification model used was a Support Vector Machine (SVM), and it demonstrated a higher classification performance compared to ICBHI's baseline methodology, with 78% accuracy for normal respiratory sounds and 20% for abnormal sounds.

Minami et al. presented a deep learning methodology that can learn features on its own without human intervention as much as possible in the process of extracting features from breathing sounds. They applied Short-Time Fourier Transform (STFT) and Wavelet Transform algorithms to each respiratory sound sample to extract spectrogram and scalogram features, respectively, converting these features into images. The classification model was trained using the VGG-16 model, pre-trained on ImageNet data, which was fine-tuned for this task. Each VGG-16 model was independently trained on the feature images, extracting features from each image, which were then combined and passed through a fully connected layer to classify the respiratory sounds. They demonstrated improved performance compared to traditional machine learning methodologies, with a classification accuracy of 81% for normal respiratory sounds and 28% for abnormal respiratory sounds.

They also extracted spectrogram and wavelet matrix features from each respiratory sound sample and converted them into images. They independently trained ResNet models on each image. Similar to their methodology, the features extracted from the ResNet models were combined and passed through a fully connected layer to classify the respiratory sounds. They achieved a classification accuracy of 69.2% for normal respiratory sounds and 31.12% for abnormal respiratory sounds, indicating a higher accuracy for abnormal sounds compared to the methodology proposed by them. These deep learning methodologies, which convert extracted features from respiratory sounds into images, eliminate the complex preprocessing and feature extraction steps required by traditional machine learning methodologies.

Currently, the most recommended study in ICBHI is Bae et al.'s research[4]. Their study shows that pre-trained models on large-scale visual and auditory datasets can generalize to breath sound classification tasks. We also introduced a simple patch mix function that uses the Audio Spectrogram Transformer (AST) to randomly mix patches between different samples. They proposed a novel and effective Patch-Mix Contrastive Learning to distinguish mixed representations in latent space, and their method achieved state-of-the-art performance on the ICBHI dataset, achieving 62.37% which is a 4.08% improvement over the previous best score.

The second high-profile study is Respirenet by Gairola et al[34]. In this work, they proposed a simple CNN-based model with a set of novel techniques such as device-specific fine-tuning, connection-based augmentation, blank region clipping, and smart padding. They performed extensive evaluations on the ICBHI dataset and improved the state-of-the-art results on grade 4 classification by 2.2%. However, they pointed out that there are still limitations due to the small dataset.

The most recently published study is Niizumi et al.'s study They introduced Masked Modeling Duo (M2D) and Extended M2D for X (M2D-X) to increase the efficiency of small medical datasets. M2D-X uses self-supervised learning with mask prediction to improve general-purpose and special-purpose audio representations. Their approach shows potential application to medical audio analysis by exploiting background noise and additive operations to improve the learning process for small and diverse datasets.

However, the studies mentioned above focused on the classification accuracy of respiratory sounds, such as crackles and wheezes, rather than on the actual diagnosis of conditions like asthma and pneumonia. This limitation is mainly due to the small size of the datasets, making disease diagnosis challenging. In summary, research on respiratory sound analysis has been actively conducted under the ICBHI datasets due to issues with small datasets and imbalance in disease diagnosis classes, but it has not led to actual diagnoses. To address this issue, some are exploring ways to expand the dataset by integrating it with other respiratory sound datasets.

3.4 Insights from Mixed Database Studies

The Coswara dataset, emerging in response to the COVID-19 pandemic, has been instrumental in advancing the field of respiratory sound analysis. This dataset includes not just typical respiratory sounds but also annotated data capturing variations across different stages of COVID-19 and other respiratory conditions. Utilizing this dataset, researchers have researched novel AI-driven approaches to enhance the diagnostic capabilities of sound-based respiratory disease detection systems. The below section reviews critical studies that have successfully utilized these mixed datasets(ICBHI and Coswara), demonstrating enhanced diagnostic accuracy and illuminating the potential and challenges of these methodologies.

In Ramasubramanian's study(2022), a machine learning approach is developed to monitor lung conditions and diagnose COVID-19 severity stages through the analysis of breath sounds[35]. The study systematically outlines the progression of COVID-19 into three distinct stages: mild cases are marked by symptoms such as fever and sore throat, with possible wheezing; moderate cases involve persistent cough and tachypnea, accompanied by fine crackles that signal the onset of lower respiratory tract involvement; severe cases require urgent medical attention and are characterized by both fine and coarse crackles, indicative of severe pneumonia and extensive lung inflammation. This comprehensive classification of symptoms and their corresponding acoustic markers enables a more focused and efficacious method for diagnosing. The preprocessing involved classifying the respiratory sounds into distinct categories: sounds with no respiratory markers were classified as either healthy or indicative of tachypnea, while sounds with crackles were further categorized into coarse and fine crackles.

Silent sound files and those with high background noise were excluded to ensure data quality. The reclassification process was rigorously validated by two different doctors to ensure its medical relevance. Following classification, each .wav file from the dataset of 226 patients, which had been split into 6898 respiratory cycles, was segmented according to the start and end times specified in the accompanying label files. The initial method uses Google Cloud AutoML to classify respiratory cycles based on spectrograms, while the more effective second approach applies Log Mel filter-bank features to train multiple CNN models. This hierarchical model approach achieved a remarkable accuracy of 78.12%, showcasing its potential to substantially enhance the remote monitoring and management of home-isolated COVID-19 patients. In summary, while this study significantly advances the systematic classification of COVID-19 progression into three distinct stages, it predominantly focuses on identifying respiratory sound markers such as crackles and wheezes, rather than diagnosing specific conditions like COPD or asthma, which may limit its direct applicability as a medical diagnostic tool.

Chunhapran et al have researched the application of Deep Convolution Neural Networks (DCNN) to classify respiratory conditions by analyzing audio data from both the ICBHI 2017 Respiratory Sound Database and the Coswara database. This study classification using Deep Convolution Neuron Network (DCNN) to distinguish spectrogram images from the respiratory sound in Bronchiectasis, Bronchiolitis, COPD, LRTI, Pneumonia, URTI, COVID-19, and Healthy. They focused on distinguishing left and right lung sounds due to their differing anatomical characteristics and functions, which can affect respiratory symptom manifestation. For instance, the left lung has a unique shape due to the heart's position, affecting the way air flows and is distributed. This can influence the sound of breathing, wheezes, and crackles, which are key indicators of respiratory health and disease states. By creating separate models for the left and right lungs, the study aimed to enhance the accuracy and specificity of disease detection and classification, revealing that diagnostic accuracies can vary based on the side of the lung being analyzed, with overall accuracies reaching up to 86%. This study provides important insight into the complexities of deploying AI in medical diagnosis by highlighting the potential for combining datasets to improve diagnostic accuracy, as well as the need for models that can adapt to physiological differences. However, the study's use of the Coswara dataset solely to distinguish COVID-19 symptoms, without integrating it with ICBHI data for conditions like asthma or pneumonia, suggests that the data was not optimally augmented to broaden its diagnostic utility beyond COVID-19 detection.

