AUTOMATIC IDENTIFICATION OF ABNORMAL LUNG SOUNDS BY MACHINE LEARNING METHODS

RATTANATHON PHETTOM

**MASTER OF ENGINEERING
IN BIOMEDICAL ENGINEERING**

**CHIANG MAI UNIVERSITY
MARCH 2024**

AUTOMATIC IDENTIFICATION OF ABNORMAL LUNG SOUNDS BY MACHINE LEARNING METHODS

RATTANATHON PHETTOM

MASTER OF ENGINEERING
IN BIOMEDICAL ENGINEERING

CHIANG MAI UNIVERSITY
MARCH 2024

AUTOMATIC IDENTIFICATION OF ABNORMAL LUNG SOUNDS BY MACHINE LEARNING METHODS

RATTANATHON PHETTOM

A THESIS SUBMITTED TO CHIANG MAI UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING IN BIOMEDICAL ENGINEERING

CHIANG MAI UNIVERSITY
MARCH 2024

AUTOMATIC IDENTIFICATION OF ABNORMAL LUNG SOUNDS BY MACHINE LEARNING METHODS

RATTANATHON PHETTOM

THIS THESIS HAS BEEN APPROVED TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ENGINEERING
IN BIOMEDICAL ENGINEERING

Examination Committee:

……………………………. Chairman
(Prof. Dr. Kosin Chamnongthai)

……………………………. Member
(Prof. Dr. Nipon Theera-Umpon)

……………………………. Member
(Assoc. Prof. Dr. Sansanee Auephanwiriyakul)

……………………………. Member
(Assoc. Prof. Dr. Sermsak Uatrongjit)

Advisor:

…………………………………
(Prof. Dr. Nipon Theera-Umpon)

20 March 2024

To my late mother, who bravely battled cancer and succumbed to the illness one year subsequent to diagnosis and to all individuals enduring the hardships of lung cancer.

ACKNOWLEDGEMENTS

Initially, I would like to extend my deepest gratitude to my beloved mother and father for their boundless love, unwavering support, and profound guidance throughout my life. They have played an indispensable role in nurturing me and shaping my journey, imparting invaluable lessons, and instilling in me the courage and determination to navigate the complexities of the world. Their selfless sacrifices, endless encouragement, and unwavering belief in my potential have been the cornerstone of my achievements. I am profoundly grateful for their profound influence and unwavering presence, which have enriched my life in countless ways.

I would like to express my profound gratitude to my thesis advisor, Prof. Dr. Nipon Theera-Umpon at Chiang Mai University, for his unwavering moral support, invaluable guidance, and insightful suggestions throughout the course of this study. His encouragement has been instrumental in shaping this research endeavor. Additionally, I am deeply grateful to Assoc. Prof. Dr. Sansanee Auephanwiriyakul for her inspiration and guidance, which have enriched this work.

I am indebted to numerous individuals, particularly Dr. Ritipong Wongkhuenkaew, whose invaluable contributions have been instrumental in the culmination of this thesis. I extend my heartfelt gratitude to my examination committee for their invaluable guidance, encouragement, and unwavering patience throughout this transformative journey. Their sage advice and steadfast support have been a source of immense inspiration and motivation, guiding me towards the successful completion of this endeavor.

Special thanks are due to the Computational Intelligent Research Laboratory, Faculty of Engineering, Chiang Mai University, for generously providing the necessary resources and infrastructure, without which this study would not have been feasible. I am also grateful to my colleagues at the Computational Intelligent Research Lab for their support and camaraderie.

Furthermore, I am deeply grateful to my family for their unwavering support, patience, and love. Their constant encouragement and inspiration have been a source of strength throughout this endeavor. I am particularly thankful to my parents for their unending support and motivation. Words cannot adequately express my gratitude to them.

Rattanathon Phettom

หัวข้อวิทยานิพนธ์ การระบุเสียงผิดปกติของปอดแบบอัตโนมัติด้วยวิธีการเรียนรู้ของเครื่อง

ผู้เขียน นาย รัตนธร เพ็ชรถม

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมชีวการแพทย์)

อาจารย์ที่ปรึกษา ศ. ดร. นิพนธ์ ธีรอำพน

บทคัดย่อ

งานวิจัยนี้นำเสนอการระบุเสียงผิดปกติของปอดจากสัญญาณเสียงแบบอัตโนมัติโดยใช้การวิเคราะห์ความถี่เชิงเวลา และโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยใช้สัญญาณเสียงที่บันทึกได้จากการใช้สเต็ทโตสโคป ซึ่งจะถูกกำจัดสัญญาณรบกวนด้วยการใช้ตัวกรองความถี่แถบผ่าน จากนั้นนำไปสกัดคุณลักษณะเด่นโดยใช้การแปลงฟูริเยร์แบบช่วงเวลาสั้นเพื่อให้ได้องค์ประกอบทางความถี่ในรูปแบบสเปกโตรแกรม จากนั้นสเปกโตรแกรมจะถูกนำมาตรวจหาเพื่อแบ่งรอบการหายใจโดยอาศัยการหายอดที่สูงที่สุดและต่ำที่สุด เพื่อให้ทราบจำนวนรอบการหายใจตลอดสัญญาณเสียง จากนั้นรอบการหายใจทั้งหมดจะถูกแบ่งเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ โดยโครงข่ายประสาทเทียมแบบคอนโวลูชันจะอาศัยข้อมูลฝึกสอนดังกล่าวในการเรียนรู้เพื่อให้ได้ต้นแบบที่ดีที่สุด จากผลการทดลองด้วยวิธีการที่นำเสนอ สามารถจำแนกเสียงการหายใจแบบหวีด เสียงแซมการหายใจ และเสียงการหายใจแบบปกติ ได้อย่างมีประสิทธิภาพโดยมีความถูกต้องที่ระดับร้อยละ 85.34 ร้อยละ 68.20 และร้อยละ 60.64 ตามลำดับ

Thesis Title	Automatic Identification of Abnormal Lung Sounds by Machine Learning Methods
Author	Mr. Rattanathon Phettom
Degree	Master of Engineering (Biomedical Engineering)
Advisor	Prof. Dr. Nipon Theera-Umpon

ABSTRACT

This study introduces an automated approach for identifying abnormal lung sounds from audio recordings utilizing time-frequency analysis and convolutional neural networks. Acoustic signals captured via a stethoscope are subjected to noise removal using a bandpass filter. Subsequently, distinctive features are extracted via a short-time Fourier transform to represent frequency components in the form of a spectrogram. The spectrogram facilitates the segmentation of breathing cycles by identifying the highest and lowest peaks, thereby quantifying the number of breathing cycles within the audio signal. Following this segmentation, the breathing cycle is partitioned into training and test datasets, with the convolutional neural networks trained on the former to optimize model performance. Experimental findings demonstrate that the proposed method effectively achieves the accuracies of 85.34 percent, 68.20 percent, and 60.64 percent for wheezing sounds, crackle sounds, and normal sounds, respectively.

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBERVIATIONS

ALTD	Adaptive Local Trigonometric Decomposition
AMIE_SEG	Adaptive Multi-Level In-Exhale Segmentation
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
Aux	Auxiliary classifiers
CAD	Coronary Artery Disease
CLSA	Computerized Lung Sound Analysis
CNNs	Convolutional Neural Networks
COPD	Chronic Obstructive Pulmonary Diseases
CRV	Cross Validation
CV	Conventional Validation
DFSS	Dynamic Fuzzy Neural Networks
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
DTFT	Discrete-Time Fourier transform
DWT	Discrete Wavelet Transform
EGST	Enhanced Generalized S-Transform
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
FCSM	Fisher's Class Separability Measure
FFT	Fast Fourier Transform
GDA	Generalized Discriminant Analysis

GMM	Gaussian Mixture Models
HHS	Hilbert-Huang Spectrum
HRV	Heart Rate Variability
IDFT	Inverse Discrete Fourier Transform
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IMD	Intrinsic Mode Function
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficients
MLP	Multi-Layer Perceptron
MSWP	Multiscale Wavelet Packet
NSR	Normal Sinus Rhythm
PE	Permutation Entropy
PSD	Power Spectral Density
SBC	Subband Based Cepstral
SDR	Software-Defined Radio
Self_NSR	Self-recorded data as Normal Sinus Rhythm
SSA	Singular Spectrum Analysis
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TFR	Time-Frequency Representation
VQ	Vector Quantization
WPD	Wavelet Packet Decomposition
WPE	Wavelet Packet Energy
WPT	Wavelet Packet Transform

CHAPTER 1

Introduction

1.1 Background and Motivation

Respiratory ailments are often assessed by analyzing lung sounds, employing a diagnostic tool as a stethoscope instrument to detect irregular sounds. Precise identification of these sounds is essential for accurate diagnosis and subsequent treatment. Although auscultation via the stethoscope is a straightforward method, precise diagnosis demands expertise, posing potential difficulties for inexperienced practitioners. In light of this, our objective is to devise an algorithm that can autonomously distinguish between normal breathing sounds and abnormal breathing sounds, particularly categorizing them as either wheezes or crackles. The deployment of such an algorithm offers numerous potential advantages, such as streamlining the subjective assessment process for healthcare professionals in discerning normal and abnormal lung sounds, and diminishing the bias linked with subjective evaluations reliant on observer expertise. This study delves into the prevalence and clinical relevance of crackles and wheezing [1] in respiratory disorders, leveraging recorded lung sound signals for analysis.

Breath sounds represent a vital sign originating from the thoracic region during the process of inhalation and exhalation [2]. These sounds are readily perceptible in tranquil surroundings or when individuals consciously attend to their breathing. The act of respiration is orchestrated by the synchronized movements of breathing muscles, generating breath sounds as air traverses the air passages. Each complete sequence of inhalation and exhalation constitutes a breathing cycle. The principal muscle responsible for respiration is the diaphragm, which contracts during inhalation, thereby augmenting lung volume and facilitating air intake. Conversely, during exhalation, as the diaphragm releases, lung volume diminishes, and air flows out from the lungs. Consequently, inhalation and exhalation emerge, delineating distinct phases within the breathing cycle.

The diagnostic procedure for respiratory conditions typically entails auscultation, a method in which healthcare practitioners assess the lungs by listening to specific areas on the chest wall, including the anterior, posterior, and lateral regions. The presence of unusual sounds during this examination, known as adventitious sounds (abnormal sounds), includes manifestations like crackles and wheezing, characterized by fluctuations in frequency, pitch, intensity, and energy. Analyzing these adventitious sounds offers crucial insights for diagnosing lung conditions. Therefore, the principal aims of this study are twofold: firstly, to develop an algorithm capable of differentiating between normal and abnormal breathing sounds by identifying lung sound cycles. Secondly, to accurately classify abnormal sounds, such as crackles or wheezing, based on recorded lung sound data sourced from a respiratory dataset. To ensure precision, all sound files undergo noise elimination processes before being segmented into individual lung sound cycles.

Finally, Convolutional Neural Networks (CNNs) are deployed to discern abnormal sounds within these cycles. This algorithmic strategy facilitates the swift and accurate detection of anomalous breath sounds, thereby enhancing diagnostic proficiency in the realm of respiratory medicine.

This thesis is structured as follows: Chapter 2 provides background information on respiratory sound processing and reviews related works in the field. In chapter 3, the experimental framework employed in this research is elaborated upon. Chapter 4 presents the experimental results and subsequent discussion. Ultimately, the concluding remarks of this thesis are encapsulated in the concluding chapter 5.

1.2 Literature Review

The automatic acoustic identification of respiratory sounds holds promise for aiding healthcare professionals in the classification of diseases pertaining to the human respiratory system, such as pneumonia, asthma, and Chronic Obstructive Pulmonary Diseases (COPD) [3]. These diseases manifest distinct acoustic patterns discernible during the auscultation of lung sounds. The classification tasks in this domain can be broadly delineated into two categories: disease classification [3] and abnormal sound classification [4]. Concurrently, there is a burgeoning need for novel and simplified methodologies aimed at detecting respiratory diseases from lung sounds. Examples include the development of a robust deep learning framework [5], the isolation of lung sounds from mixed heart and lung sounds [6], and the investigation into the non-linearity and non-stationarity nature of lung sound signals [7].

In light of the challenges posed, researchers have endeavored to devise innovative algorithms aimed at discerning between normal breath sounds and abnormal breath sounds, notably crackles and wheezing. While identifying normal breath sounds poses relatively fewer complexities, crackle and wheezing sounds can manifest in various lung lobes, including the trachea. Wheezing sounds, characterized by extended duration and heightened loudness, typically exhibit frequencies ranging from 250 to 400 Hz. In contrast, crackle sounds manifest as continuous popping sounds throughout the breath cycle and can occur across a broad spectrum of frequencies within the lung sound spectrum [8]. A plethora of studies have been conducted to differentiate between normal and abnormal lung sounds, employing diverse techniques and algorithms. Examples include the utilization of the Hough transform of spectrograms [9], Wavelet Packet Decomposition (WPD) [10], the Adaptive Multi-level In-Exhale Segmentation (AMIE_SEG) technique [11], and time-expanded waveform analysis [12]. These advancements have laid the groundwork for more precise and automated analyses of respiratory sounds, thereby enhancing the diagnosis and treatment of respiratory diseases.

Several research endeavors have concentrated on diagnosing diseases based on lung sounds, with a particular emphasis on mitigating noise interference, notably from heart sounds [13, 14]. Moreover, investigations have delved into noise removal methodologies, particularly in scenarios where heart sounds encroach upon lung sounds [15-17]. These scholarly inquiries have played a pivotal role in advancing noise elimination techniques and precise disease diagnosis via lung sounds, effectively tackling the hurdles posed by overlapping heart sounds and diverse ambient noises.

Moreover, the normal breath sounds, also known as vesicular breath sounds, exhibit distinct characteristics that differ from abnormal lung sounds such as crackles and wheezing, as illustrated below [8].

In Figure 1.1, the normal breath sound is distinguished by its gentle, low-pitched quality with a rustling characteristic. During inhalation, it becomes more audible and persists longer compared to exhalation, with a ratio of expiration to inspiration approximately at 1:3. As depicted in Figure 1.2, crackle sounds are typified by their popping, low-pitched features accompanied by a bubbling quality. They exhibit increased volume and duration compared to fine crackles and can be perceived during both inhalation and exhalation. Wheezing, illustrated in Figure 1.3, manifests as a continuous sound that can vary in pitch, ranging from high-pitched (squeaking) to low-pitched (snoring or moaning). This phenomenon arises from airway constriction, resulting in an elongated wheezing phase experienced during both inhalation and exhalation, where Y axis is an amplitude of normalized in dB and X axis is time in second.