In the study by Wall et al., the ICBHI and Coswara datasets were strategically integrated to create a robust framework for lung abnormality classification[36]. This integration process involved extensive data preprocessing, including noise reduction and volume normalization, to standardize the diverse audio recordings from both datasets. Advanced feature extraction techniques, such as Mel-frequency cepstral coefficients (MFCCs) and Chroma features, were applied to harness the acoustic diversity available across both datasets. The researchers also employed audio augmentation techniques like pitch shifting and time stretching to enhance the generalizability of their models. By using a combination of CRNN, BiLSTM, BiGRU, and CNN models optimized with Particle Swarm Optimization, they significantly improved the models' ability to distinguish COVID-19 from other respiratory diseases. This meticulous integration allowed for comprehensive model testing and validation across varied audio inputs, ensuring the developed models' applicability in diverse real-world environments. However, similar to the study by Chunhapran et al., the model primarily focuses on distinguishing the presence or absence of COVID-19 symptoms, and thus, it has not been substantially extended to serve as a practical diagnostic tool for a broader range of respiratory diseases.

In summary, this literature review addresses the limitations of private datasets and ICBHI. To overcome these limitations, the review investigated significant research done in respiratory sound analysis by using both ICBHI and Coswara datasets together. However, it was confirmed that no research has been conducted yet on diagnosing respiratory diseases directly by integrating the Coswara and ICBHI datasets.

In this study, the ResNet-34 model is used to analyze the integrated dataset and directly diagnose respiratory diseases. This research aims to go beyond just classifying abnormal respiratory sounds; it applies transfer learning with the ResNet-34 model to develop an AI model that can accurately diagnose respiratory diseases from sounds. This model has the potential to revolutionize early detection and diagnosis, particularly in remote or resource-limited settings, by providing medical professionals with a reliable and automated tool.

4 Methods

This chapter introduces the datasets utilized and describes the methods employed for data pre-processing as well as the ResNet-34 models used in this study. All computational tasks were performed on the Hábrók High Performance Computing (HPC) Cluster at the University of Groningen. For the training of the models, NVIDIA A100 GPUs within this HPC environment were used.

4.1 Datasets

By merging Coswara with the ICBHI dataset, which contains detailed recordings of different respiratory sounds, I aim to create a more robust and comprehensive dataset. This will address data scarcity and imbalance, allowing for the more accurate and reliable AI model.

4.1.1 ICBHI Datasets

The ICBHI dataset includes 920 recordings from 126 patients, accumulating to a total duration of 5.5 hours and 6,898 respiratory cycles. Experts have annotated each breathing cycle in these recordings into one of four categories: normal, crackle, wheeze, or both (crackle and wheeze). These recordings were gathered using four different devices in hospitals located in Portugal and Greece. For each patient, data was recorded from seven distinct body locations. The diagnosis from this dataset consists of 64 instances of Chronic Obstructive Pulmonary Disease (COPD), 26 instances of healthy subjects, 6 instances of pneumonia, 14 instances of Upper Respiratory Tract Infection (URTI), 1 instance of asthma, 2 instances of Lower Respiratory Tract Infection (LRTI), 7 instances of bronchiectasis, and 6 instances of bronchiolitis.

4.1.2 Coswara Datasets

The Coswara dataset, developed in response to the COVID-19 pandemic, provides labeled respiratory sound recordings, including samples from healthy individuals and those with various respiratory conditions[9]. The Coswara includes respiratory sounds recorded from 2635 individuals, consisting of 1819 COVID-19 negative, 674 positive, and 142 recovered subjects. The dataset features nine categories of sounds: two types of breathing (deep and shallow), two types of coughing (deep and shallow), three types of vowel phonation ([u], [i], [æ]), and two types of continuous speech (counting numbers at normal and fast paces). In this paper, of the two types of breathing, only the deep-breathing sound, which is most similar to the ICBHI data in terms of breathing duration, was used. These recordings were accompanied by metadata, including demographic, health, and COVID-19 status information, and have been manually annotated for audio quality.

The dataset includes a significant number of non-COVID subjects who exhibited COVID-19-like symptoms, such as fever (225 individuals), cough (173 individuals), and cold (90 individuals). Additionally, there were several non-COVID subjects with pre-existing respiratory conditions, including pneumonia (83 individuals), asthma (110 individuals), and other respiratory diseases (120 individuals). This paper used data from patients with asthma, pneumonia, and colds (URTI), which were conditions also represented in the ICBHI dataset.

4.2 Data Pre-processing

Pre-processing datasets covering breath sounds is important before combining them to ensure data integrity and consistency. This process includes normalizing data format and scale to align features such as sound amplitude and length, cleaning and removing noise or irrelevant data to improve data quality, and handling missing values to avoid bias in the analysis. The detailed pre-processing process for each data set is introduced in the subsection below.

4.2.1 ICBHI

In this study, pre-processing of the ICBHI dataset was performed as follows. Firstly, all recordings were downsampled to a uniform rate of 4 kHz, standardizing the data input and ensuring compatibility across various recording conditions. This step is essential for maintaining consistency in the feature extraction phase. Then the 5th-order Butterworth band-pass filter was applied to each downsampled recording. This filtering is crucial for removing unwanted low and high-frequency noise such as heartbeat sounds and background speech, which can obscure or distort the respiratory sounds that are critical for accurate disease diagnosis. Additionally, normalization of the audio signals was performed, adjusting the amplitude of the sound waves to a standard range of -1 to 1. This normalization is important for minimizing variability in sound intensity, which can result from differences in recording devices or environmental conditions. After completing the initial steps, additional preprocessing was conducted following the method proposed by Gairola et al[34], as outlined in Figure 1, which begins with concatenation-based augmentations to combine samples of the same respiratory class, addressing class imbalance and enriching the dataset. This followed by smart padding, which adjusts the lengths of breathing cycles by adding meaningful segments from the same or a normal class to standardize input sizes and maintain relevant data integrity. Finally, the signal was converted to a Mel spectrogram and subjected to black region clipping to remove frequency regions with no signal to improve the focus of the neural network on relevant features.

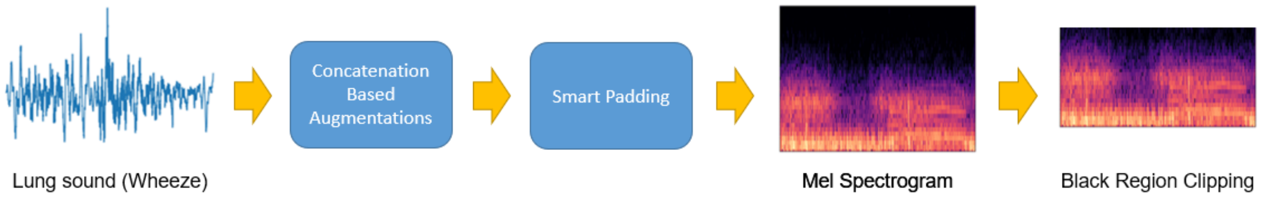


Figure 1: Overview Process of Respiratory Sounds Preprocessing (Wheeze)

4.2.2 Coswara

In the preprocessing of the Coswara dataset, the same method as in the preprocessing of the ICBHI dataset was first applied. However, considering the nature of the Coswara dataset where individuals self-record breathing sounds, additional issues such as various background noises and incomplete recordings were raised. Therefore, special attention was paid to eliminating silent or zero-second recordings that had no diagnostic value. Additionally, to ensure that only high-quality data was used to train the model, recordings that were too short or contained significant amounts of silence were excluded from the dataset. This preprocessing not only improved the dataset but also tailored it to the stringent requirements needed for effective breath sound analysis, improving the overall accuracy and performance of the subsequent machine learning process.

4.3 Transfer Learning

The best way to create a high-performing AI model is to secure a large amount of data. However, there may be cases where the amount of data is not large or it costs a lot to secure the data. Applying transfer learning was proved to be a working approach to overcome limitations of small datasets[37]. Therefore, given the size of both ICBHI and Coswara, I used the transfer learning approach in this thesis. Transfer learning is a method of using some of the abilities of a neural network learned in a specific field to learn a neural network used in a similar or completely new field[38]. It is effective when the number of training data is small and provides much higher accuracy and faster learning speed than learning without transfer learning. The learned neural network used in transfer learning is called a pre-trained model. The pretrained model which has already been trained on a first dataset, is adapted by slightly modifying its weights to align with the specifics of the new problem situation, such as the size of the second dataset and the relevance between datasets. Figure 2 shows an architecture of transfer learning. As illustrated in the figure 2, a model using the ICBHI dataset is trained and performs learning task 1 to classify 8 disease classes by learning common breath sound patterns. Afterwards, learning task 2, which classifies the 3 disease classes is performed using the Coswara dataset as an input to the pre-trained model trained for learning task 1. This addresses the limitations of small datasets and class imbalance issues, enhancing the model's diagnostic identification ability.

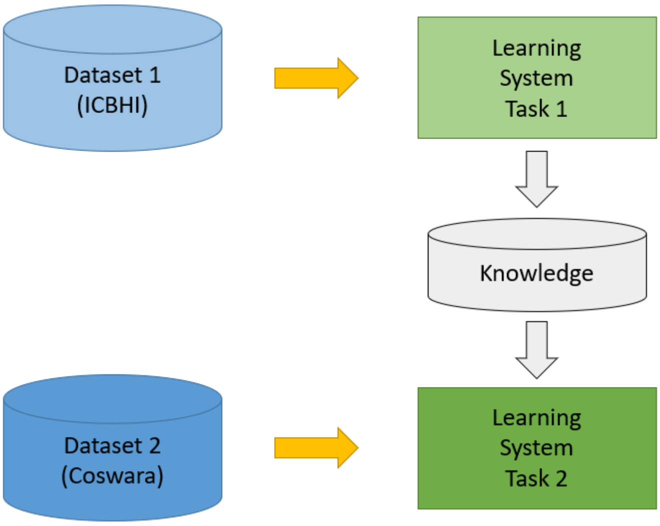


Figure 2: An Architecture of Transfer Learning

4.3.1 ResNet Model

The ResNet-34 model introduced by He et al.(2016) features 34 layers, including convolutional layers and identity shortcut connections (see Figure 3)[8] . These shortcuts, known as residual connections, allow the input from one layer to be fast-forwarded to a later layer, thus skipping certain layers in between. They help prevent the vanishing gradient problem, allowing deeper networks to be trained effectively. The ResNet-34 model consists of four main convolutional base groups, each containing

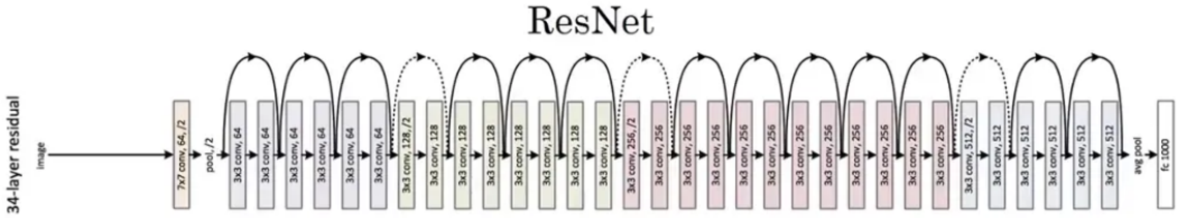


Figure 3: An Architecture of ResNet-34 Model

multiple residual blocks. The first group has 3 blocks, the second has 4, the third contains 6, and the fourth has 3 blocks, totaling 16 blocks. In Figure 3, the ResNet model's four primary block groups are distinguished by different colors. The first group, shown in light purple, is located at the beginning of the network and extracts low-level features, such as edges and textures. The second group, in light green, recognizes more complex shapes and structures. The third group, depicted in pink, processes higher-dimensional features by recognizing complex patterns such as interrelationships. The last group, in light gray, extracts the most abstract and high-dimensional features. This stage is crucial for understanding the overall context of the image and processing important information for classifying complex objects or scenes. This paper compares the stability and accuracy of the model when trained from the second, third, and fourth blocks during transfer learning.

4.3.2 Stage1: Pre-trained Model

In this stage, a ResNet-34 model was used to analyze and classify respiratory diseases using the ICBHI 2017 dataset. Unlike previous studies have focused on classifying respiratory sounds[4][34], this study uniquely targets the classification of respiratory disease diagnoses. Therefore, the model has the classifier which has the role of distinguishing diagnosis; Healthy, URTI, COPD, Bronchiectasis, Pneumonia, Bronchiolitis, Asthma, and LRTI. The model was modified for audio data by adjusting the final layers, including dropout layers with a 0.5 probability to prevent overfitting, followed by linear transformations and ReLU activations for refined classification. The model was trained for 20 epochs with a batch size of 64 and an initial learning rate of 10^{-5}, managed by an Adam optimizer.

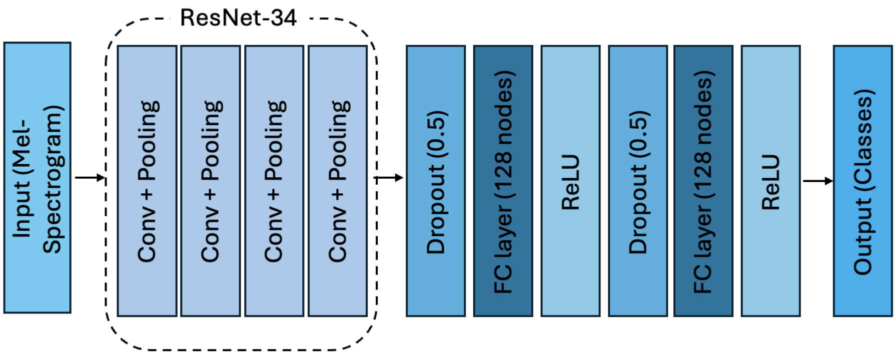


Figure 4: An Architecture of the Pre-trained Model

Figure 4 illustrates the architecture of the pre-trained model. The Mel-Spectrogram images of the respiratory sounds were used as input for the model. The model is based on the ResNet-34 architecture, consisting of four convolutional base which has multiple convolutional and pooling layers. Then dropout layers with a 0.5 probability were added after the convolutional layers to prevent overfitting. Following these, fully connected layers (FC layers) with 128 nodes each were included, with ReLU (Rectified Linear Unit) activations applied to introduce non-linearity. The final output layer provides the probability distribution over different respiratory disease classes. As mentioned earlier, the ICBHI dataset has an unbalanced distribution among diagnostic classes such as 1 asthma class and 64 COPD. So the goal was to stabilize data augmentation and enhance the robustness of the model by applying the Coswara dataset during the fine-tuning process.