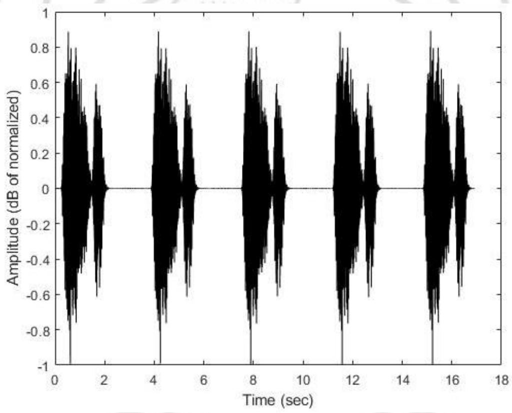

Figure 1.1 Signal of normal breath sound

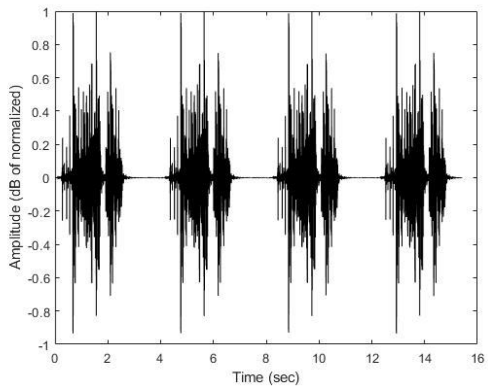

Figure 1.2 Signal of crackle sound

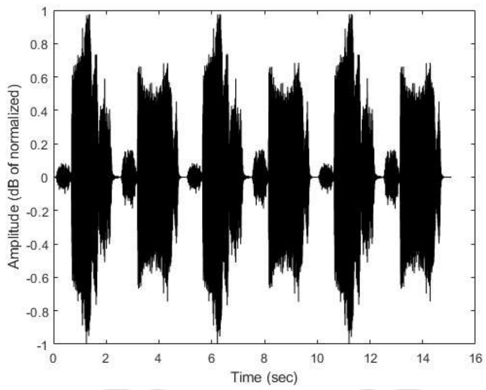

Figure 1.3 Signal of expiration wheezing sound

These descriptions offer a comprehensive understanding of the distinguishing features of vesicular breath sounds, crackles, and wheezing. Familiarity with these distinct sound profiles can significantly assist healthcare professionals in accurately identifying and diagnosing various respiratory conditions during the auscultation process.

The subsequent investigations presented diverse objectives, with certain endeavors aimed at contrasting machine learning methodologies for the detection of various respiratory phenomena such as wheezing, crackling, simultaneous occurrence of crackling and wheezing, cardiac and pulmonary disorders, inspiratory and expiratory events, or instances of coughing. Others sought to delineate methodologies concerning the differentiation between normal and abnormal respiratory patterns, feature extraction techniques, noise reduction strategies, and classification methodologies.

Oweis RJ, Abdulhay EW, Khayal A, and Awad A [7] showed the comparison of Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) toolboxes that were applied to 10 lung sounds to classify them. The result showed that accuracy, specificity, and sensitivity of ANN was better than ANFIS of 98.6, 100, and 97.8%, respectively.

Zhang KX, Long Z, Wang XF, and Zhao H [9] used Hough transform of spectrogram to detected wheezing sound. The approach initially employed Canny edge detection operator to identify edges followed by used Hough transform to analyze wheezing from the international shared lung sound files. The results of detection show 87% accuracy of 60 wheezing cases, and 74% accuracy of 70 normal cases.

Zhang J, Wang HS, Zhou HY, Dong B, Zhang L, Zhang F, et al. [10] performed feature extraction WPD of crackles and wheeze sounds. And then data were trained by a Support Vector Machine (SVM). Results show an accuracy, sensitivity, and specificity of 90.3, 88.3, and 92.3% of crackles and 87.1, 86.7, and 87.5% of wheezing, respectively.

Chen H, Yuan X, Li J, Pei Z, and Zheng X [11] enhanced the features of wheezing for classification by AMIE_SEG that is for extracting inspiration and expiration. And wheezing will be extracted by Enhanced Generalized S-Transform (EGST) feature. After that the authors employed three machine learning-based classifiers: SVM, Extreme Learning Machine (ELM) and K-Nearest Neighbor (KNN). The results show KNN is the best method with accuracy, sensitivity, specificity as 98.62%, 95.9% and 99.3% in average respectively.

Sovijarvi ARA, Helisto P, Malmberg LP, Kallio K, Paajanen E, Saarinen A, et al. [12] conducted a study on a respiratory sound analyzer system designed to automatically analyze crackle sounds. The analysis involved several techniques, including phono pneumography, time-expanded waveform analysis, spectral analysis utilizing time-averaged Fast Fourier Transform (FFT), frequency analysis in the time domain (sonogram), and automatic detection and waveform analysis of crackles. The study concluded that the median frequency exhibited the best repeatability among quartile frequencies of breath sounds.

Chowdhury SK and Majumder AK [13] used FFT to describe for spectrum analysis of respiratory sounds of six normal cases and six patients with tuberculosis. The result shows that the amplitude of tuberculosis patients was shift to 1,000 Hz from normal cases at 250 Hz.

Fraiwan L, Hassanin O, Fraiwan M, Khassawneh B, Ibnian AM, and Alkhodari M [14] identified asthma, heart failure, pneumonia, bronchiectasis, bronchitis, and COPD. Initially, features such as Shannon Entropy, Logarithmic Energy Entropy, and Spectrogram-Based Spectral Entropy were extracted. These features were then used to train both baseline models, including SVM, KNN, Decision Tree and Linear Discriminant Analysis and ensemble models which are Bagged Decision Tree, Bagged Linear Discriminants, Boosted Decision Trees, and Boosted Linear Discriminants. After that performed evaluation of accuracy, sensitivity, specificity, F1-score, and Cohen's kappa correlation. The results show the best performance is the boosted decision tree model of ensemble classification as demonstrated by its highest 98.20% accuracy, 91.50% sensitivity and 98.55% specificity.

Ghaderi F, Mohseni HR, and Sanei S [15] located heart sound components in mixed between heart and respiratory sound by using Singular Spectrum Analysis (SSA) to show the frequency of heart and lung sound signals that normally overlap in each other. The result compared with well-established methods; wavelet transform, and entropy of the signal and result of heart detection showed that wavelet transform was slightly better than entropy-based method.

Iyer VK, Ramamoorthy PA, Fan H, and Ploysongsang Y [16] investigated technique using adaptive filtering to reduce heart sound out from mixed between breath sound and heart sound. The result showed percent of heart sound removing is 50 to 80 percent.

Suzuki A, Sumi C, and Nakayama K [17] used adaptive filtering technique to remove ambient noises; environment sound, device noise, human voice, etc. that disrupted while recording the lung sounds. This method can remove the noises by about 30 dB using a 256-tap filter with the convergence time of several seconds and, it is very effective for a lung sound analysis preprocessing tool by real-time processing.

Kandilogiannakis G, Mastorocostas P, and Varsamis D [18] separated the abnormal sounds of the lungs from the vesicular breath sounds by a computational intelligence-based filter that used two operating in parallel Dynamic Fuzzy Neural Networks (DFNN) to perform the tasks.

Kok XH, Anas Imtiaz S, and Rodriguez-Villegas E [19] used feature extractions and Wilcoxon Rank Sum statistical test to identify respiratory disease from recording files. The results of training achieved accuracy of 87.1%, sensitivity of 86.8% and specificity of 93.6%.

Li S and Liu Y [20] present a vector of feature extraction of normal, pneumonia and asthma lung sounds based on bispectrum that is 2-D Fourier transform of third order cumulants.

Manir SB, Karim M, and Kiber MA [21] performed Digital Signal Processing (DSP) methods to perform various features such as RMS, Zero Crossings, Turn Count, Mean, Variance and Form Factor.

Bahoura M and Pelletier C [22] classified between normal and wheezing lung sounds with Gaussian Mixture Models (GMM) method. Mel-Frequency Cepstral Coefficients (MFCC) or Subband Based Cepstral (SBC) parameters characterize overlapped signal segments and then compare with other Vector Quantization (VQ) and Multi-Layer Perceptron (MLP) neural networks.

Hsu FS, Huang CJ, Kuo CY, Huang SR, Cheng YR, Wang JH, et al. [23] demonstrated a lung sound labeling algorithms to classify inspiration, expiration, and abnormal sounds with six feature vectors. The result show F1-scores of 86.0% on inspiration task, 51.6% on continuous abnormal sound task and 71.4% on discontinuous abnormal sound task.

Neili Z, Fezari M, and Redjati A [24] compared the ability of ELM and KNN machine learning algorithms between normal and abnormal lung sounds. First, the authors used Empirical Mode Decomposition (EMD) to analyze lung sounds. Then into Intrinsic Mode Functions (IMD). The Hjorth descriptors (Activity) and Permutation Entropy (PE) are the features from each IMFs and then combined. The results show an accuracy of 90.71 and 95.00% using ELM and KNN, respectively.

Kumar A, Vincent DRPM, Srinivasan K, Chang C-Y [25] combined Deep Learning and Machine Learning techniques in infant cry activities; hungry, pain, and sleep cries base on SVM, Naïve Bayes and KNN. The results show accuracy of SVM at 93.3%, Naïve Bayes at 86.6% and KNN at 88.3%.

Vashkevich R and Azarov E [26] used pitch-invariant convolutions on frequency axis of amplitude spectrum in speech processing to detect activity of voice. The result shows a comparison with publicly available voice activity detection model from the WebRTC showed higher F1 scores (0.94 versus 0.87).

Reyes BA, Charleston-Villalobos S, González-Camarena R, and Aljama-Corrales T [27] sought a technique to obtain a Time-Frequency Representation (TFR) of thoracic sound by comparing general goodness-of-fit criteria in different simulated thoracic sounds scenarios. Time-frequency patterns of thoracic sounds; heart, normal tracheal and adventitious lung sounds were assessed by mathematical functions to find the best TFR. Results showed that the Hilbert-Huang Spectrum (HHS) had a superior performance as compared with other techniques.

Palaniappan R, Sundaraj K, Sundaraj S, Huliraj N, and Revadi S S [28] employed the Wavelet Packet Transform (WPT) to extract energy and entropy features from lung sound signals. The study reported maximum accuracies of 97.36% and 98.37% for Conventional Validation (CV) of the energy and entropy, respectively. Additionally, Cross Validation (CRV) yielded accuracies of 96.80% and 97.91% for energy and entropy, respectively. Furthermore, ensemble features achieved accuracies of 98.25% for CV and 99.25% for CRV, respectively.

Yan J, Shen X, Wang Y, Li F, Xia C, Guo R, et al. [29] performed WPT and SVM algorithm to analysis. The authors employed WPD at level 6 to split more elaborate frequency bands of the auscultation signals. After that analyze statistic based on the extracted Wavelet Packet Energy (WPE) features from WPD coefficients. In additional, mixed subject's statistical feature values of sample groups through SVM was used to be separated by the pattern recognition. Finally, the results showed that the classification accuracies were at a high level.

Singh RS, Saini BS, and Sunkaria RK [30] proposed a novel method for detecting coronary artery disease (CAD) utilizing Heart Rate Variability (HRV) signals. Their approach involved employing Multiscale Wavelet Packet (MSWP) transform and entropy feature extraction to decompose the HRV signals. The detection performance was evaluated using the Fisher ranking method, Generalized Discriminant Analysis (GDA), and a binary classifier known as Extreme Learning Machine (ELM). Results indicated that the proposed approach outperformed other methods, particularly when utilizing the top ten ranked entropy features for dataset combination. The datasets included self-recorded data representing Normal Sinus Rhythm (Self_NSR), healthy Normal Sinus Rhythm (NSR), and CAD patients sourced from a standard database. Notably, the multiquadric method achieved an approximate detection accuracy of 100%, surpassing ELM and linear discriminant analysis.

Ono M, Arakawa K, Mori M, Sugimoto T, and Harashima H [31] identified fine crackle sounds from vesicular sounds by using a nonlinear digital filter that separate nonstationary which is a characteristic of crackle sounds from stationary signals in six participants who were diagnosed with pulmonary fibrosis. The result showed that this method is useful enough in clinical medicine.

Ademovic E, Pesquet JC, and Charbonneau G [32] used the Adaptive Local Trigonometric Decomposition (ALTD) in the time-frequency domain with a lattice in time to identify wheezing from lung sound signals.

Kiyokawa H, Greenberg M, Shirota K, and Pasterkamp H [33] investigated lung fine, medium, and coarse crackle sound detector. The authors computerized analysis of lung sounds within and between physical observers. The results showed the conditions of failed detection that was more common in 1) higher intensity background lung sounds compared to lower intensity background lung sounds, 2) coarse or medium crackles compared to fine crackle and 3) small amplitude compared to large amplitude of crackle sounds.

Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, and Malmurugan N [34] demonstrated that lung sound signals should not use the conventional method of frequency analysis because, its classification is not successful. And the authors showed the wavelet transform analysis method of lung sound signals with ANN classification and trained by the resilient backpropagation algorithm.

Guntupalli KK, Alapat PM, Bandi VD, and Kushnir I [35] investigated to detect wheezing pattern from dynamic image of lung sound on spectral analysis using a computerized automatic stethoscope compared to the physicians in seven subjects with 100 sound files. The overall results showed 84% of the sensitivity inter-rater agreement.

Wang Z and Xiong YX [36] used acoustic device to estimate lung sound patterns using computerized analysis in acute congestive heart failure and improvement patients, normally, this disease presents the adventitious sounds. The result showed the homogenous distribution of lung vibration energy was more increase geographical area of the vibration energy image. And this analysis may be useful to track in acute congestive heart failure recurrence.

Gurung A, Scrafford CG, Tielsch JM, Levine OS, and Checkley W [37] traced meta-analysis of Computerized Lung Sound Analysis (CLSA) for the best specific respiratory disease detectors. The authors forecasted the sensitivity and specificity of CLSA and, found that electret microphones or piezoelectric sensors for auscultation, and Fourier Transform and Neural Network algorithms for analysis and automated classification of lung sounds mostly used. The overall result, sensitivity, and specificity for the detection of wheezes or crackles was 80% and 85% respectively.

Ellington LE, Emmanouilidou D, Elhilali M, Gilman RH, Tielsch JM, Chavez MA, et al. [38] found that distinct spectral and septotemporal signal parameters of age, height, and weight do not make lung sounds difference with genders. Moreover, younger children had a slower decaying spectrum than older children. In conclusion, lung sound characteristics of lung sound features of children varied significantly.

Kosasih K, Abeyratne UR, and Swarnkar V [39] used wavelet analysis of a range up to 90 kHz that above the human perception of 90 cough sound samples from 4 patients. The result showed the R^2 of 77 to 82% at 15 to 90 kHz frequencies and, the R^2 increased to 85 to 90% at frequencies that below 15 kHz.

Kosasih K, Abeyratne UR, Swarnkar V, and Triasih R [40] utilized wavelet features in conjunction with other features such as Mel Cepstral coefficients and non-Gaussian index to detect childhood pneumonia from a dataset comprising 815 cough sounds. Their findings revealed a sensitivity of 94% and specificity of 63%. Furthermore, when combined with the findings of previous research (High frequency analysis of cough sounds in pediatric patients with respiratory diseases, 2012 [39]), the sensitivity increased to 94% and the specificity to 88%.

Haider NS, Joseph J, and Periyasamy R [41] investigated the statistical significance of five different spectral lung sounds (maximum frequency, dominant frequency and spectral centroid that identified from spectra and, median frequency and spectral roll off that computed from the Power Spectral Density (PSD)) that are stridors, wheezing, bronchial, vesicular and crackle using Analysis of Variance (ANOVA) and Fisher's Class Separability Measure (FCSM). The result of preprocessing showed P-values at a confidence level of 0.05 for dominant frequency, maximum frequency and median frequency, spectral roll off and spectral centroid of 0.0386, 0.7508, 0.0197, 0.055 and 0.6979, respectively. And FCSM and ANOVA are 0.1242, 0.0192, 0.1498, 0.1112 and 0.0222, respectively. The median frequency comparatively is more significant than the other four.

Habukawa C, Ohgami N, Matsumoto N, Hashino K, Asai K, Sato T, et al. [42] made a features algorithm following the Computerized Respiratory Sound Analysis guidelines to identify wheezing sounds in 214 children, 2 mouths to 12 years. There were 65 wheezing sounds and 149 without wheezing sounds. The results showed sensitivity, specificity, positive predictive value, and negative predictive value of the wheeze recognition algorithm of 100, 95.7, 90.3, and 100%, respectively.

Naqvi SZH and Choudhry MA [43] developed a novel framework for diagnosing COPD, pneumonia, and normal breath sounds. Their approach involved integrating time domain, cepstral, and spectral features using the back-elimination method, while denoising and segmenting the pulmonic signal were achieved through techniques based on EMD and Discrete Wavelet Transform (DWT). Experimental results demonstrated an impressive accuracy of 99.70% when employing selected fused features.

This section encapsulates key summaries from the literature review, depicting comparative analyses of machine learning approaches in Table 1.1. Additionally, Table 1.2 outlines findings related to the detection of wheezing, crackling, simultaneous occurrences of crackling and wheezing, cardiac and pulmonary disorders, inspiratory and expiratory events, as well as instances of coughing. Furthermore, Table 1.3 provides insights into methods pertaining to the differentiation between normal and abnormal lung sounds, feature extraction, noise reduction, and classification.

Table 1.1 Summary review of machine learning comparing

Literature	Concept	Algorithm	Result
An alternative respiratory sounds classification system utilizing ANN. [7]	Comparing ANNs and ANFIS to classify lung sounds.	ANNs and ANFIS.	ANN is the best by accuracy, specificity, and sensitivity of 98.6, 100.0, and 97.8%, respectively.

Table 1.1 Summary review of machine learning comparing (continue)

Literature	Concept	Algorithm	Result
ELM and K-nn machine learning in classification of Breath sounds signals. [24]	Comparing ELM and KNN to classify lung sounds.	ELM and KNN.	Accuracy ELM of 90.71% and KNN of 95.00%.
Deep CNNs based Feature Extraction with optimized Machine Learning Classifier in Infant Cry Classification. [25]	Combined ML techniques in infant cry activities; hungry, pain, and sleep cries.	SVM, Naïve Bayes and KNN	Accuracy of SVM at 93.3%, Naïve Bayes at 86.6% and KNN at 88.3%.
Classification of pulmonary pathology from breath sounds using the WPT and an ELM. [28]	Comparing energy and entropy feature extraction of lung sounds.	WPT, CV and CRV.	Max accuracy CV of energy and entropy of 97.36% and 98.37%, CRV of energy and entropy of 96.80% and 97.91%, and CV and CRV ensemble feature of 98.25% and 99.25%, respectively.

Table 1.2 Summary review of detection or separation

Literature	Concept	Algorithm	Result
Wheezing Detection			
Detection of Wheeze Based on Hough Transform of Spectrogram. [9]	Separating normal and wheezing lung sounds.	Hough Transform of Spectrogram.	Accuracy for wheezing detection of 87% and for normal lung sounds detection of 74%.

Table 1.2 Summary review of detection or separation (continue)

Literature	Concept	Algorithm	Result
Wheezing Detection			
Automatic Multi-Level In-Exhale Segmentation and EGST for wheezing detection. [11]	Identifying wheezing lung sounds.	AMIE_SEG and EGST feature extraction. And, be trained by SVM, ELM, and KNN.	KNN is the best at accuracy, sensitivity, and specificity of 98.62%, 95.9%, and 99.3% by mean.
Respiratory sounds classification using cepstral analysis and GMM. [22]	Showing method to separate normal and wheezing lung sounds.	GMM (MFCC or SBC vectors) and compare to VQ and MLPNW.	The performance is better than nonuse.
Wheezing lung sounds analysis with adaptive local trigonometric transform. [32]	Showing method to identify wheezing from lung sound signals.	ALTD in time-frequency domain.	Notice that wheeze (0.5 s <t< 2 s) is well rebuilt and composed of large segments. And the inspiration signal just after (2.1 s <t< 3.7 s).

Table 1.2 Summary review of detection or separation (continue)

Literature	Concept	Algorithm	Result
Wheezing Detection			
Validation of automatic wheeze detection in patients with obstructed airways and in healthy subjects. [35]	Detecting wheezing from dynamic image of lung sound on spectral analysis computer compared to physicians.	Fourier transform (spectral analysis)	Inter-rater sensitivity of 84%.
A wheeze recognition algorithm for practical implementation in children. [42]	Identifying wheezing using feature algorithm following the Computerized Respiratory Sound Analysis guidelines.	Wheezes features following the Computerize d Respiratory Sound Analysis guidelines	Sensitivity, specificity, positive predictive value, and negative predictive value of 100%, 95.7%, 90.3%, 100%, respectively.
Crackle Detection			
A new versatile PC-based lung sound analyzer with automatic crackle analysis (HeLSA); repeatability of spectral parameters and sound amplitude in healthy subjects. [12]	Investigating system to analyze crackle sounds.	FFT.	Median frequency is the best repeatability of quartile frequencies of breath sounds.

Table 1.2 Summary review of detection or separation (continue)

Literature	Concept	Algorithm	Result
Crackle Detection			
Separation of Fine Crackles from Vesicular Sounds by a Nonlinear Digital Filter. [31]	Identifying fine crackle sound from vesicular sound.	Nonlinear digital filter.	It is useful enough in clinical medicine.
Auditory detection of simulated crackles in breath sounds. [33]	Investigating crackle detector and analyze within and between physical observers.	MATLAB; MathWorks; Natick, MA (simulated)	Result showed condition of filed detections.
Crackle and Wheezing Detection			
Real-World Verification of Artificial Intelligence Algorithm-Assisted Auscultation of Breath Sounds in Children. [10]	Identifying crackles and wheezing lung sounds.	Feature extraction WPD and be trained by SVM.	Accuracy, sensitivity, and specificity of crackles of 90.3%, 88.3%, and 92.3% and wheezing of 87.1%, 86.7, and 87.5%, respectively.
Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis. [37]	Detecting wheezing and crackles using CLSA.	Fourier Transform and Neural Network algorithms.	Sensitivity of 80% and specificity of 85%.

Table 1.2 Summary review of detection or separation (continue)

Literature	Concept	Algorithm	Result
Heart and Lung Sounds Detection			
Localizing heart sounds in respiratory signals using SSA. [15]	Separating heart and lung sounds.	SAA (wavelet transform and entropy)	The best algorithm for separating is wavelet transform.
Reduction of Heart Sounds from Lung Sounds by Adaptive Filtering. [16]	Reducing heart sound out of breath sounds.	Adaptive filter.	Removing heart sound of 50% to 80%.
Respiratory Diseases Detection			
Digital Spectrum Analysis of Respiratory Sound. [13]	Separating spectrum of normal and tuberculosis lung sounds.	Fourier Transform.	Normal spectrum is 250 Hz and tuberculosis is 1,000 Hz.
Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. [14]	Identifying lung diseases (asthma, heart failure, pneumonia, bronchiectasis, bronchitis, and COPD).	Shannon Entropy, Logarithmic Energy Entropy and Spectrogram-Based Spectral Entropy.	Boosted decision tree of ensemble class is the best accuracy of 98.20%, sensitivity of 91.50%, and specificity of 98.55%.
A Novel Method for Automatic Identification of Respiratory Disease from Acoustic Recordings. [19]	Identifying lung diseases.	Feature extraction and WRS.	Accuracy of 87.1%, sensitivity of 86.8%, and Specificity of 93.6%.

Table 1.2 Summary review of detection or separation (continue)

Literature	Concept	Algorithm	Result
Inspiration and expiration Separation			
Development of a respiratory sound labeling software for training a deep learning-based respiratory sound analysis model. [23]	Separating inspiration, expiration and abnormal lung sounds.	Six feature extractions.	F1-scor inspiration and expiration of 86.0%, continuous abnormal of 51.6%, and discontinuous abnormal of 71.4%.
Coughing Detection			
High frequency analysis of cough sounds in pediatric patients with respiratory diseases. [39]	Detecting coughing frequency.	Wavelet analysis.	R^2 of 77% - 82% at 15 - 90 kHz and it increased to 85% - 90% at below 15 kHz.
Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis. [40]	Detecting coughing sound and combined with other features to identify childhood pneumonia.	Wavelet features and others feature extraction.	Sensitivity of 94% and specificity of 63%. And, combining with (39)'s work sensitivity and specificity were improved to 94% and 88%, respectively.

Table 1.3 Summary review of relevant method

Literature	Concept	Algorithm	Result
Adaptive Cancelling of Ambient Noise in Lung Sound Measurement. [17]	Showing method to remove ambient noises while recording lung sound.	Adaptive filter.	It is very effective for lung sound analysis preprocessing tool by real-time.
A computational intelligence-based filter for lung sound separation. [18]	Showing method of identifying normal and abnormal lung sounds.	DFNN	Filter was efficient separation performance and capable in real-time.
Feature Extraction of Lung Sounds Based on Bispectrum Analysis. [20]	Showing method of identifying normal and lung disease.	2-D Fourier transform of third order cumulants.	Proper features can be extracted from bispectrum of lung sounds to form the feature vector for classification.
Assessment of Lung Diseases from Features Extraction of Breath Sounds Using Digital Signal Processing Methods. [21]	Showing feature extraction of lung sounds.	RMS, zero crossing, turn count, mean, variance, and form factor.	System can diagnose lung conditions by integrating artificial intelligence.

Table 1.3 Summary review of relevant method (continue)

Literature	Concept	Algorithm	Result
Assessment of TFR techniques for thoracic sounds analysis. Computer methods and programs in biomedicine. [27]	Finding technique to obtain TFR of chest region sounds and compare to goodness-of-fit criteria in different simulated thoracic sounds scenarios.	The general class, TVAR modeling and the instantaneous power spectrum, the scalogram, and the Hilbert–Huang spectrum.	HHS is the best technique as compared to others technique.
Detection of CAD by reduced features and ELM. [30]	Showing method of detection of CAD from HRV.	MSWP and entropy feature extraction.	Multiquadric accuracy of 100% as compared to ELM and linear discriminant analysis.
An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis. [43]	Preforming a new framework to diagnose COPD, pneumonia, and normal breath sound.	Time domain, cepstral, and spectral through the back-elimination method next EMD and DWT-based techniques used to denoise and segment.	Accuracy of 99.7%.

1.3 Purposes of The Study

The purposes are to devise an algorithm capable of discerning lung sound cycles, effectively distinguishing between normal and abnormal lung sounds. Specifically, in instances of abnormal lung sounds, the algorithm will adeptly classify them as either crackles or wheezing based on the recorded lung sound data.

1.4 Research Scopes

1.4.1 Focus on respiratory sounds, specifically wheezing sounds, crackle sounds, and normal sounds.

1.4.2 The dataset employed for the analysis of respiratory sounds in this study originates from collaborative efforts between two research teams in Portugal and Greece, known as the "A Respiratory Sound Database for the Development of Automated Classification."

1.5 Education Advantages

The primary aim of this study is to streamline the process of differentiating between normal and abnormal lung sounds, thereby reducing the subjective examination time required by medical personnel. Additionally, the study seeks to alleviate any bias associated with observer experience during the assessment of lung sounds. Additionally, the proposed methodology has the potential to evolve into a continuous monitoring device that utilizes chest wall-attached stethoscopes. This innovation obviates the necessity for medical staff to manually operate a stethoscope, which proves particularly advantageous in scenarios involving contagious diseases. By minimizing the proximity of medical personnel to patients, the risk of disease transmission is significantly mitigated, thereby enhancing overall safety measures.

1.6 Research Methodologies

1.6.1 Thesis Preparation: Define research question or problem, literature review, hypothesis or thesis statement, research methodology, proposal defense.

1.6.2 Data Acquisition: Source of data, ethical considerations.

1.6.3 Data Preparation: Cleaning and preprocessing, data formatting.

1.6.4 Feature Extraction: Select relevant features, feature engineering.

1.6.5 Training and Classification: Model selection, training, classification.

1.6.6 Performance Evaluation Criteria: Metrics, evaluation plan.

1.6.7 Evaluation, Publication, and Thesis Writing: Evaluation, publication, thesis writing, revisions, final defense.

1.7 Organization of Thesis

This thesis is meticulously organized into five comprehensive chapters, each contributing uniquely to the overarching exploration. The breakdown of chapters is as follows: Chapter 1 provides a contextual introduction, laying the groundwork for the ensuing exploration. It offers a concise overview of the research problem, the rationale for its significance, and an outline of the subsequent chapters. Chapter 2 delves into the foundational principles and theories intrinsic to the study. Specifically, it elucidates the theoretical underpinnings related to digital signal and image processing techniques. This chapter serves as a theoretical anchor, establishing the conceptual framework for the subsequent empirical investigations. Chapter 3 meticulously delineates the research designs employed and expounds upon the intricacies of the proposed methodology. It provides a detailed account of the chosen research designs, outlining the steps taken in the pursuit of the research objectives. This chapter serves as a bridge between theory and application, elucidating the strategies employed in the study. Chapter 4 dedicates to the comprehensive presentation and analysis of the experimental results derived from the application of the proposed methodology to a standard dataset. It provides insights into the empirical outcomes, facilitating a nuanced understanding of the method's effectiveness in practice. The final chapter, chapter 5, serves as the culminating section of the thesis. Herein, conclusions drawn from the entire research endeavor are presented. This chapter encapsulates the key findings, their implications, and potential avenues for future research. It provides a thoughtful synthesis of the entire study, underscoring the significance of the undertaken research in the broader academic landscape.

CHAPTER 2

Background and Motivation

In this section, the investigators elucidate the fundamental principles governing various types of digital filters, encompassing low-pass, high-pass, band-pass, and band-stop filters, alongside the utilization of the Butterworth filter, which was employed in this research to attenuate heart sounds and seamlessly eliminate environmental noise. Additionally, logarithmic compression was employed to not only compress low-amplitude signals, which typically lack informative content, but also amplify signals containing pertinent information, thereby enhancing the visibility of lung sound cycles in the time domain. Moreover, owing to the discrete nature of the digital data utilized in this study, the researchers employed the Discrete Fourier transform (DFT) to facilitate the conversion of signals from the time domain to the frequency domain. Furthermore, given the discrete nature of the digital data, a discrete-time STFT was employed to identify lung cycles. For the training phase of the study, the GoogLeNet model, a pre-trained CNNs tool, was utilized. Subsequently, performance evaluations were conducted utilizing a confusion matrix, which facilitated the extraction of key metrics including accuracy, precision, sensitivity, specificity, F1-score, and correlation.

2.1 Digital Filters

Digital filters play a crucial role in DSP [44, 45], serving two primary functions: signal separation and signal restoration. Signal separation becomes necessary when a signal becomes contaminated with interference, noise, or other signals. For instance, consider a scenario where a device measures the electrical activity of a baby's heart while in the womb; the raw signal is likely to be corrupted by the mother's breathing and heartbeat. In such cases, filters are employed to isolate these signals, enabling individual analysis.

Signal restoration, on the other hand, is employed when a signal has undergone distortion. For instance, audio recordings made with subpar equipment may undergo filtration to accurately represent the original sound. Similarly, deblurring images captured with improperly focused lenses or shaky cameras necessitates signal restoration.

Both analog and digital filters can address these challenges, yet digital filters offer significantly superior performance. For instance, a low-pass digital filter may exhibit a gain of 1 ± 0.0002 from Direct Current to 1000 Hz, with a gain of less than 0.0002 for frequencies above 1001 Hz, all within a narrow transition band of just 1 Hz. In contrast, analog filters are limited by factors such as the accuracy and stability of the electronic components, such as resistors and capacitors.

In DSP, it is conventional to refer to a filter's input and output signals as being in the time domain, given that signals are typically sampled at regular time intervals. However, sampling can also occur in space, where readings are taken at equal spatial intervals. Despite this, time domain remains the most prevalent in DSP, with the term "time domain" often encompassing any domain in which the samples are collected.

The most direct method of implementing a digital filter is by convolving the input signal with the filter's impulse response, allowing for the creation of all possible linear filters. Filter designers often refer to the impulse response as the "filter kernel" when employing it in this manner.