4.3.3 Stage2: Fine-tuning Process

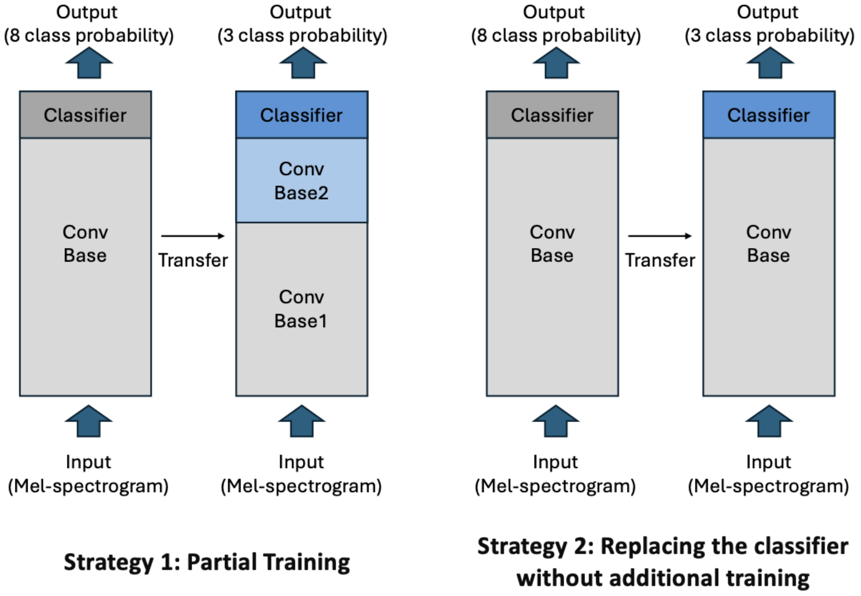


Figure 5: The Strategies of the Fine-tuning

Fine-tuning refers to a method of learning by transforming the architecture to suit a new purpose based on an existing model and finely adjusting the weights of the already learned model[39]. In the stage 1, the classifier of the pre-trained model was configured to distinguish between 8 different diagnosis classes using the ICBHI dataset. In the stage 2, two strategies were employed using Coswara dataset. Figure 5 illustrates the two fine-tuning strategies employed. At this stage 2, both strategies changed the classifier to distinguish between three diseases: URTI (common cold), asthma, and pneumonia. In other words, the main change is that it was reduced from 8 diagnostic classes to 3 by considering the sample labels of the Coswara dataset.

The strategy 1, ”Partial Training”, involves partially freezing the convolutional base while training the remaining layers and the classifier. The convolutional base, composed of multiple layers of convolutions and pooling, serves as the feature extraction component. The classifier, typically consisting of fully connected layers, uses the features extracted by the convolutional base to classify each image sample into the correct class (image classification). This strategy is suitable for the datasets with high similarity, where overfitting is less of a concern due to the dataset size, and the pre-trained model can effectively utilize previously learned knowledge due to the high similarity of the data[39]. In this strategy, three different settings were tested by freezing first, second and third convolutional base of the ResNet-34 model. This strategy is used to confirm the possibility of data augmentation by verifying the high similarity between the two datasets.

The second strategy is to freeze the convolutional base and train only the classifier. It focuses on modifying the classifier part of the architecture without further training the convolutional base. This approach was employed when the task at hand closely resembles the dataset used for the pre-trained model[39]. The existing convolutional base was utilized as a fixed feature extractor, and only the classifier section was adapted and trained to accommodate the new task. This has the advantage of shortening the time required for transfer learning.

4.4 Repository and Execution Guidelines

All code used in this study has been backed up in the repository linked below. Please follow the guidelines provided in the README.md file to execute the code. The repository does not include the data used in the study. Please visit the official websites of ICBHI[7] and Coswara[9] to directly download the data.

The project's source code is available at `https://github.com/Gyeongaa/VT_Graduation_Thesis/tree/main`.

5 Results

This chapter presents the results obtained from the application of models on the augmented respiratory sound datasets comprising ICBHI and Coswara data.

5.1 Pre-trained Model

The performance of the pre-trained ResNet model is illustrated in Figure 6. The model, initially trained with a learning rate of 10^{-5}, a batch size of 64, and throughout 20 epochs, demonstrated significant improvements in learning respiratory diseases.

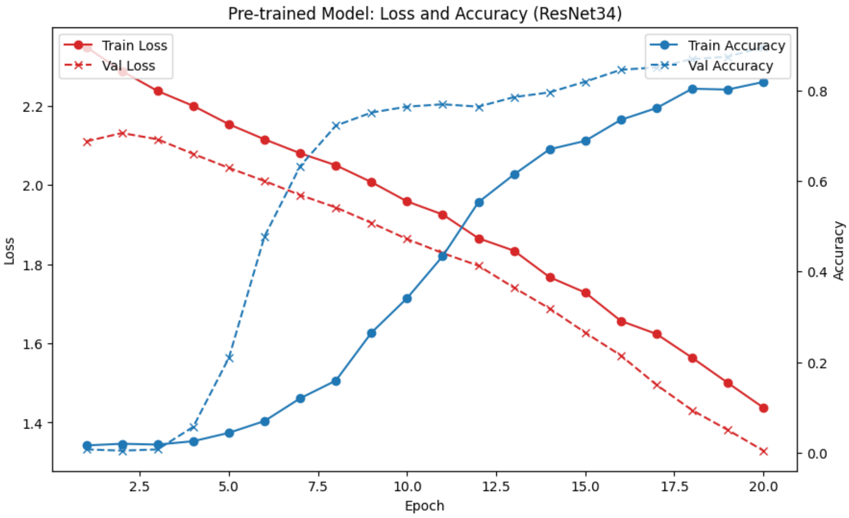


Figure 6: The Pre-trained models' loss and accuracy

As illustrated in Figure 6, the initial phase of training the pre-trained model showed results in tackling respiratory diseases, with a reduction in training loss from approximately 2.2 to below 1.6 and an increase in accuracy to above 0.8. Validation loss and accuracy trends indicate some challenges in generalizing the learned features to unseen data. The relatively high validation accuracy observed in the model's performance can be attributed to the distribution of class samples within the training dataset. Especially, the model tends to predict the COPD class more accurately, likely because this class has the most samples available. Conversely, classes like asthma, which are represented by only one sample, are not sufficiently learned by the model. This imbalance in class distribution leads to the model achieving higher validation accuracy more easily, as it becomes adept at recognizing the more frequently represented classes while struggling with underrepresented ones. To address this imbalance, the initial ICBHI split was abandoned, and the data was randomly redistributed while maintaining an 8:2 ratio for training and testing. This reorganization helped alleviate the class imbalance and increased the model's accuracy from 74.8% to 89.5%, demonstrating the better model's performance.

5.2 Fine-tuning

5.2.1 Strategy 1

Freezing convolutional base during the fine-tuning of neural networks preserves the generic features learned from the original dataset while adapting the model to new data. This approach is particularly useful for transfer learning, where the goal is to leverage previously acquired knowledge to improve performance on another task without having to start from scratch. In ResNet models, the initial layers often capture basic visual patterns such as edges and textures, which are fundamental across various tasks[8]. Freezing these layers can maintain these essential features while the deeper layers, closer to the output, adapt to specifics of the new dataset.