2.1.1 Low-Pass filter

A digital low-pass filter is a type of filter used in DSP to attenuate high-frequency components of a signal while allowing low-frequency components to pass through. It's commonly used for smoothing signals, removing noise, and performing anti-aliasing in applications such as audio processing, communications, and control systems. The general equation [46] for a digital low-pass filter in the time domain:

$$y(n) = \sum_{k=0}^{N-1} b_k \cdot x(n-k), \tag{1}$$

where $y(n)$ is the output signal at sample n, $x(n)$ is the input signal at sample n, b_k are the filter coefficients, and N is the filter order.

2.1.2 High-Pass, Band-Pass and Band-Stop Filters

High-pass, band-pass, and band-stop filters are typically designed by initially creating a low-pass filter and then transforming it into the desired response. Consequently, discussions on filter design often focus on low-pass filters, with examples provided accordingly. The conversion from low-pass to high-pass filters can be accomplished through two methods: spectral inversion and spectral reversal, both of which are equally effective.

2.1.3 Butterworth Filter

The Butterworth filter [47], also known as a maximally flat magnitude filter, is a signal processing filter designed to maintain a frequency response that is as flat as possible within the passband. This filter was initially introduced by British engineer and physicist Stephen Butterworth in 1930 through his paper titled "On the Theory of Filter Amplifiers."

In his research, Butterworth demonstrated that by increasing the number of filter elements with appropriate values, progressively closer approximations to the desired response could be achieved. During that era, filters often exhibited significant ripple within the passband, and the selection of component values involved considerable interaction. Butterworth's key insight was the design of a low-pass filter with a normalized cutoff frequency of 1 radian per sec, resulting in a frequency response (gain: G) expressed by the following equation:

$$G(\omega) = \frac{1}{\sqrt{1+\omega^{2n}}}, \qquad (2)$$

where ω represents the angular frequency in radians per sec, and n denotes the number of poles in the filter, which is equivalent to the number of reactive elements in a passive filter. When ω equals 1, the amplitude response of this filter type in the passband is $\frac{1}{\sqrt{2}}$, approximately 0.7071, corresponding to half power or −3 dB.

In his paper, Butterworth exclusively focused on filters with an even number of poles. It is possible that he was unaware of the potential to design such filters with an odd number of poles. Butterworth constructed his higher-order filters by combining 2-pole filters separated by vacuum tube amplifiers. The frequency response plots of 2, 4, 6, 8, and 10 pole filters are represented as A, B, C, D, and E, respectively, in the original graph depicted in Figure 2.1 [47] of his work.

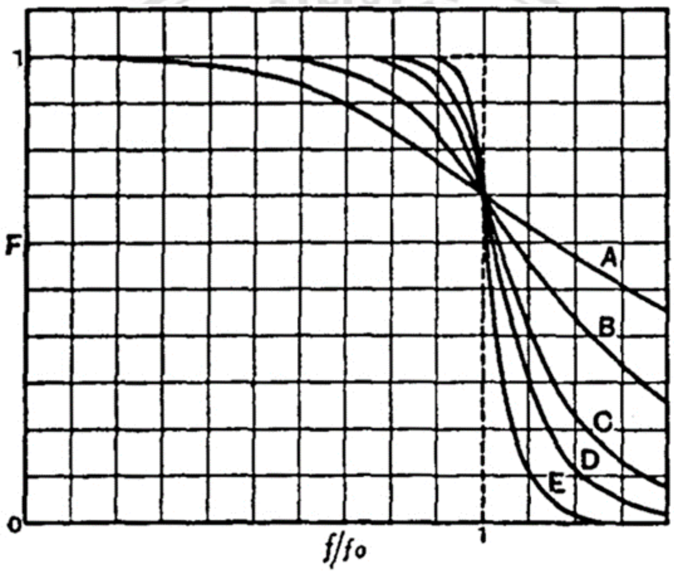

Figure 2.1 The frequency response plot [47]

Butterworth's work further demonstrated that the fundamental low-pass filter could be adapted to provide low-pass, high-pass, band-pass, and band-stop functionality.

In this study, the second-order band-pass Butterworth filter was employed to attenuate frequencies below 200 Hz associated with heart sounds and frequencies above 2,000 Hz related to environmental noise, ensuring a smooth removal process.

2.2 Logarithmic Compression

In the realm of music signal processing [48], representations like spectrograms or chromograms encounter a challenge due to their values exhibiting a wide dynamic range.

This can lead to small yet significant values being overshadowed by larger ones. To address this issue, a dB scale is often employed to mitigate the discrepancy, aiming to reduce the gap between large and small values while accentuating the latter. Additionally, alternative logarithm-based functions may be applied, a process commonly known as logarithmic compression.

Let γ denote a real positive adjustable parameter (where higher γ values result in more aggressive compression) and let y_γ represent the compression output function. It is defined by the following equation:

$$y_\gamma = \log(1+\gamma x), \tag{3}$$

where x is the input signal. In this study, logarithmic compression was employed not only to compress low amplitudes, which typically contain less information, but also to amplify the amplitudes containing significant information. By doing so, the lung sound cycle in the time domain could be observed more clearly, enhancing the discernibility of relevant features.

2.3 Fourier Transform

The Fourier transform is a fundamental analysis technique that decomposes a complex-valued function into its constituent frequencies and their corresponding amplitudes. Its inverse process, synthesis, reconstructs the original function from its transformed representation. In this study, given the digital nature of the data, characterized by discrete signals, the researchers utilized the DFT to transition the signal from the time domain to the frequency domain.

The DFT operates by converting a finite sequence of equidistant samples of a function into an equivalent sequence of equidistant samples of the Discrete-Time Fourier transform (DTFT), a complex-valued function of frequency. The sampling interval for the DTFT is inversely proportional to the duration of the input sequence. An Inverse Discrete Fourier Transform (IDFT) represents a Fourier series, with the DTFT samples serving as coefficients of complex sinusoids at corresponding DTFT frequencies. Consequently, the IDFT yields the same sample values as the original input sequence. The DFT thus serves as a frequency domain representation of the original input sequence. If the original sequence spans all non-zero values of a function, its DTFT is continuous (and periodic), and the DFT provides discrete samples of one cycle. Conversely, if the original sequence represents one cycle of a periodic function, the DFT offers all non-zero values of one DTFT cycle.

The discrete FFT [46] transforms a sequence of N complex numbers, $\{x_n\} = x_0, x_1, x_2, ..., x_{N-1}$, into another sequence of complex numbers, $\{X_k\} = X_0, \mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_{N-1}$, as defined by the following equation:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-i2\pi kn}{N}}, \tag{4}$$

where $k = 0, 1, 2, ..., N-1$.

2.4 Short-Time Fourier Transform

The STFT [46] is a Fourier-related technique employed to analyze the sinusoidal frequency and phase components of localized segments within a signal, capturing how they evolve over time. Practically, STFT computation involves dividing a longer duration signal into shorter segments of uniform length, upon which individual Fourier transforms are computed independently. This process unveils the Fourier spectrum associated with each short segment. Subsequently, the evolving spectra are typically plotted against time, forming a spectrogram or waterfall plot, commonly utilized in spectrum displays for Software-Defined Radio (SDR) applications. In this particular investigation, considering the discrete nature of the data, the researchers employed a discrete-time variant of the STFT to identify lung cycles.

In the discrete-time scenario, the data to be transformed is segmented into frames or chunks, often with overlapping regions to mitigate artifacts at segment boundaries. Each segment undergoes Fourier transformation, resulting in a complex output that is aggregated into a matrix, documenting the magnitude and phase for every point across time and frequency. Mathematically, this process can be represented as follows:

$$X(j,k)=\sum_{n=0}^{N-1}x(n)w(n-j)e^{\frac{-i2\pi kn}{N}} \tag{5}$$

where $X(j,k)$ is the complex-valued output for each time-frequency bin, $x(n)$ is the signal and $w(n-j)$ is the structuring element function applied to each frame to control spectral leakage, and N signifies the total number of samples in each frame.

Likewise, in the majority of typical applications, the STFT is executed on a computer leveraging FFT algorithm. Consequently, both the variables involved, discreate and continuous, are discretized and quantized. This discretization process enables efficient computation of the STFT within the digital domain, facilitating analysis of signals in discrete time and frequency intervals.

2.5 Classification

In this research, the data comprises images representing lung cycles in the frequency domain. To analyze and classify these samples, the researchers employed the GoogLeNet model, which serves as a pre-trained CNNs model.

GoogLeNet is a deep CNNs architecture consisting of 144 layers. It offers the capability to load pre-trained versions of the network that have been trained on large-scale datasets such as ImageNet or Places365. The version trained on ImageNet is designed to classify images into 1000 distinct object categories, encompassing a diverse array of objects including various animals, household items, and natural elements. Conversely, the version trained on Places365 is similar in structure but specializes in categorizing images into 365 different place categories, spanning environments such as parks, streets, and interiors.

Pretrained networks have acquired sophisticated feature representations through extensive training on vast collections of images. Notably, these networks accept images with an input size of 224 by 224 pixels. This standardized input size ensures compatibility with the pretrained models, facilitating seamless integration into the research workflow for image classification tasks.

GoogLeNet, also known as Inception V1, is a CNNs architecture developed by Google. It was the winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014 [49]. GoogLeNet introduced several innovations to CNNs architecture, including the Inception module, which allows for efficient use of computational resources and deeper networks.

In Figure 2.2 and layer details below are the brief overview of the main components and layers of the GoogLeNet architecture:

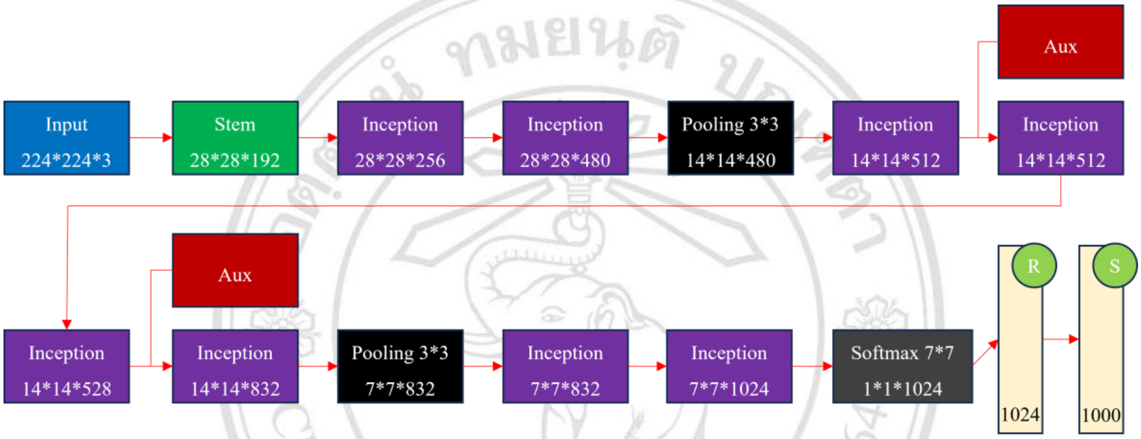

Figure 2.2 The brief overview of GoogLeNet CNNs architecture

1. Input Layer: This layer takes the input image, typically in RGB format, with a predefined size (224x224x3 pixels).

2. Convolutional Layers: The network starts with several convolutional layers that perform feature extraction. These layers use filters (kernels) to convolve across the input image, extracting low-level features such as edges, corners, and textures.

3. Inception Modules: The core innovation of GoogLeNet is the Inception module, which replaces the traditional single convolutional layer with a combination of parallel convolutional layers of different sizes (1x1, 3x3, 5x5), along with pooling operations. This allows the network to capture features at multiple scales efficiently.

4. Pooling Layers: Pooling layers, such as max pooling or average pooling, are used to down sample the feature maps obtained from the convolutional layers. They reduce the spatial dimensions of the feature maps while retaining important features.
5. Fully Connected Layers: After several convolutional and pooling layers, the feature maps are flattened into a vector and passed through one or more fully connected layers. These layers perform high-level feature extraction and classification. They may also incorporate dropout regularization to prevent overfitting.
6. Softmax Layer: The final layer of the network is a softmax layer, which produces the probability distribution over the output classes. It assigns a probability score to each class, indicating the likelihood that the input image belongs to that class.
7. Auxiliary classifiers (Aux) are inserted into intermediate layers of the network and provide additional supervision signals to the network during training. They consist of convolutional and pooling layers followed by fully connected layers and a softmax output layer. The output of these auxiliary classifiers is used as an auxiliary loss function during training, in addition to the main loss function at the end of the network.

Overall, GoogLeNet consists of many layers, including convolutional, pooling, and fully connected layers, organized into multiple Inception modules. This architecture allows for deeper networks while maintaining computational efficiency and achieving high accuracy in image classification tasks.

In the evaluation of classifier performance, a crucial tool is the confusion matrix, denoted as $A=[A(j,\mathrm{i})]$, as depicted in Table 1. Within this matrix, each element $A(j,\mathrm{i})$ represents the count of data points that belonged to the true class label i and were classified as belonging to class j.

Table 2.1 The elements of a confusion matrix A

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

From the confusion matrix A, various performance metrics can be directly derived, with accuracy being one of the fundamental measures. The properties of a classification system can be derived from the confusion matrix, enabling the calculation of important evaluation metrics. These metrics include accuracy, precision, sensitivity, specificity, F1-score, and correlation, each of which provides valuable insights into the performance of the classifier [50-54] using the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{2TP}{TN + FP} \tag{9}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{10}$$

$$Correlation = \frac{Accuracy - P_e}{1 - P_e}, \tag{11}$$

where P_e is defined as $P_e = \dfrac{(TP + FP)(TP + FN) + (TN + FN)(TN + FP)}{(TP + TN + FP + FN)^2}$

CHAPTER 3

Methods

The experimental framework utilized in this study comprises five key phases including data acquisition, data preparation, classification, testing, and performance evaluation, as depicted in Figure 3.1. Each phase involves specific steps, which are outlined as follows:

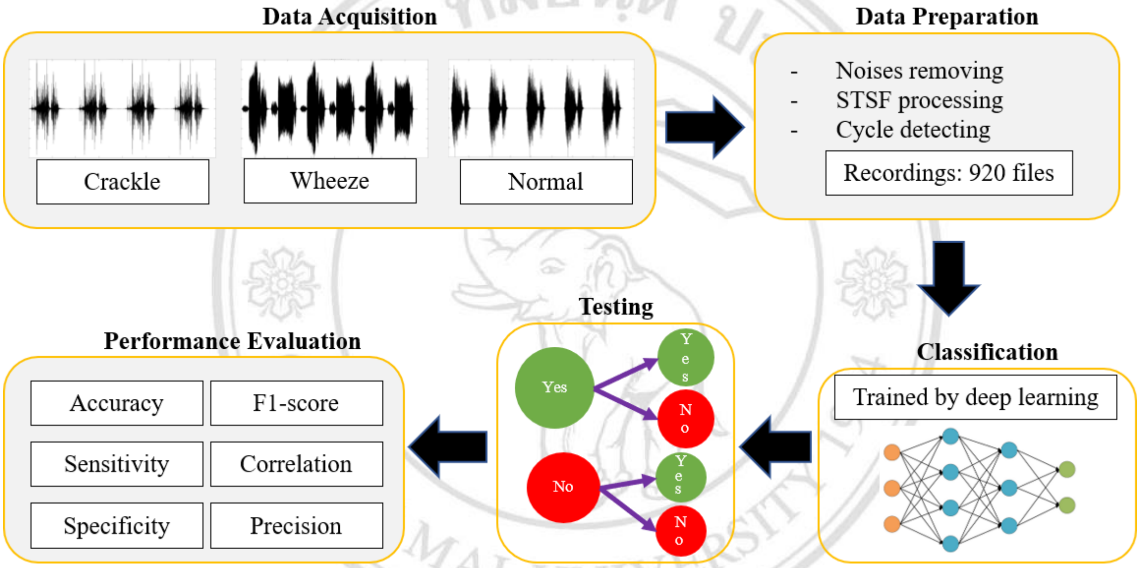

Figure 3.1 Framework block diagram of this study

3.1 Data Acquisition

We employed the "A Respiratory Sound Database for the Development of Automated Classification" dataset [2], which encompasses 920 lung sound files collected from 126 subjects. The files exhibit diverse durations, spanning from 10 seconds to 90 seconds, resulting in a total recording duration of 5.5 hours. This dataset encompasses 6,898 respiratory cycles, with 1,864 cycles containing crackle sounds, 886 cycles containing wheeze sounds, and 506 cycles containing both crackle and wheeze sounds, as illustrated in Table 3.1. The dataset incorporates recordings from individuals across various age demographics, encompassing children, adults, and the elderly. It comprises both clean respiratory sounds and noisy recordings, mimicking real-life conditions to provide a comprehensive representation of lung sound variations.

Table 3.1 Number of lung sound cycles

Lung sounds	Among of cycle
Crackle	1,864
Wheeze	886
Both crackle and wheeze	506
Normal	3,642
Total	6,898

The data collection utilized four recording devices: the AKG C417L Microphone, 3M Littmann Classic II SE, 3M Litmmann 3200 Electronic Stethoscope, and WelchAllyn Meditron Master Elite Electronic Stethoscope. These devices are capable of capturing sounds within the frequency range of 20 to 44,100 Hz.

3.2 Data Processing

3.2.1 Noise Removing

In certain instances, heart sounds posed a significant interference, overshadowing the lung sounds and obscuring the breathing cycle, particularly given their higher frequency of occurrence. To mitigate this issue, a 5^{th} order Butterworth bandpass filter was implemented with lower and upper cut-off frequencies set at 200 Hz and 2,000 Hz, respectively. This filter effectively attenuated the low-frequency components associated with heart sounds, as illustrated in Figure 3.2. Notably, lung sounds typically exhibit frequencies below 1600 Hz, while frequencies exceeding this range are often indicative of environmental noise. Furthermore, any speech artifacts within the breathing cycle, characterized by a mixture of low and high frequencies, were also removed to enhance the clarity of lung sound signals [55]. In addition, logarithmic compression was employed in this study not only to compress low-amplitude signals, which typically contain less relevant information, but also to amplify the amplitude of informative signals. As a result, the lung sound cycle in the time domain became more prominently visible, facilitating clearer analysis and interpretation of the data.

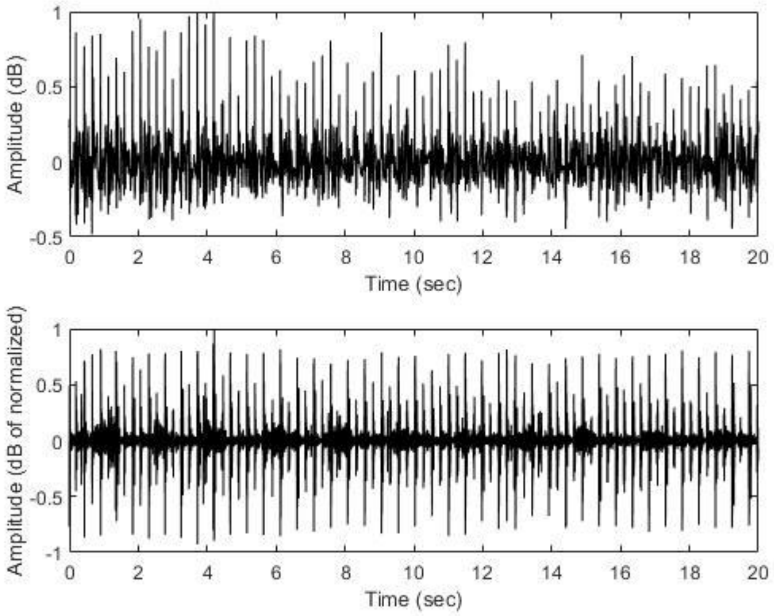

Figure 3.2 Top: Signal before using bandpass filter
Bottom: Signal after using bandpass filter

3.2.2 Short-Time Fourier Transform Processing

Following the noise removal process, the signal was subjected to analysis in the time domain by computing its absolute value. Subsequently, thresholding was applied to isolate the breathing cycle. However, the resultant signal exhibited spikes, rendering the identification of the cycle challenging. To resolve this issue, we employed STFT according to Equation (5). STFT offers a frequency depiction of the signal across various time intervals [56]. This method enabled us to examine the prickly signal as frequencies within the cycle. The spectrogram of the signal representation of the cycle is typically more distinct than its time domain counterpart, as illustrated in Figures 3.3 and 3.4. Both figures depict time on the X-axis in seconds, with the Y-axis of Figure 3.3 representing amplitude, while the Y-axis of Figure 3.4 represents frequency in Hertz.

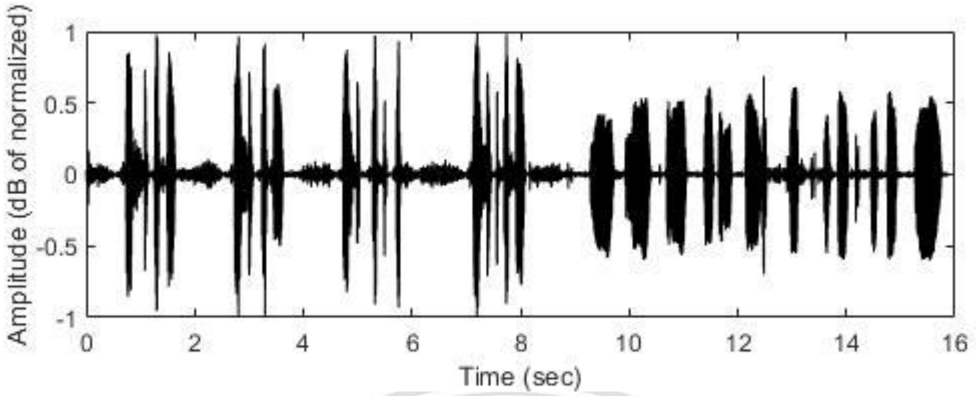

Figure 3.3 Time domain signal

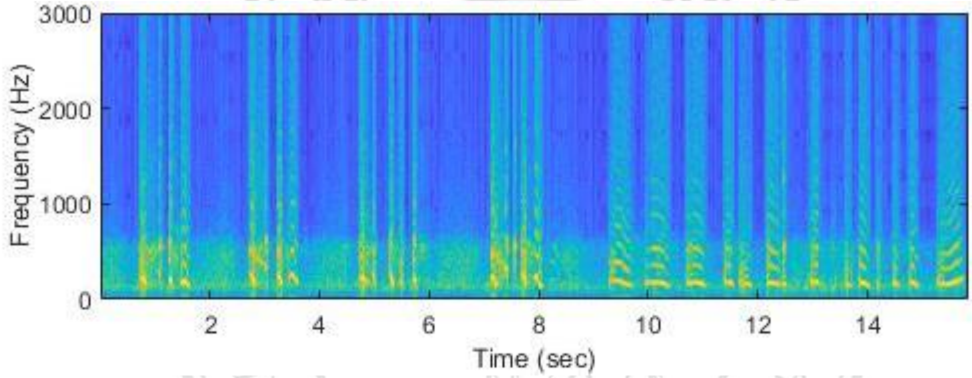

Figure 3.4 Frequency domain or spectrogram signal

3.2.3 Cycle Detecting

To identify the lung cycles from the 2-D STFT representation, each column of STFT signal was summed to get the 1-D graph. Then a 3rd order Gaussian filter with 21 pixels of window was applied to smooth the graph. Nevertheless, despite smoothing the graph, some jagged portions persisted. To address this issue, we applied a morphological opening operator to further refine the graph. This operator helped to eliminate small irregularities and smooth out the overall shape of the graph, resulting in a more accurate representation of the underlying data. The maximum and minimum point detection algorithms were applied in the graph to identify true peaks (inhalation or exhalation points) and true minimum points (covering the breathing cycle) on the graph in Figure 3.5. Following this, the processed graph was combined with the STFT representation to pinpoint the initiation and termination points of the breathing cycles, as depicted in Figure 3.6. This integration facilitated the accurate identification of the beginning and end of each respiratory cycle, thereby enabling the segmentation of the lung sound signals into individual breath cycles for further analysis. Finally, the first and the last cycle were eliminated to prevent a non-cycle. The steps to get the breathing cycles will show an example below.

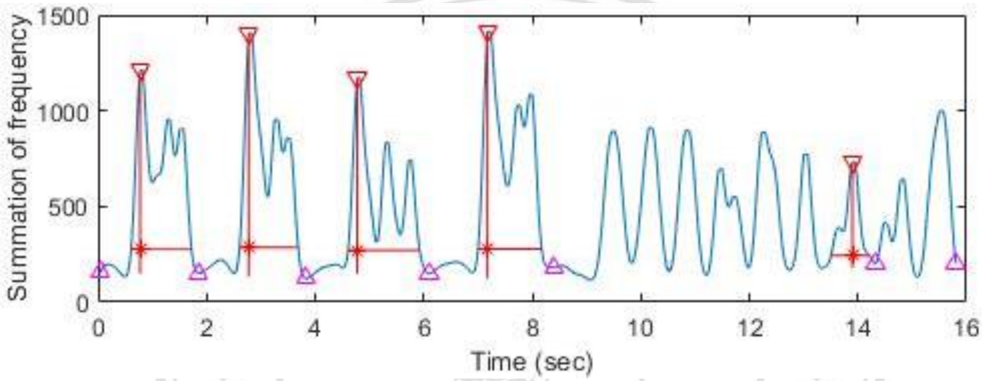

Figure 3.5 The graph represents the smoothed and detected cycle points

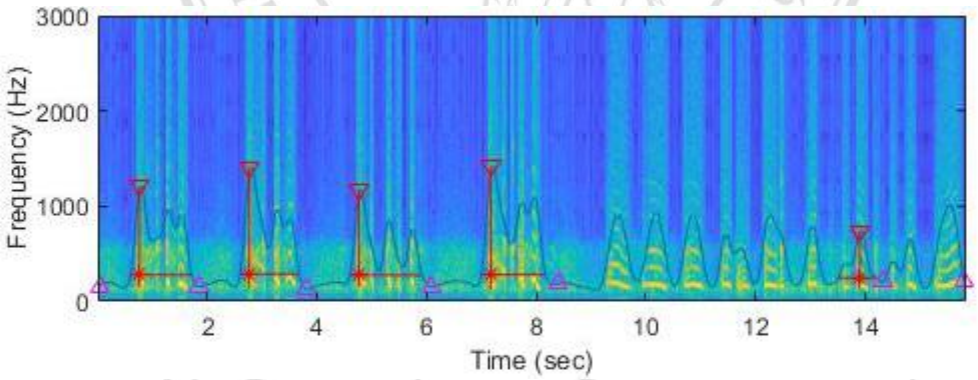

Figure 3.6 The merged graph depicts the smoothed and detected cycles in the frequency domain, alongside the frequency domain signal

For example, case number 104 (1) employed a 5^{th} order Butterworth bandpass filter within the frequency range of 200 to 2,000 Hz and logarithmic compression normalization also be applied, yielding outcomes depicted in Figure 3.7. Subsequently, it will undergo STFT with the outcomes depicted in Figure 3.8.

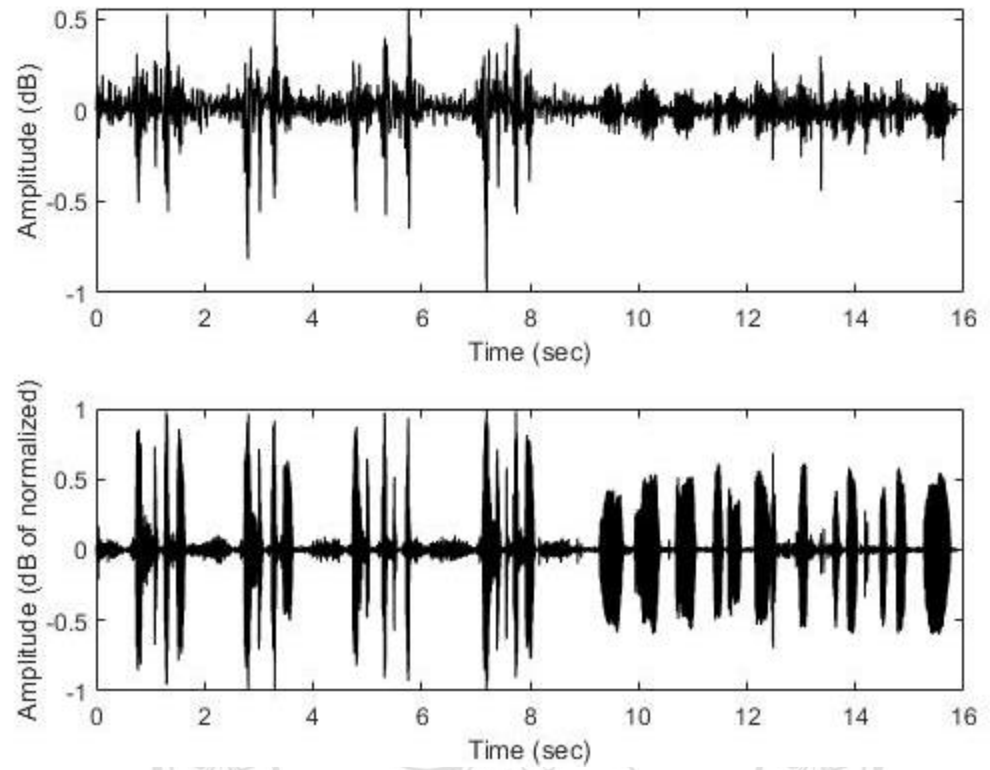

Figure 3.7 Top: Original signal

Bottom: Signal after applying bandpass filter and logarithmic compression normalization

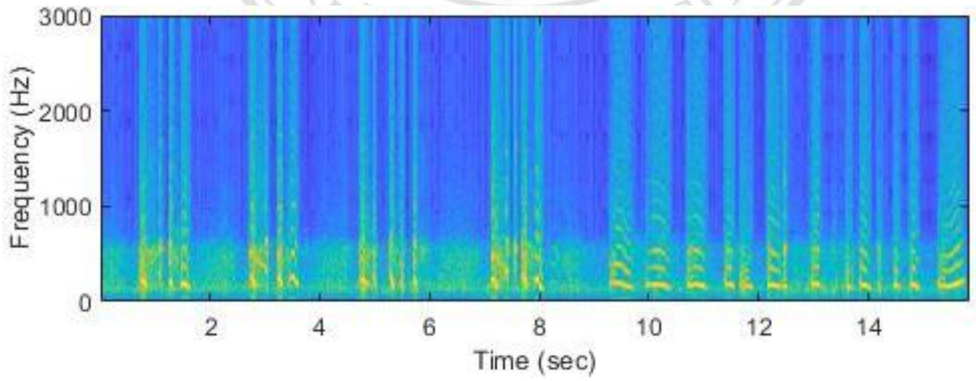

Figure 3.8 Signal after applying STFT

Afterward, the magnitude of spectrogram in each column was summed to obtain the magnitude of spectrogram graph shown in Figure 3.9 and apply smoothing to the graph using a 3rd order Gaussian filter with 21 pixels of window. Subsequently, we performed an opening operation to fill in any small holes and enhance the smoothness of the orange graph. The refined results are illustrated in Figure 3.10.