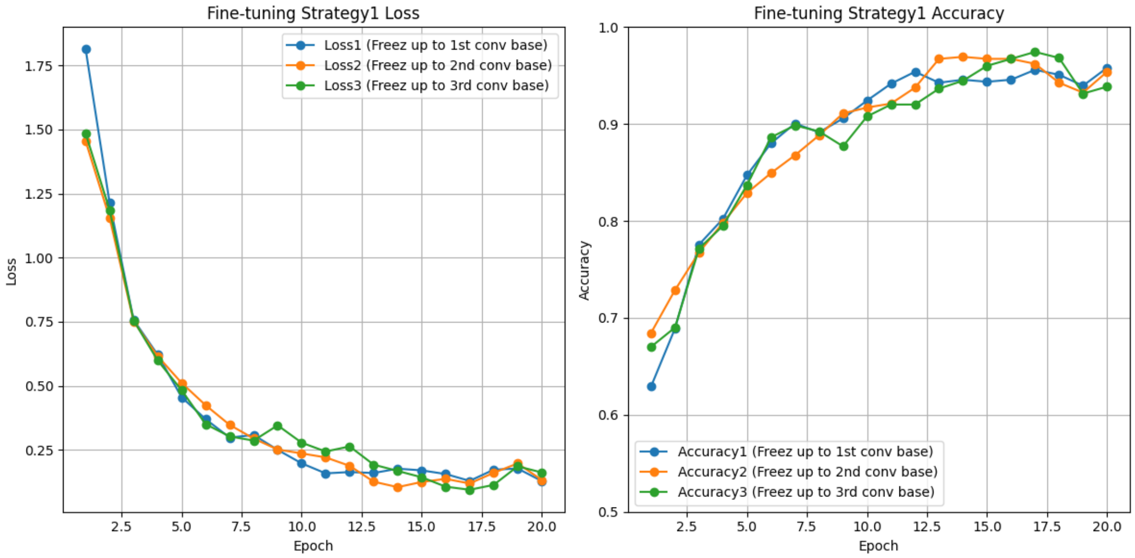


Figure 7: Epoch-wise Loss and Accuracy Trends with Variable Layer Freezing

In the fine-tuning strategy 1, three different settings were tested by freezing up to the first, second, and third convolutional base of the ResNet-34 model. These settings are referred to as Loss1 and Accuracy1 (freeze up to the first convolutional base), Loss2 and Accuracy2 (freeze up to the second convolutional base), and Loss3 and Accuracy3 (freeze up to the third convolutional base), respectively. In this paper, the three different settings discussed will be referred to for convenience as fine-tuning strategies 1-1, 1-2, and 1-3. As shown in the left graph of Figure 7, all three strategies exhibited a sharp decrease in loss within the initial epochs. After this initial drop, the loss values continued to decrease gradually, converging to a stable state around epoch 12.5. Loss3 showed slightly more variation in later epochs compared to Loss1 and Loss2, reaching its minimum at 0.0947 by the 17th epoch and closing at 0.1622 by the final epoch. The right graph in Figure 7 displays the accuracy trends for each setting. All strategies demonstrate a rapid increase in accuracy during the initial epochs, reflecting the model's rapid adaptation to the fine-tuning process. The accuracy then plateaus after approximately 15 epochs, with each setting achieving similar high-performance levels by the end of training. For instance, Accuracy1 started at 62.91% and rose significantly to peak at 95.59% by the 17th epoch, finishing the training at 95.80%. Accuracy2 commenced at 68.44%, surged to 96.93% by the 14th epoch, and ended at 95.39%. Lastly, Accuracy3 finished at 93.75%. From the results, it can be inferred that allowing more layers to learn by freezing less extensively enables the model to better adapt to new datasets. However, this can vary depending on the quality of the dataset. In the case of Coswara, which consists of respiratory sounds recorded personally by participants, there is significant variability in sample quality. Therefore, as evidenced in the graph above, strategy 1-2 is seen to be the most consistently trained across the different strategies.

Figure 8 below displays the results of the first three fine-tuning strategies in confusion matrix format, detailing the classification results for diseases such as URTI, pneumonia, and asthma across the

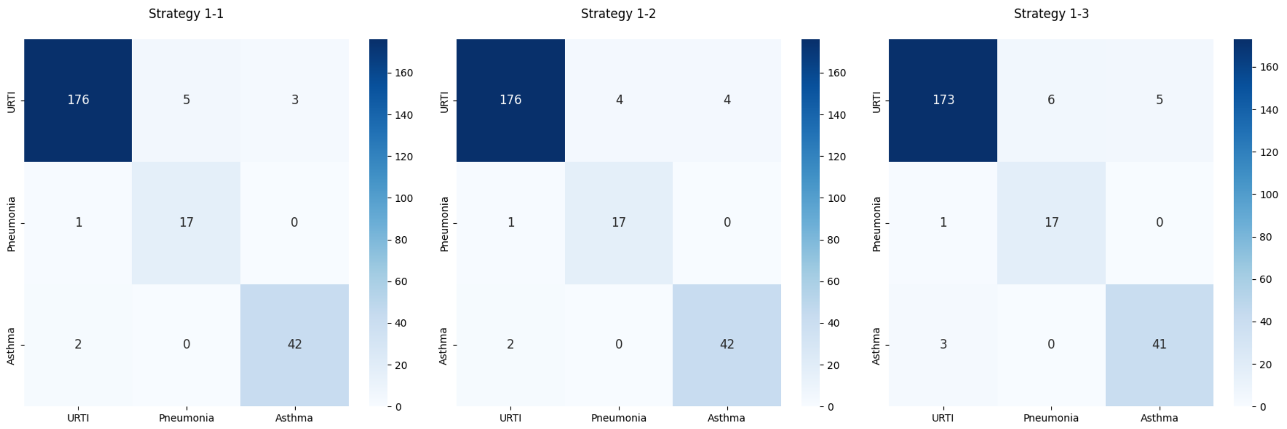
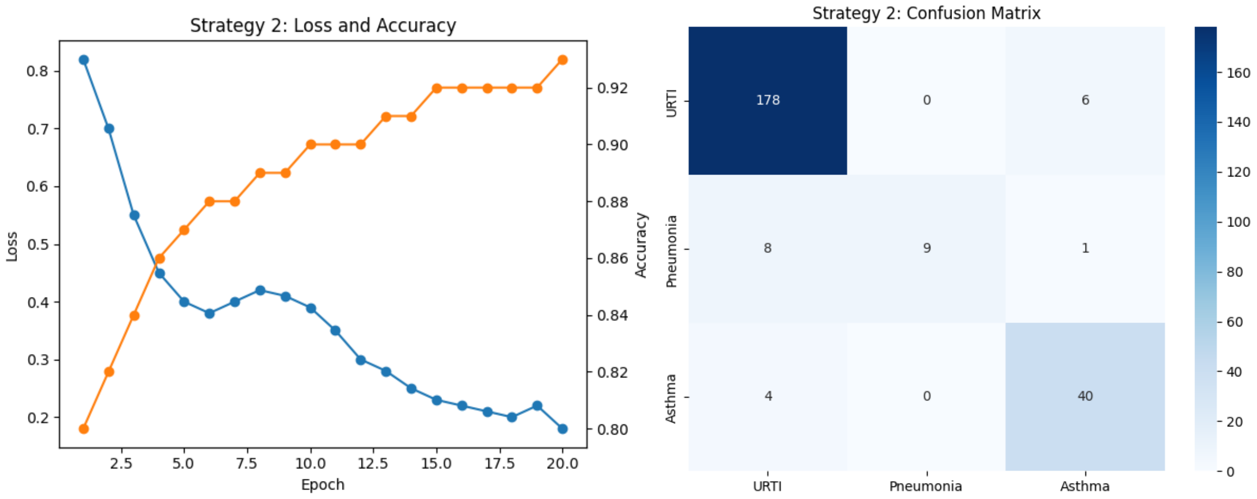