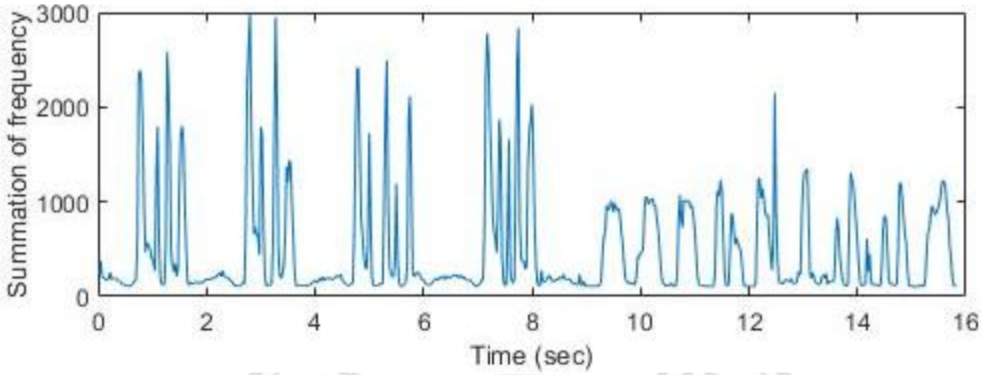

Figure 3.9 Graph summation of each column of its STFT

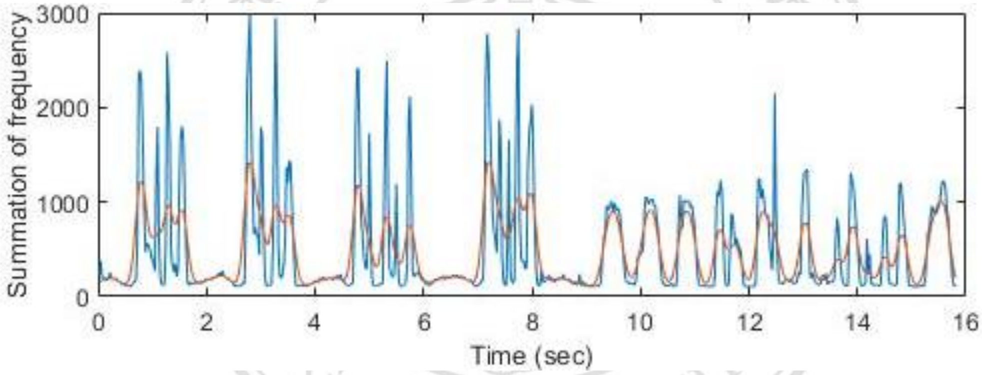

Figure 3.10 Smoothed graph (orange line)

Next, the smoothed graph was identified the maximum and minimum points. The highest points correspond to each cycle of inhalation and exhalation, with some instances featuring both inhalation and exhalation, while others have only one phase. This pattern is depicted in Figure 3.11. Subsequently, we established a prominence line (vertical line) as depicted in Figure 3.12 and determined peak widths at 50% of the prominence. Following this, the points with excessively long or oscillating peak widths are removed and generated a new width line at 88%, as illustrated in Figure 3.13.

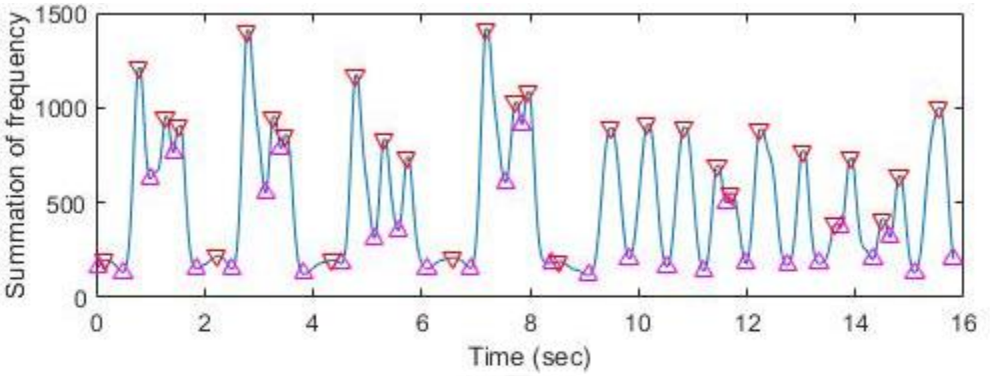

Figure 3.11 Smoothed graph with all peaks and all minimum points

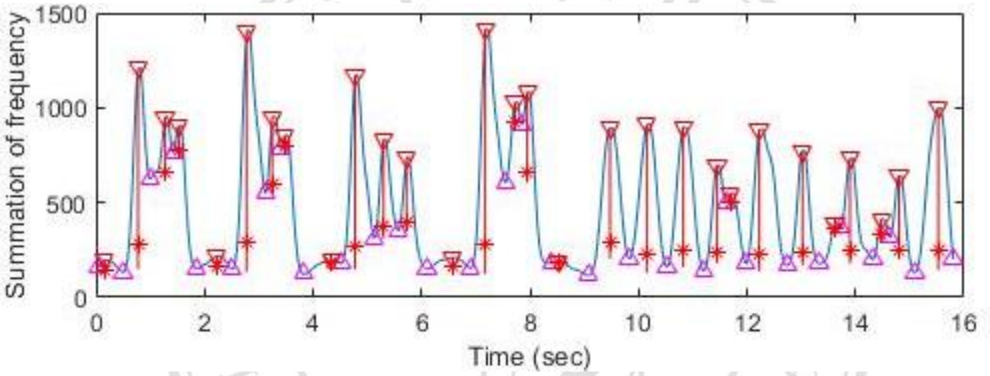

Figure 3.12 Smoothed graph with all peaks, all minimum points, and all prominence lines

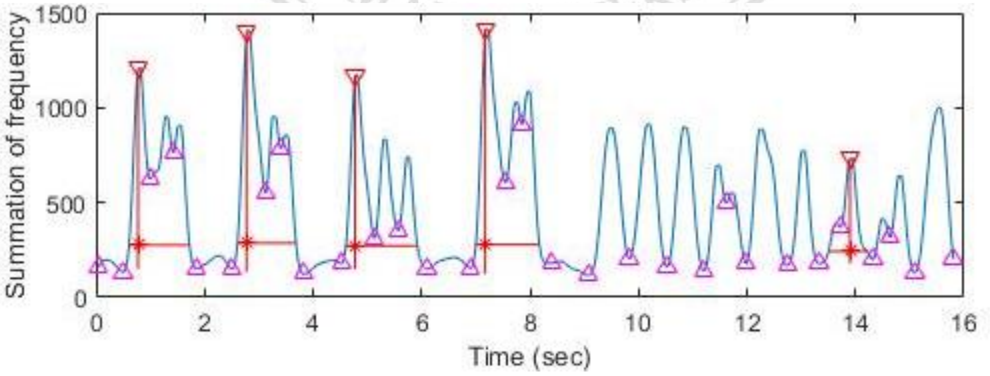

Figure 3.13 Smoothed graph with truth peak of each cycle, all minimum points, and truth peak prominence lines

Subsequently, all peak points within the range of the peak widths were eliminated to isolate the most maximum point in each segment of the graph. Following this, the minimum points were retained that were closest to the prominence before and after, resulting in Figure 3.14.

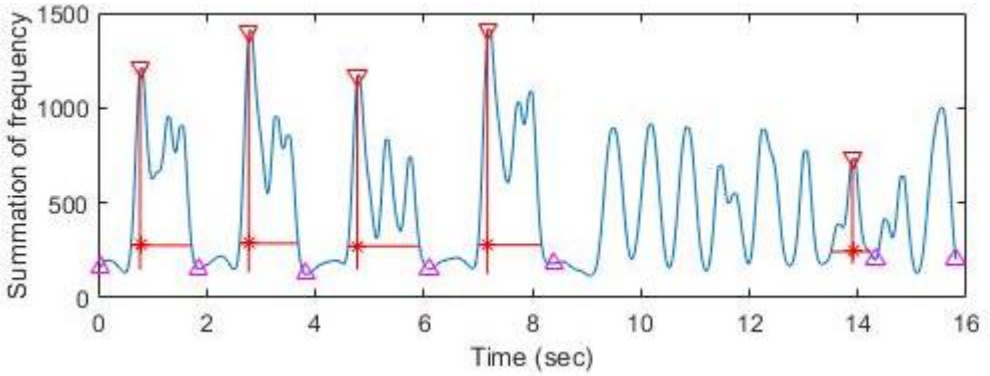

Figure 3.14 Smoothed graph with truth peak, truth peak prominence line and minimum points as the edges of breathing cycle of each cycle

Then, the graph was proceeded to compare the length of time from each nearest minimum point. If the time interval was similar to the duration of one breath taken by the patient, we retained it and proceeded to examine the next point. However, if the time interval was less than 25 percent, we incremented the count by one more point, totaling three points. This adjustment was made because the second point might only represent the transition between inhaling and exhaling. Consequently, the cycle of inhalation and exhalation could be identified from the first lowest point to the next, constituting the first round of breathing in and out. The subsequent points, from the second to the third, represented subsequent rounds. This process continued until all points were examined. Finally, we superimposed these findings onto the spectrogram of the signal itself, as depicted in Figure 3.15.

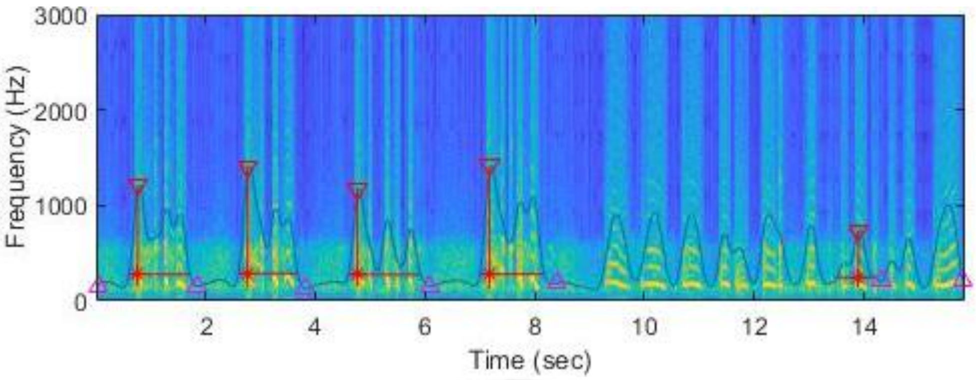

Figure 3.15 Merged graph of cycle detection and its STFT

Lastly, the breathing cycles were segmented at the lowest point to isolate each individual breathing cycle and eliminated the first and the last cycles for preventing the non-complete cycles of lung sounds, resulting in the depiction shown in Figure 3.16.

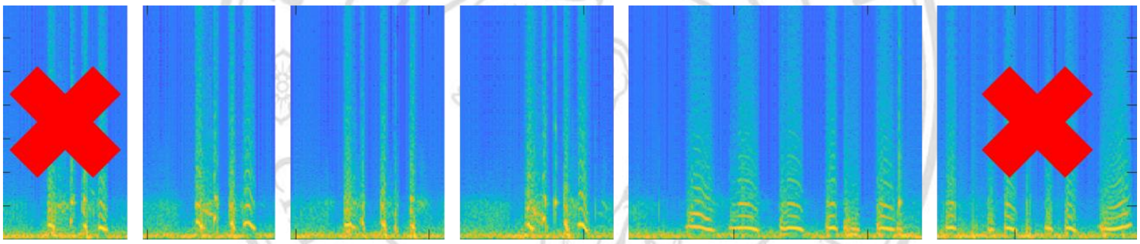

Figure 3.16 Each breathing cycle

3.3 Classification

After getting all samples of 5,076 cycles that were crackle of 1,908 cycles, wheeze of 780 cycles, and normal of 2,388 cycles from this cycle detection technique. For each class, 80% of the cycle sound types were randomly used for training (1,590 of crackle, 650 of wheeze, and 1,990 of normal) as show in Table 3.2. A pretrained GoogLeNet model (one of pretrained CNNs) was employed in this study. The GoogLeNet model received STFT images as inputs, which covered the breathing cycle detected in the previous step. The GoogLeNet model was configured to produce outputs corresponding to four distinct classes: 1) crackle versus wheeze, 2) crackle versus normal, 3) wheeze versus normal, and 4) a combination of all three (crackle versus wheeze versus normal). The number of epochs and the learning rate were set to 60 and 0.001, respectively. The remaining 20% of the data were reserved for testing the algorithm model specific to each class.

Table 3.2 Number of samples for training and testing

Lung sounds	80 percent	20 percent
Crackle	1,590	318
Wheeze	650	130
Normal	1,990	398
Total	4,230	846

3.4 Performance Evaluation

After the samples underwent classification by the GoogLeNet model, the subsequent step involved the calculation of performance evaluation metrics, which included True Positives, True Negatives, False Positives, and False Negatives values. These metrics were then utilized to compute several key performance indicators, including accuracy, precision, sensitivity, specificity, F1-Score, and correlation. These calculations were performed using Equations (6) through (11), respectively, to perform the necessary calculations, facilitating the assessment of the model's capability to differentiate between distinct categories of lung sounds.

3.4.1 Accuracy

Accuracy is a performance metric that measures the proportion of correctly classified instances out of the total number of instances evaluated by a predictive model. It provides an overall assessment of the model's correctness and is calculated by dividing the number of correctly predicted instances by the total number of instances. Accuracy is particularly useful when the classes in the dataset are balanced, but it may not provide an accurate representation of performance in the presence of class imbalance.

3.4.2 Precision

Precision is a performance metric that evaluates the accuracy of positive predictions made by a model. It measures the proportion of true positive predictions (correctly predicted positive instances) out of all positive predictions made by the model, including both true positives and false positives. Precision focuses on the quality of positive predictions and is calculated by dividing the number of true positives by the sum of true positives and false positives.

3.4.3 Sensitivity

Sensitivity, also known as recall or true positive rate, assesses the ability of a model to correctly identify positive instances from the total number of actual positive instances in the dataset. It measures the proportion of true positive predictions out of all actual positive instances and is calculated by dividing the number of true positives by the sum of true positives and false negatives. Sensitivity is crucial in scenarios where correctly identifying positive instances is of utmost importance, such as medical diagnostics.

3.4.4 F1-Score

The F1-Score is a performance metric that provides a balanced measure of a model's precision and sensitivity. It is the harmonic mean of precision and sensitivity, calculated by taking the reciprocal of the average of the reciprocals of precision and sensitivity. The F1-Score considers both false positives and false negatives and is especially useful when there is an imbalance between positive and negative instances in the dataset.

3.4.5 Correlation

Correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. It ranges from -1 to 1, where 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. Correlation is commonly used to assess the association between variables and is particularly useful in identifying patterns and dependencies in data analysis and predictive modeling.

CHAPTER 4

Experimental Results and Discussions

4.1 Experimental Results

To offer a thorough assessment of the results obtained from the test data, the confusion matrices for all four classes, namely "crackle versus wheeze versus normal," "crackle versus wheeze," "crackle versus normal," and "wheeze versus normal," are provided in Tables 4.1 to 4.4, respectively. Additionally, Table 4.5 presents the performance evaluation metrics for each case, encompassing accuracy, precision, sensitivity, F1-score, and correlation.