Figure 8: Fine-tuning Strategy1 Comparison via Confusion Metrics

different strategies. Each matrix shows how the models perform with varying degrees of layer freezing—up to the 1st, 2nd, and 3rd convolutional base, respectively. For example, in the first strategy, the model successfully predicts URTI with 176 true positives but confuses it with pneumonia and asthma in 5 and 3 cases, respectively. The Coswara dataset is primarily composed of URTI cases, which has led to class imbalance issues. This imbalance has been observed to cause a tendency for the model to misclassify pneumonia and asthma cases as URTI. Nonetheless, addressing the class imbalance in the pre-trained model improved diagnostic classification accuracy for asthma and pneumonia.

5.2.2 Strategy 2



(a) The Loss and Accuracy

(b) The Confusion Matrix

Figure 9: The Results from Fine-tuning Strategy 2

In the fine-tuning strategy 2, the model uses the same core structure but updates only the final decision-making part (the classifier) to adapt to the new data from the Coswara dataset. This method does not change the layers that handle basic feature extraction from respiratory sounds. It focuses on improving the model's output mechanisms to better classify diseases based on previously learned features. The performance of the model is detailed in Figure 9. As shown in 9a, the model completes 20 training epochs, beginning with a training loss of 0.82, which consistently decreases, concluding at 0.18 by the last epoch. However, the loss intermittently increases between epochs 6 and 10, and again at epoch 19. The accuracy metric starts strong at 0.8 and gradually improves throughout the training period, peaking at 0.93 in the last epoch. Despite these improvements in overall accuracy, the model continues to face limitations. First, because only the classifier was changed for fine-tuning with the Coswara dataset, the model shows more instability compared to Strategy 1. Second, there is ongoing difficulty in classifying disease due to the class imbalance in the Coswara dataset. Especially, this strategy is influenced heavily by the imbalanced sample structure as it proceeds without prior model learning on the dataset. Nevertheless, it is important to note that the accuracy of the model improves by up to 93% and successfully classifies the disease in a shorter training period.

6 Discussion

In this chapter, the presented research questions are reviewed and the hypothesis are verified based on the results derived in Chapter 5. The current study aimed to answer two research questions below. The results help answer these questions and support the initial hypotheses of this research.

6.1 Validation of the Hypothesis

6.1.1 First Hypothesis

The first hypothesis posited that different strategies of freezing convolutional bases during the transfer learning process would significantly impact the model's accuracy and loss metrics, with the prediction that earlier freezing would preserve essential features better, leading to lower loss and higher accuracy. The empirical results presented strategy 1 in Chapter 5 validate this hypothesis to a considerable extent.

In the fine-tuning strategies applied to the pre-trained ResNet model, freezing earlier convolutional bases (up to the first and second one) resulted in more stable and consistent learning curves as Figure 7. Strategy 1-2, which involved freezing up to the second convolutional base, demonstrated the most balanced performance, achieving peak accuracy of about 96.93%, and maintaining lower loss values throughout the training epochs. This strategy was notably effective in addressing the initial high variability in sample quality within the Coswara dataset and achieving a more generalized performance across the augmented dataset. Comparatively, Strategy 1-3, which froze up to the third convolutional base, showed slightly higher variation in loss in later epochs and did not achieve as high accuracy as the other strategies. This suggests that while deeper freezing can lead to over-specialization on training data, freezing earlier layers preserves the model's ability to generalize across new, unseen data from the respiratory sound datasets.

These findings underscore the importance of selecting the appropriate depth of layer freezing in transfer learning to balance between feature preservation and adaptability, supporting the initial hypothesis that earlier freezing benefits the model's performance metrics.

6.1.2 Second Hypothesis

The second hypothesis predicted that changing the classifier when fine-tuning with different datasets would otuperform in terms of both accuracy and loss metrics. This hypothesis was grounded in the assumption that the Coswara and ICBHI datasets, while different, contain sufficiently similar samples that could benefit from a refreshed classification approach to address class imbalance more effectively. This hypothesis was tested through Strategy 2, where only the classifier component of the model was updated.

Results from Strategy 2 show a mixed outcome. While this strategy achieved a significant peak accuracy of 93%, it also displayed higher variability and instability in loss metrics between epochs, suggesting that solely focusing on the classifier could introduce some volatility in the model's learning process. Specifically, the training loss started at 0.82 and ended at 0.18, but it experienced irregular increases. This outcome highlights that while the classifier adjustment can effectively utilize learned features to optimize performance, it may also be prone to overfitting or underfitting if not complemented by adjustments in the convolutional bases. Comparing this with the freezing strategies, where Strategy 1-2 demonstrated the most consistent and highest performance, it appears that the second hypothesis is not entirely supported. The results suggest that a combination of freezing certain layers to preserve generic features and modifying the classifier to adapt to specific dataset characteristics offers a more balanced approach to fine-tuning transfer learning models.

In conclusion, while changing the classifier presents an effective method for refining model performance, especially in terms of accuracy, its implementation should be carefully managed to avoid instabilities in loss metrics. This indicates a partial validation of the second hypothesis, proposing that an integrated approach to fine-tuning might be more effective than any single strategy in isolation.

6.2 Limitation

This study investigated a new territory by focusing on training diagnosis models, an area not covered in existing literature, while it also addresses the effects of various fine-tuning strategies on transfer learning models for respiratory sound datasets. However, it encounters several limitations that are for placing these findings within a broader research context. One primary limitation is the representativeness and diversity of the datasets used, namely ICBHI and Coswara. As noted in the literature review, the Coswara dataset, emerging during the COVID-19 pandemic, includes respiratory sounds that may not comprehensively represent all respiratory conditions. This limitation was similarly highlighted in the studies by Chunhapran et al.[40] and Wall et al.[36], who attempted to utilize this dataset for broader diagnostic applications but acknowledged that dataset focuses on COVID-19 symptoms. Additionally, despite efforts to mitigate class imbalance through different fine-tuning strategies, this study acknowledges that class imbalance remains a important challenge. This issue is not unique to this study but is prevalent across the field, as discussed in the studies by Gorkem et al.[5], where small and biased datasets frequently hindered the generalizability and accuracy of diagnostic models. For instance, it can be observed in Figure 9b that the model tends to select URTI, which is the majority class, when predicting incorrect disease diagnoses. Furthermore, the focus on accuracy and loss metrics, while informative, does not encompass other important model evaluation metrics such as specificity, sensitivity, and predictive value, crucial for clinical applicability. These metrics were extensively analyzed in studies like those by Ramasubramanian et al.[35], which focused on enhancing diagnostic precision beyond mere classification accuracy. Additionally, the reliance on accuracy and loss metrics may lead to an incomplete assessment of a model's clinical utility, potentially overlooking how it performs under varied real-world conditions where diverse diagnostic outcomes are critical. Lastly, the study's reliance on supervised learning approaches, which depend heavily on well-labeled data, poses a limitation in scenarios where labeling is costly or expertise-dependent. This challenge was also noted in the research by Gairola et al.[34], who pointed out the difficulties in applying models trained on small or inadequately labeled datasets.