Table 4.1 Confusion matrix of crackle versus wheeze versus normal

		Actual		
		Crackle	Wheeze	Normal
Predicted	Crackle	217	11	157
	Wheeze	19	82	57
	Normal	82	37	184

Table 4.2 Confusion matrix of crackle versus wheeze

		Actual	
		Crackle	Wheeze
Predicted	Crackle	283	31
	Wheeze	35	99

Table 4.3 Confusion matrix of crackle versus normal

		Actual	
		Crackle	Normal
Predicted	Crackle	210	157
	Normal	108	241

Table 4.4 Confusion matrix of wheeze versus normal

		Actual: Wheeze	Actual: Normal
Predicted	Wheeze	89	90
	Normal	41	308

Table 4.5 Evaluation of each category's performance includes comparisons between crackle versus wheeze versus normal, crackle versus wheeze, crackle versus normal, and wheeze versus normal lung sounds classifications

Classes	Crackle versus wheeze versus normal	Crackle versus wheeze	Crackle versus normal	Wheeze versus normal
Accuracy	57.09%	85.27%	62.99%	75.19%
Precision	58.28%	90.13%	57.22%	49.72%
Sensitivity	66.74%	88.99%	66.04%	68.46%
Specificity	46.23%	76.15%	60.55%	77.39%
F1-Score	62.23%	89.56%	61.31%	57.61%
Correlation	0.13	0.65	0.26	0.41

Table 4.5 reveals that the crackle versus wheeze classification achieved the highest metrics across all tested scenarios, with accuracy at 85.27%, precision at 90.13%, sensitivity at 88.99%, F1-score at 89.56%, and correlation at 0.65. Conversely, the wheeze versus normal classification displayed the highest specificity at 77.39%. Notably, distinguishing crackle sounds proved more feasible compared to normal sounds, resulting in relatively lower performance in the crackle versus normal classification. This challenge likely stems from the overlapping frequency characteristics between crackles and normal lung sounds, posing difficulty in distinguishing both classes using STFT features.

Moreover, the algorithm exhibited the robust performance in distinguishing between crackles, wheezes, and normal lung sounds, even in challenging classification scenarios involving multiple sound categories.

Table 4.6 Accuracy of each case in class of crackle versus wheeze versus normal

Cases	TP	FP	FN	TN	Accuracy
Crackle	217	168	101	360	68.20%
Wheeze	82	76	48	640	85.34%
Normal	184	119	214	329	60.64%

Furthermore, Table 4.6 presents the individual accuracy of crackle sounds, wheeze sounds, and normal sounds for the classification scenario of crackle versus wheeze versus normal model belonging to Table 4.1 by 68.20%, 85.34%, and 60.64%, respectively.

Table 4.7 Confusion matrix of crackle versus wheeze versus normal of pretrained model

		Actual		
		Crackle	Wheeze	Normal
Predicted	Crackle	0	0	0
	Wheeze	0	1	0
	Normal	327	149	477

From Table 4.7 presents the confusion matrix for the class of crackle versus wheeze versus normal, depicting the performance of the original pretrained model (the model prior to learning). The matrix illustrates that the model predominantly classified the data into the normal case across all possible scenarios. This contrasts with the findings observed in Table 4.1, which showcases the confusion matrix for the same class but with the pretrained model (the model after learning). The disparity between the two matrices underscores the influence of model learning on prediction outcomes, highlighting the significance of the learning process in shaping the model's predictive capabilities.

4.2 Discussions

The findings of this study illuminate the promising potential of automated lung sound analysis in reshaping respiratory healthcare. By amalgamating sophisticated signal processing techniques with cutting-edge machine learning algorithms, the proposed methodology provides a dependable and efficient means of identifying abnormal lung sounds.

The impressive accuracy rates achieved by the algorithm underscore its clinical significance in supporting healthcare practitioners with the diagnosis and monitoring of respiratory conditions. Accurately classifying abnormal lung sounds, including crackles and wheezes, has the potential to greatly improve diagnostic precision and streamline treatment decision-making procedures.

However, the study also uncovers several challenges and limitations that merit attention. The interference of heart sounds emerges as a significant hurdle in accurately isolating lung sounds, necessitating the development of more advanced noise removal techniques. Additionally, issues concerning precise cycle detection and differentiation between similar sound patterns emphasize the need for continual methodological refinement.

Looking ahead, future research endeavors could concentrate on tackling these challenges through the exploration of innovative noise reduction strategies and the enhancement of algorithmic approaches for sound classification. Furthermore, integrating real-time monitoring capabilities into portable medical devices harbors immense potential for enhancing respiratory healthcare delivery, enabling timely interventions and improved patient outcomes.

We faced challenges in addressing noise interference, particularly from heart sounds. The use of a high-pass filter to eliminate heart sound noise inadvertently led to the removal of low-frequency lung sound information, as both heart and lung sounds share overlapping frequency ranges. Additionally, the maximum frequency of heart sounds varied across different locations of the lung lobes, complicating the noise removal process. These complexities highlight the difficulty in effectively eliminating heart sound noise while preserving pertinent lung sound information [15]. Achieving accurate differentiation between the two types of sounds based solely on frequency cutoff proved challenging due to these factors.

Another issue raised concerns the algorithm's accuracy in counting breathing cycles. Some cycles were incorrectly counted as two separate cycles instead of one, while others lacked the inclusion of inspiration and expiration phases at the beginning and end, respectively, as indicated by the ground truth data. These observations highlight limitations in the algorithm's capability to accurately detect and count cycles. Moreover, the presence of prolonged coughing sounds, surpassing the typical duration of breathing cycles, posed additional challenges in precise cycle counting.

These challenges and limitations suggest potential areas for methodological refinement. For example, investigating alternative noise reduction techniques tailored to address heart sound interference and enhancing algorithms for accurate cycle detection could improve the overall performance of the classification system. This would enhance its reliability in effectively distinguishing between various categories of lung sounds.

Moving forward, this study compares crackle vs wheeze lung sounds, focusing on predicting cases where the input consists solely of crackle or wheezing sounds. Similarly, comparisons between crackle vs normal sound and wheeze vs normal sound are conducted to predict only two possible cases learned by the system. When confronted with lung sound inputs outside its training data, the system predicts based solely on the learned cases. However, attempts to train the model with crackle vs wheeze vs normal cases yielded the lowest performance, suggesting diminishing returns with the addition of more cases. Furthermore, it was found that the STFT may not provide sufficient features for identifying these lung sounds.

Several suggestions arise from these findings. Firstly, utilizing a new and improved dataset is recommended due to various limitations in the current dataset.

1) Some sample files contain various types of noise, including coughing sounds, medical device noise, speech noise, electronic noise, etc., as illustrated in Figure 4.1.
2) Certain samples contain heart sounds rather than lung sounds, which predominantly occur in the lower left lobe of the lungs, as depicted in Figure 4.2.

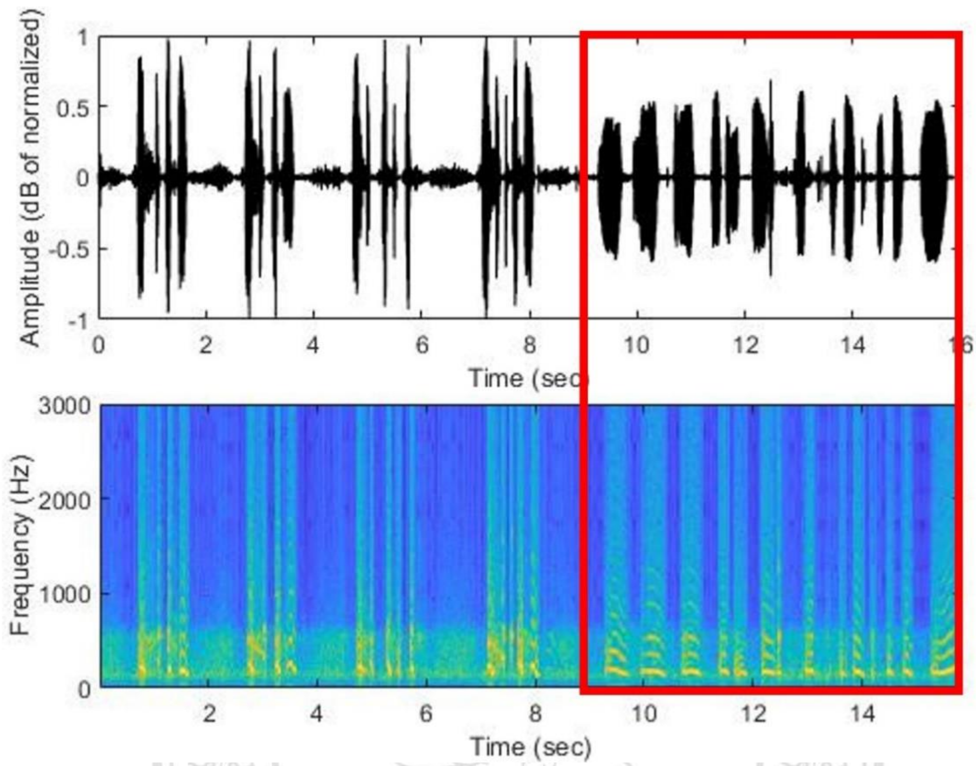

Figure 4.1 Speech noise signal

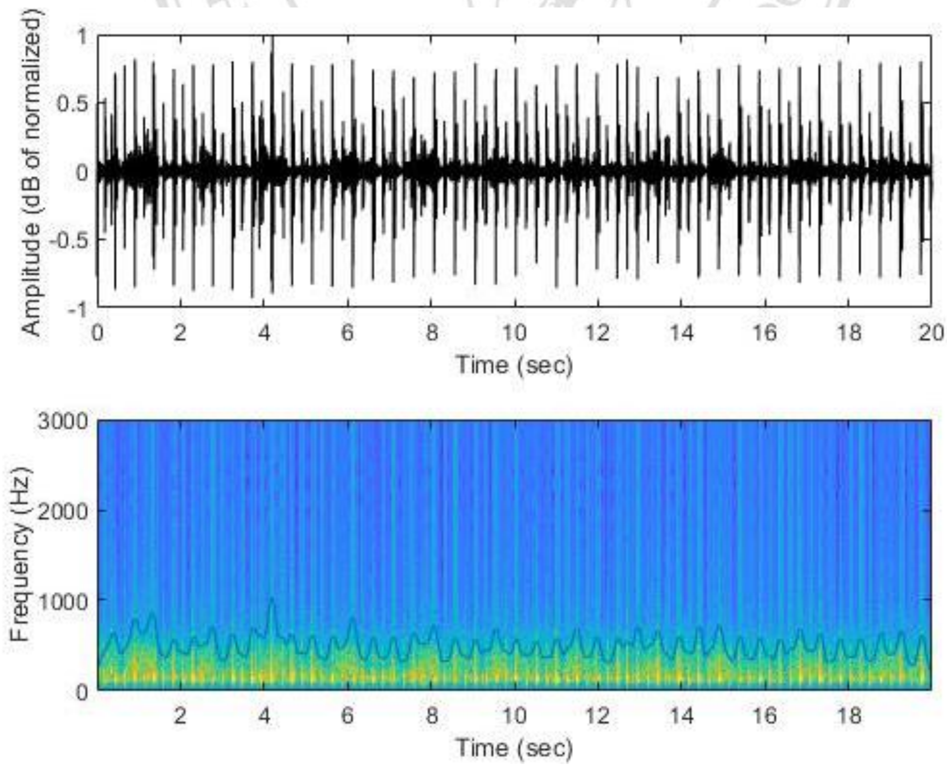

Figure 4.2 Heart sound signal

3) In some parts of certain sample files, there is no discernible sound, yet they are labeled as crackle, wheeze, both (crackle and wheeze), or normal lung sound, as shown in Figure 4.3.

4) The dataset is collected from four different types of recording devices (AKG C417L Microphone, 3M Littmann Classic II SE Stethoscope, 3M Littmann 3200 Electronic Stethoscope, and WelchAllyn Meditron Master Elite Electronic Stethoscope) and two different modes (sequential/single channel and simultaneous/multichannel), potentially leading to unequal resolution and information within the data.

5) In clinical settings, crackle lung sounds can be further divided into fine crepitation and coarse crepitation, which should also be labeled to improve the prediction system.

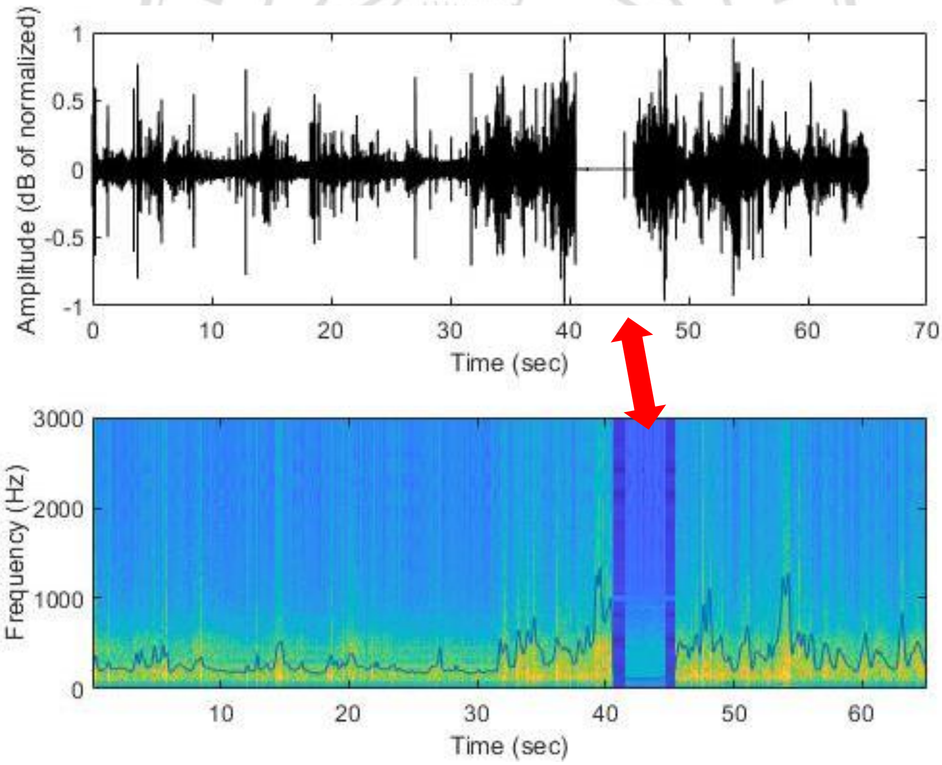

Figure 4.3 No sound information part

Finally, considering the characteristics of the data, alternative classification techniques such as Long Short-Term Memory (LSTM), SVM, RNN, PCA, among others, designed for one-dimensional signal classification, may be more appropriate than CNNs, which are typically used for two-dimensional signal or image classification. Additionally, in this thesis, during the training step, the data were trained only once for each class. It could be advantageous to employ k-fold cross-validation instead of one time trained.

In summary, while the proposed algorithm demonstrates impressive performance in automating abnormal lung sound identification, continued research and development efforts are essential to further enhance its reliability and efficacy in clinical settings. By leveraging the synergies between signal processing and machine learning, the field of respiratory medicine stands poised to benefit from transformative advancements in diagnostic technologies, ultimately improving patient care and outcomes.

CHAPTER 5

Conclusion

This study aims to fulfill the pressing need for automated detection of abnormal lung sounds by employing a novel fusion of signal processing and machine learning methodologies. Through the development of a robust algorithm capable of accurately discerning between normal and abnormal breathing sounds, specifically targeting crackles and wheezes, this research marks a significant advancement in respiratory medicine diagnostics.

The proposed methodology initiates with the acquisition of lung sound data from a meticulously curated dataset, encompassing a diverse range of recordings spanning various respiratory conditions and demographic profiles. Subsequent data processing involves meticulous noise removal, with particular emphasis on eliminating heart sounds, followed by the transformation of signals into the frequency domain utilizing STFT. This process facilitates the identification and segmentation of breath cycles, laying the groundwork for precise classification.

Harnessing the power of CNNs, specifically leveraging the GoogLeNet model, enables robust classification of abnormal lung sounds based on extracted features from STFT images. The algorithm demonstrates remarkable accuracy rates across multiple classification scenarios, notably excelling in differentiating between crackles and wheezes. These findings underscore the effectiveness of the proposed approach in facilitating precise diagnosis and monitoring of respiratory conditions.