In summary, this research is meaningful as it advances beyond previous studies that focused primarily on analyzing respiratory sounds, moving towards training models that lead to actual disease diagnosis. It examines the impacts of various fine-tuning strategies on transfer learning models and addresses the challenges posed by dataset limitations and class imbalances. However, it still faces challenges such as class imbalance, small datasets, the reliance on supervised learning and limited time, indicating that there is room for further research in this field. Future studies should utilize more diverse datasets, integrate a broader range of evaluation metrics, and experiment with unsupervised or semi-supervised learning models to overcome these limitations and enhance the robustness and applicability of the findings.

7 Conclusion

The current research investigated the methods to enhance the accuracy of diagnosing respiratory diseases using transfer learning through data augmentation to overcome the limitation of small datasets. The results demonstrated that layer freezing and classifier adjustments are effective strategies to improve the respiratory disease diagnosis classification. The findings of fine-tuning strategy 1 illustrated substantial improvements, with the model achieving an accuracy peak of 95.80% by selectively freezing the initial layers. This method effectively preserved essential low-level features while adapting to new data, proving its efficacy in enhancing diagnostic precision. The results of fine-tuning strategy 2 improved the model's accuracy to 93%, demonstrating the potential benefit of focusing tuning on the decision-making component of the model when datasets have similar characteristics.

The novelty of this study is the expansion of two respiratory datasets. Additionally, transfer learning was performed using the ResNet architecture which promotes fast and effective learning tuning due to its deep residual learning framework. This approach not only addressed the class imbalance inherent in the respective datasets but also improved the stability and overall performance of the model. However, the effectiveness of the model depends on the quality and diversity of the datasets. In case of Coswara dataset which is composed of respiratory sounds recorded by patients without professional equipment, it presented variability that limited the amount of usable data. Therefore, it is important to build high-quality respiratory sound datasets such as ICBHI.

As AI tools become more integrated into medical diagnostics, resolving ethical and regulatory issues becomes increasingly important. Collaborating with the medical community to test and validate these AI models in real clinical settings could ensure their practical effectiveness. By addressing these challenges, future research can build upon this foundation to enhance the capabilities of AI in improving respiratory disease diagnosis, ensuring that technological advancements are matched with practical and ethical applications in healthcare settings.

Future Research

The findings of this thesis open several avenues for future research in applying AI to respiratory disease diagnostics. One critical area is the expansion of datasets used for training AI models. Incorporating a broader array of respiratory sounds from diverse populations will help develop more robust models, capable of accurate performance across different demographic groups. This effort should include data from underrepresented groups to enhance the equity and reduce biases in AI diagnostics. Another important direction is improving the interpretability of AI models within healthcare. Developing methods to make model decisions more transparent is essential for building trust and understanding among healthcare providers. Future studies could explore techniques such as model visualization, feature importance analysis, and simplifying model architectures to achieve this goal. Furthermore, to ensure the practical effectiveness of AI models, they should be tested and validated in real-world clinical settings. Collaborations with hospitals and clinics to implement these models in live environments could provide invaluable feedback for refining the AI tools. Research could also explore the development of multi-label classification systems capable of diagnosing multiple respiratory conditions from a single respiratory sound sample. Such systems would enhance the efficiency and comprehensiveness of diagnostics, providing a significant clinical value. Lastly, as AI tools become more integral to healthcare diagnostics, addressing ethical and regulatory challenges will be paramount. Future work should aim to establish guidelines for the ethical use of AI in healthcare, ensuring that these tools benefit patients without causing unintended harm. By pursuing these research directions, future studies can build upon the findings of this thesis to further enhance the capabilities of AI in improving respiratory disease diagnosis, ensuring technical advancements are matched with practical and ethical application in healthcare settings.

8 Ethics

In this study, publicly available datasets consisting of anonymized respiratory sound recordings, ICBHI[7] and Coswara[9], were utilized. The details on the ethics behind data collection are described in the respective publications. By using anonymized datasets, the risk of privacy breaches was mitigated. Anonymized data were used only by Hábrók (high performance computing cluster system) to protect the data from unauthorized access and to use the data only for research purposes that benefit public health.

The deployment of AI technologies in healthcare, especially those related to diagnostics, always raises important ethical issues. The reason for such concerns is that important health decisions are being left to AI. To make these AI systems transparent, I ensured that they were written in plain language so that users, including clinicians and patients, could understand the decision-making process. Additionally, another important ethical consideration is to ensure that AI models are free from potential bias. The ICBHI dataset used was collected from a specific population, whereas the Coswara dataset was collected from all over the world, so the developed AI model has diversity across demographic groups. By considering these ethical considerations, I tried to integrate AI into healthcare by ensuring data security and addressing potential bias.

Bibliography

[1] World-Health-Organization, “The global impact of respiratory diseases (2nd edition),” *Forum of International Respiratory Societies (FIRS)*, 2017.

[2] Y. M. Arabi, E. Azoulay, H. M. Al-Dorzi, J. Phua, J. Salluh, C. H. A. Binnie, D. C. Angus, M. Cecconi, and B. Du, “How the covid-19 pandemic will change the future of critical care,” *Intensive care medicine*, vol. 47, pp. 282–291, 2021. DOI: 10.1007/s00134-021-06352-y.

[3] L. Arts, E. H. T. Lim, P. M. van de Ven, L. Heunks, and P. R. T. ., “The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: a meta-analysis.,” *Sci Rep*, vol. 10, 2020. DOI: 10.1038/s41598-020-64405-6.

[4] S. Bae, J.-W. Kim, W.-Y. Cho, H. Baek, S. Son, B. Lee, C. Ha, K. Tae, S. Kim, and S.-Y. Yun, “Patch-Mix Contrastive Learning with Audio Spectrogram Transformer on Respiratory Sound Classification,” in *Proc. INTERSPEECH 2023*, pp. 5436–5440, 2023. DOI: 10.21437/Interspeech.2023-1426.

[5] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. U. Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. P. Paiva, I. Chouvarda, and P. de Carvalho, “An open access database for the evaluation of respiratory sound classification algorithms,” *Physiol Meas*, vol. 40, 2019. DOI: 10.1088/1361-6579/ab03ea.

[6] Y. Ma, X. Xu, Q. Yu, Y. Li, J. Zhao, and G. Wang, “Lungbrn: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm,” *Biomedical Circuits and Systems Conference*, vol. 10, 2019. DOI: 10.1109/BIOCAS.2019.8919021.

[7] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, and N. Maglaveras, “A respiratory sound database for the development of automated classification,” *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017*. DOI: 10.1007/978-981-10-7419-6_6.

[8] X. Z. K. He, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. DOI: 10.1109/CVPR.2016.90.

[9] D. Bhattacharya, N. K. Sharma, D. Dutta, S. R. Chetupalli, P. Mote, S. Ganapathy, C. Chandrakiran, S. Nori, K. K. Suhail, S. Gonuguntla, and M. Alagesan, “Coswara: A respiratory sounds and symptoms dataset for remote screening of sars-cov-2 infection,” *Scientific Data*, vol. 10, p. 397, 2023. DOI: 10.1038/s41597-023-02266-0.

[10] A. Bohadana, G. Izbicki, and S. S. Kraman, “Fundamentals of lung auscultation,” *Pattern Analysis and Applications*, vol. 370, no. 8, 2014. DOI: 10.1056/NEJMra1302901.

[11] S. Leng, R. Tan, Chai, and K. et al, “The electronic stethoscope,” *BioMed Eng OnLine 14*, vol. 66, 2015. DOI: 10.1186/s12938-015-0056-y.

[12] D. Bardou, K. Zhang, and S. M. Ahmad, “Lung sounds classification using convolutional neural networks,” *Artif Intell Med*, vol. 88, 2018. DOI: 10.1016/j.artmed.2018.04.008.

[13] A. Gurung, C. G. Scrafford, J. M. Tielsch, O. S. Levine, and W. Checkley, "Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis," *Respir Med*, vol. 105, 2011. DOI: 10.1016/j.rmed.2011.05.007.

[14] M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra, "Auscultation of the respiratory system," *Ann Thorac Med*, 2015. DOI: 10.4103/1817-1737.160831.

[15] M. Aykanat, Özkan Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image Video Process*, vol. 2017, no. 65, 2017. DOI: 10.1186/s13640-017-0213-2.

[16] I. Weisman, "Erratum: Ats/accp statement on cardiopulmonary exercise testing," *Sci Rep*, vol. 167, 2020. DOI: 10.1164/ajrccm.167.10.952.

[17] Z. Dokur, "Respiratory sound classification by using an incremental supervised neural network," *Pattern Analysis and Applications*, vol. 12, 2009. DOI: 10.1007/s10044-008-0125-y.

[18] M. Munakata, H. Ukita, I. Doi, Y. Ohtsuka, Y. Masaki, Y. Homma, and Y. Kawakami, "Spectral and waveform characteristics of fine and coarse crackles," *Thorax*, vol. 46, no. 9, 1991. DOI: 10.1136/thx.46.9.651.

[19] H. Melbye, L. Garcia-Marcos, P. Brand, M. Everard, K. Priftis, and H. Pasterkamp, "Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings," *BMJ Open Respir Res*, vol. 3, no. 1, 2016. DOI: 10.1136/bmjresp-2016-000136.

[20] S. Ohshimo, T. Sadamori, and K. Tanigawa, "Innovation in analysis of respiratory sounds," *Annals of Internal Medicine*, vol. 164, no. 9, 2016. DOI: 10.7326/L15-0350.

[21] J. H. Lee, J.-Y. Kim, J. S. Choi, and J. O. Na, "Respiratory reviews in asthma 2022," *Tuberc Respir Dis (Seoul)*, vol. 85, no. 4, 2022. DOI: 10.4046/trd.2022.0097.

[22] A. Cattamanchi, "What to know about asthma and wheezing." `https://www.medicalnewstoday.com/articles/asthma-wheezing#causes`, 2020. Accessed: October 4, 2023.

[23] Mayo Clinic Staff, "Pneumonia: Symptoms & causes." `https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204`, 2021. Accessed: October 4, 2023.

[24] Y. Kim, Y. Hyon, S. S. Jung, S. Lee, G. Yoo, C. Chung, and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, vol. 11, p. 17186, 2021. DOI: 10.1038/s41598-021-96724-7.

[25] W. MacNee, "Pathology, pathogenesis, and pathophysiology," *BMJ*, vol. 332, no. 7551, p. 1202–1204, 2006. PMCID: PMC1463976.

[26] M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra, "Auscultation of the respiratory system," *Ann Thorac Med*, vol. 10, no. 3, pp. 158–168, 2015. DOI: 10.4103/1817-1737.160831.

[27] D. E. Pappas, "The common cold," in *Principles and Practice of Pediatric Infectious Diseases*, pp. 199–202.e1, Elsevier, 2018. DOI: 10.1016/B978-0-323-40181-4.00026-8.

[28] E. O. Cathain and M. M. Gaffey, “Upper airway obstruction.” StatPearls [Internet], 2024. Updated 2022 Oct 17; Available from: https://www.ncbi.nlm.nih.gov/books/NBK564399/.

[29] R. L. H. Murphy, A. Vyshedskiy, V.-A. Power-Charnitsky, D. S. Bana, P. M. Marinelli, A. Wong-Tse, and R. Paciej, “Automated lung sound analysis in patients with pneumonia,” *Respir Care*, vol. 49, no. 12, 2004.

[30] T. Grzywalski, M. Piecuch, M. Szajek, A. Breborowicz, H. Hafke-Dys, J. Kociński, A. Pastusiak, and R. Belluzzo, “Practical implementation of artifcial intelligence algorithms in pulmonary auscultation examination,” *Eur J Pediatr*, vol. 178, 2019. DOI: 10.1007/s00431-0 9-03363-2.

[31] G. Serbes, C. O. Sakar, Y. P. Kahya, and N. Aydin, “Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 178, 2011. DOI: 10.1109/IEMBS.2011.6090899.

[32] G. Altan, Y. Kutlu, and N. Allahverdi, “Deep learning on computerized analysis of chronic obstructive pulmonary disease,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, 2020. DOI: 10.1109/JBHI.2019.2931395.

[33] S. Aras, M. Öztürk, and A. Gangal, “Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, 2018. DOI: 10.3906/elk-1705-16.

[34] S. Gairola, F. Tom, N. Kwatra, and M. Jain, “Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting,” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2021, 2021. DOI: 10.1109/embc46164.2021.9630091.

[35] C. Ramasubramanian, “Diagnosing the stage of covid-19 using machine learning on breath sounds,” *PHM Society European Conference*, 2021. DOI: 10.36001/phme.2021.v6i1.2858.

[36] C. Wall, L. Zhang, Y. Yu, A. Kumar, and R. Gao, “A deep ensemble neural network with attention mechanisms for lung abnormality classification using audio inputs,” *Sensors 22*, no. 15, p. 5566, 2022. DOI: 10.3390/s22155566.

[37] H. W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, “Deep learning for emotion recognition on small datasets using transfer learning,” p. 443–449, Association for Computing Machinery, 2015. DOI: 10.1145/2818346.2830593.

[38] A. H. Ali, M. G. Yaseen, M. Aljanabi, and S. A. Abed, “Transfer learning: A new promising techniques,” *Mesopotamian Journal of Big Data*, vol. 2023, pp. 31–32, 02 2023. DOI: 10.58496/MJBD/2023/004.

[39] I. Moran, D. T. Altilar, M. K. Ucar, C. Bilgin, and M. R. Bozkurt, “Deep transfer learning for chronic obstructive pulmonary disease detection utilizing electrocardiogram signals,” *IEEE Access*, vol. 11, pp. 40629–40644, 2023. DOI: 10.1109/ACCESS.2023.3269397.

[40] O. Chunhapran, S. Vongananansup, T. Yampaka, and R. Burirat, “Covid-19 and respiratory diseases classification using deep convolution neuron network,” in *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1–6, 2022. DOI: 10.1109/JCSSE54890.2022.9836259.