However, the study also sheds light on several challenges and limitations that warrant further investigation and refinement. Addressing issues such as noise interference, precise cycle detection, and differentiation between similar sound patterns presents avenues for methodological enhancement. Investigating alternative methods for noise reduction and advancing algorithmic sophistication have the potential to elevate the overall performance and dependability of the classification system.

Looking ahead, the research holds promising implications for the development of a portable medical device capable of real-time lung sound analysis. Such a device has the potential to revolutionize respiratory healthcare by providing immediate access to diagnostic insights, thereby facilitating timely interventions and improving patient outcomes. Moreover, the integration of advanced technologies into medical practice underscores the transformative role of interdisciplinary research in advancing human health and well-being.

In conclusion, this study represents a significant stride towards the automation of abnormal lung sound identification, offering valuable contributions to the field of respiratory medicine and paving the way for future innovations in healthcare diagnostics and monitoring.

REFERENCES

[1] vbookshelf. (2021, Jun. 1). *Respiratory Sound Database Use audio recordings to detect respiratory diseases.* [internet]. Available: https://www.kaggle.com/vbookshelf/respiratory-sound-database?select=demographic_info.txt&fbclid=IwAR2kqg_VLZjnRVUhz9BsxPfus87P5wfScYV2MSJqdZp59t-lF3b85L6H0Bs

[2] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, et al., "A Respiratory Sound Database for the Development of Automated Classification," in *Precision Medicine Powered by pHealth and Connected Health*, Singapore, 2018, pp. 33-37.

[3] S. Ntalampiras, and I. Potamitis, "Automatic acoustic identification of respiratory diseases," *Evolving Systems* April 2020.

[4] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique," *Computers in Biology and Medicine,* vol. 104, pp. 175-182, January 2019.

[5] L. Pham, I. McLoughlin, H. Phan, M. Tran, T. Nguyen, and R. Palaniappan, "Robust Deep Learning Framework for Predicting Respiratory Anomalies and Diseases," in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 2020, pp. 164-167.

[6] Q. Zhang, C. O. Francisco, M. Kabir, J. Zhang, N. Montazeri, B. Taati, et al., "Noise Removal of Tracheal Sound Recorded during CPET to Determine Respiratory Rate," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 4650-4653.

[7] R. J. Oweis, E. W. Abdulhay, A. Khayal, and A. Awad, "An alternative respiratory sounds classification system utilizing artificial neural networks," *Biomedical Journal,* vol. 38, no. 2, pp. 153-161, March 2015.

[8] MEDZCOOL. (2021, Mar. 18). *Heart and Lung Sound Library* [web page]. Available: https://www.medzcool.com/auscultate

[9] K. X. Zhang, Z. Long, X. F. Wang, and H. Zhao, "Detection of Wheeze Based on Hough Transform of Spectrogram," *Dongbei Daxue Xuebao/Journal of Northeastern University,* vol. 38, no. 11, pp. 1534-1537, 2017.

[10] J. Zhang, H. S. Wang, H. Y. Zhou, B. Dong, L. Zhang, F. Zhang, et al., "Real-World Verification of Artificial Intelligence Algorithm-Assisted Auscultation of Breath Sounds in Children," *Frontiers in Pediatrics,* vol. 9 March 2021.

[11] H. Chen, X. Yuan, J. Li, Z. Pei, and X. Zheng, "Automatic Multi-Level In-Exhale Segmentation and Enhanced Generalized S-Transform for wheezing detection," *Comput Methods Programs Biomed,* vol. 178, pp. 163-173, September 2019.

[12] A. R. A. Sovijarvi, P. Helisto, L. P. Malmberg, K. Kallio, E. Paajanen, A. Saarinen, et al., "A new versatile PC-based lung sound analyzer with automatic crackle analysis (HeLSA); repeatability of spectral parameters and sound amplitude in healthy subjects," *Technology and Health Care,* vol. 6, no. 1, pp. 11-22, January 1998.

[13] S. K. Chowdhury, and A. K. Majumder, "Digital Spectrum Analysis of Respiratory Sound," *IEEE Transactions on Biomedical Engineering,* vol. BME-28, no. 11, pp. 784-788, November 1981.

[14] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers," *Biocybernetics and Biomedical Engineering,* vol. 41, no. 1, pp. 1-14, January 2021.

[15] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Transactions on Biomedical Engineering,* vol. 58, no. 12 PART 1, pp. 3360-3367, December 2011.

[16] V. K. Iyer, P. A. Ramamoorthy, H. Fan, and Y. Ploysongsang, "Reduction of Heart Sounds from Lung Sounds by Adaptive Filtering," *IEEE Transactions on Biomedical Engineering,* vol. BME-33, no. 12, pp. 1141-1148, December 1986.

[17] A. Suzuki, C. Sumi, and K. Nakayama, "Adaptive Cancelling of Ambient Noise in Lung Sound Measurement," *Japanese journal of medical electronics and biological engineering,* vol. 31, no. 4, pp. 354-359, 1993.

[18] G. Kandilogiannakis, P. Mastorocostas, and D. Varsamis, "A computational intelligence-based filter for lung sound separation," in *22nd Pan-Hellenic Conference on Informatics*, New York, NY, United States, 2018, pp. 107-112.

[19] X. H. Kok, S. Anas Imtiaz, and E. Rodriguez-Villegas, "A Novel Method for Automatic Identification of Respiratory Disease from Acoustic Recordings," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 2589-2592.

[20] S. Li, and Y. Liu, "Feature Extraction of Lung Sounds Based on Bispectrum Analysis," in *Third International Symposium on Information Processing*, Qingdao, China, 2010, pp. 393-397.

[21] S. B. Manir, M. Karim, and M. A. Kiber, "Assessment of Lung Diseases from Features Extraction of Breath Sounds Using Digital Signal Processing Methods," in *Emerging Technology in Computing, Communication and Electronics (ETCCE)*, Bangladesh, 2020, pp. 1-6.

[22] M. Bahoura, and C. Pelletier, "Respiratory sounds classification using cepstral analysis and Gaussian mixture models," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, CA, USA, 2004, pp. 9-12.

[23] F. S. Hsu, C. J. Huang, C. Y. Kuo, S. R. Huang, Y. R. Cheng, J. H. Wang, et al., "Development of a respiratory sound labeling software for training a deep learning-based respiratory sound analysis model," in *International Forum on Medical Imaging in Asia*, Taipei, Taiwan, 2021.

[24] Z. Neili, M. Fezari, and A. Redjati, "ELM and K-nn machine learning in classification of Breath sounds signals," *International Journal of Electrical and Computer Engineering,* vol. 10, no. 4, pp. 3528-3536, August 2020.

[25] A. Kumar, D. R. P. M. Vincent, K. Srinivasan, and C.-Y. Chang, "Deep Convolutional Neural Network based Feature Extraction with optimized Machine Learning Classifier in Infant Cry Classification," in *International Conference on Decision Aid Sciences and Application (DASA)*, Sakheer, Bahrain, 2020, pp. 27-32.

[26] R. Vashkevich, and E. Azarov, "Pitch-invariant Speech Features Extraction for Voice Activity Detection," in *22nd International Conference on Digital Signal Processing and its Applications (DSPA)*, Moscow, Russia, 2020, pp. 1-4.

[27] B. A. Reyes, S. Charleston-Villalobos, R. González-Camarena, and T. Aljama-Corrales, "Assessment of time-frequency representation techniques for thoracic sounds analysis," *Comput Methods Programs Biomed,* vol. 114, no. 3, pp. 276-90, May 2014.

[28] R. Palaniappan, K. Sundaraj, S. Sundaraj, N. Huliraj, and S. S. Revadi, "Classification of pulmonary pathology from breath sounds using the wavelet packet transform and an extreme learning machine," *Biomed Tech (Berl),* vol. 63, no. 4, pp. 383-394, July 2018.

[29] J. Yan, X. Shen, Y. Wang, F. Li, C. Xia, R. Guo, et al., "Objective research of auscultation signals in Traditional Chinese Medicine based on wavelet packet energy and support vector machine," *Int J Bioinform Res Appl,* vol. 6, no. 5, pp. 435-48, January 2010.

[30] R. S. Singh, B. S. Saini, and R. K. Sunkaria, "Detection of coronary artery disease by reduced features and extreme learning machine," *Clujul Med,* vol. 91, no. 2, pp. 166-175, May 2018.

[31] M. Ono, K. Arakawa, M. Mori, T. Sugimoto, and H. Harashima, “Separation of Fine Crackles from Vesicular Sounds by a Nonlinear Digital Filter,” *IEEE Transactions on Biomedical Engineering,* vol. 36, no. 2, pp. 286-291, February 1989.

[32] E. Ademovic, J. C. Pesquet, and G. Charbonneau, “Wheezing lung sounds analysis with adaptive local trigonometric transform,” *Technology and Health Care,* vol. 6, no. 1, pp. 41-51, February 1998.

[33] H. Kiyokawa, M. Greenberg, K. Shirota, and H. Pasterkamp, “Auditory detection of simulated crackles in breath sounds,” *Chest,* vol. 119, no. 6, pp. 1886-1892, June 2001.

[34] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, “Neural classification of lung sounds using wavelet coefficients,” *Computers in Biology and Medicine,* vol. 34, no. 6, pp. 523-537, September 2004.

[35] K. K. Guntupalli, P. M. Alapat, V. D. Bandi, and I. Kushnir, “Validation of automatic wheeze detection in patients with obstructed airways and in healthy subjects,” *Journal of Asthma,* vol. 45, no. 10, pp. 903-907, 2008.

[36] Z. Wang, and Y. X. Xiong, “Computerized lung sound analysis following clinical improvement of pulmonary edema due to congestive heart failure exacerbations,” *Chinese Medical Journal,* vol. 123, no. 9, pp. 1127-1132, May 2010.

[37] 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,” *Respiratory Medicine,* vol. 105, no. 9, pp. 1396-1403, September 2011.

[38] L. E. Ellington, D. Emmanouilidou, M. Elhilali, R. H. Gilman, J. M. Tielsch, M. A. Chavez, et al., “Developing a Reference of Normal Lung Sounds in Healthy Peruvian Children,” *Lung,* vol. 192, no. 5, pp. 765-773, June 2014.

[39] K. Kosasih, U. R. Abeyratne, and V. Swarnkar, “High frequency analysis of cough sounds in pediatric patients with respiratory diseases,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, 2012, pp. 5654-5657.

[40] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, “Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis,” *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 4, pp. 1185-1194, April 2015.

[41] N. S. Haider, J. Joseph, and R. Periyasamy, “An investigation on the statistical significance of spectral signatures of lung sounds,” *Biomedical Research (India),* vol. 28, no. 6, pp. 2801-2810, 2017.

[42] C. Habukawa, N. Ohgami, N. Matsumoto, K. Hashino, K. Asai, T. Sato, et al., “A wheeze recognition algorithm for practical implementation in children,” *PLoS ONE,* vol. 15, no. 10 October 2020.

[43] S. Z. H. Naqvi, and M. A. Choudhry, “An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis,” *Sensors (Switzerland),* vol. 20, no. 22, pp. 1-23, November 2020.

[44] B. A. Shenoi, *Introduction to Digital Signal Processing and Filter Design*: John Wiley and Sons, 2005.

[45] S. W. Smith, *The scientist and engineer's guide to digital signal processing*, 1st ed. ed.San Diego, Calif.: California Technical Pub., 1997.

[46] N. Theera-Umpon, in *Digital Signal and Image Processing Theories and Applications*, Chiang Mai, Thailand: ST2 Mo Space design, 2018.

[47] S. Butterworth, “On the Theory of Filter Amplifiers,” *Experimental Wireless and the Wireless Engineer,* vol. 7, pp. 536-541, 1930.

[48] M. Müller, "Fundamentals of Music Processing," Springer International Publishing, 2021.

[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[50] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology,* vol. 215, no. 3, pp. 403-410, October 1990.

[51] T. Joachims, "A Support Vector Method for multivariate performance measures," in *22nd international conference on Machine learning*, New York, NY, United States, 2005, pp. 377-384.

[52] J. Kaufman, B. Birmaher, D. Brent, U. Rao, C. Flynn, P. Moreci, et al., "Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data," *Journal of the American Academy of Child and Adolescent Psychiatry,* vol. 36, no. 7, pp. 980-988, July 1997.

[53] G. Kresse, and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B - Condensed Matter and Materials Physics,* vol. 54, no. 16, pp. 11169-11186, October 1996.

[54] K. J. Livak, and T. D. Schmittgen, "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method," *Methods,* vol. 25, no. 4, pp. 402-408, December 2001.

[55] N. Theera-Umpon, S. Chansareewittaya, and S. Auephanwiriyakul, "Phoneme and tonal accent recognition for Thai speech," *Expert Systems with Applications,* vol. 38, no. 10, pp. 13254-13259, September 2011.

[56] K. J. Lim, and A. V. Oppenheim, *Advanced Topics in Signal Processing,* Prentice Hall, 1988.

LIST OF PUBLICATION

1 R. Phettom, N. Theera-Umpon and S. Auephanwiriyakul, "Automatic Identification of Abnormal Lung Sounds Using Time-Frequency Analysis and Convolutional Neural Networks," 2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 26th-27th October 2023, pp. 317-322, 2023, doi: 10.1109/ICITEE59582.2023.10317776.

CURRICULUM VITAE

Author's Name	Mr. Rattanathon Phettom	
Place of Birth	Phrae Province, Thailand	
Education	Academic Year 2015	Bachelor of Science: Associated Medical Sciences majoring in Physical Therapy, Chiang Mai University, Chiang Mai
	Academic Year 2020	Master of Engineering (Biomedical Engineering): Biomedical Engineering Institute, Chiang Mai University, Chiang Mai
Scholarship	2020-2021 The Teaching Assistant and Research Assistant Scholarships from Graduate School of Chiang Mai University provided during the first two semesters of this study greatly contributed to the successful completion of this research.	
Publication	Rattanathon Phettom, Nipon Theera-Umpon and Sansanee Auephanwiriyakul, "Automatic Identification of Abnormal Lung Sounds Using Time-Frequency Analysis and Convolutional Neural Networks," 2023 15^{th} International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 26^{th}-27^{th} October 2023, pp. 317-322, 2023, doi: 10.1109/ICITEE59582.2023.10317776.	
Experience	August 2015-March 2019:	A physical therapy intern in the Associated Medical Sciences Faculty, Chiang Mai University, Chiang Mai, Thailand.

-Used advanced clinical reasoning skills and manual assessment techniques to provide accurate diagnosis.

-Diagnosed conditions through consultations, reviewing treatment methods, then developed tailored recovery plans.

-Reduced patients' pain and swelling through a variety of techniques including hot and cold packs, taping and splinting.

-Retrained patients to walk and use devices such as walking frames, crutches, and wheelchairs to improve mobility.

-Improved patients' range of motion and assisting in muscle re-education and strengthening through techniques including manual and electrotherapy.

-Educated patients, their families, and the wider community to lead healthy lifestyles to prevent injury and disability.

-Assessed patients' physical condition to identify issues and plan an appropriate treatment regimen.

August 2020-March 2024: A biomedical engineer candidate in the Computational Intelligence laboratory, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand.

-Digital signal processing on automatic identification of abnormal lung sounds using time-frequency analysis and convolutional neural networks.

-Analyzed data, reviewed results from tests conducted and generated reports using different types of software programs.

-Collaborated with medical, science, engineering, and research communities to propose solutions promoting quality healthcare.

-Evaluated new medical devices to test safety before releasing to public.

-Designed medical equipment, including prosthetics, artificial organs, and surgical instruments.