UNIVERSIDADE D
COIMBRA

Daniel Filipe Simões Bastos

Differential Diagnosis of Respiratory Pathologies: a Data-Driven Approach based on Respiratory Sounds

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with the specialization in Clinical Informatics and Bioinformatics, supervised by Prof. Dr Paulo Fernando Pereira de Carvalho and Dr Rui Pedro Pinto de Carvalho e Paiva.

October of 2021

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Daniel Filipe Simões Bastos

Differential Diagnosis of Respiratory Pathologies: a Data-Driven Approach based on Respiratory Sounds

Supervisors:

Prof. Dr. Paulo Fernando Pereira de Carvalho

Prof. Dr. Rui Pedro Pinto de Carvalho e Paiva

Jury:

Prof. Dr. João Pedro de Almeida Barreto

Prof. Dr. César Alexandre Domingues Teixeira

Prof. Dr. Rui Pedro Pinto de Carvalho e Paiva

Thesis submitted to the University of Coimbra for the degree of Master in Biomedical Engineering

Coimbra, October 2021

Acknowledgements

I would like to start by thanking Prof. Rui Paiva and Prof. Paulo de Carvalho for their continuous assistance during this path. Since the beginning, always gave me full support and helped me in everything I needed. Without them, I know it would have been much more difficult to finish this journey.

Secondly, I would like to thank Bruno Rocha and Diogo Pessoa for always being available to help me, every time I needed it.

Special recognition to my friends, new and old, that stayed by my side, always motivating me to go further.

Last but not least, I would like to thank my mom, dad, and sister. For always being there for me, supporting, taking care of me, and always giving me everything I needed.

Resumo

Doenças respiratórias são uma das principais causas de morte no mundo, causando uma série de problemas económicos e sociais. Para tentar contrariar isso, novos e melhores métodos de diagnóstico e acompanhamento têm sido procurados, havendo um grande investimento nessa área. A análise computacional de sons respiratórios é uma área que se tem desenvolvido bastante para tentar parar este tipo de doenças.

Esta dissertação tem o objetivo de avaliar a possibilidade de criação de um sistema de diagnóstico diferencial de patologias respiratórias, usando uma abordagem baseada em dados de sons sonoros. A metodologia adotada consistiu em extrair um conjunto de features das 3 bases de dados de sons respiratórios disponíveis, onde um total de 81 features são extraídas de cada gravação de som respiratório. De seguida foram usados 3 tipos de classificadores de machine learning para tentar classificar cada gravação em diferentes classes. Para além de usar todas as features para a classificação, métodos de seleção de features também foram usados de forma a tentar selecionar as melhores features e obter melhores resultados. Os melhores resultados foram possíveis usando a base de dados do ICBHI, classificando entre a classe de patologias crónicas e patologias não crónicas. No entanto estes foram os únicos resultados satisfatórios.

Concluindo, grande parte dos resultados tiveram uma performance bastante baixa, provando-se que, usando estes métodos e materiais, não é possível criar um sistema de diagnóstico de patologias respiratórias. Como trabalhos futuros propõe-se ser testado o uso de outras features, o uso de bases de dados maiores e mais balanceadas, e tendo bases de dados maiores, o uso de técnicas de deep learning.

Palavras-chave: Sons Respiratórios, Features, Classificação, Machine Learning,

Abstract

Respiratory diseases are one of the main causes of death in the world, causing a series of economical and social problems. Trying to counter that, new and better diagnosing and monitoring methods are being searched, having a great investment in that area. Computational analysis of respiratory sounds is an area that is being developed in order to try to resolve this problem.

This dissertation has the main goal of evaluating the possibility of the creation of a differential diagnosis system of respiratory pathologies, using respiratory sounds data. The adopted methods consist of extracting a set of features from the 3 available respiratory sounds databases, where a total of 81 features are extracted from each respiratory sound recording. Then 3 types of machine learning classifiers are used to try to classify each recording in different classes. Besides using all features, feature selection methods are also used in order to try to select the best features and achieve better results. The best results are achieved using the ICBHI database, classifying between Chronic and Non-Chronic pathologies. However, these were the only satisfactory results.

In conclusion, a big part of the results obtained have a low performance, proving that, using these methods and materials, it is not possible to create a differential diagnosis system of respiratory pathologies. As future work it is proposed that new features are tested, the use of bigger and better balanced databases, and using bigger databases, the use of deep learning techniques.

Keywords: Respiratory Sounds, Features, Classification, Machine Learning,

Contents

List of Figures

List of Tables

List of Acronyms

CNN	Convolution Neural Network
COPD	Chronic Obstructive Pulmonary Disease
ICBHI	International Conference on Biomedical and Health
KNN	K-Nearest Neighbor
LRTI	Lower Respiratory Tract Infection
MFCC	Mel-Frequency Cepstral Coefficient
mRMR	Minimum Redundancy Maximum Relevance
SVM	Support Vector Machine
URTI	Upper Respiratory Tract Infection

1 Introduction

1.1 Motivation and Objectives

Respiratory diseases are one of the biggest causes of death in the world, with Chronic Obstructive Pulmonary Disease (COPD) and lower respiratory infections being the fourth and fifth leading causes of death in the world in 2019 [World Health Organization Global Health Estimates 2020], creating a lot of stress to health systems.

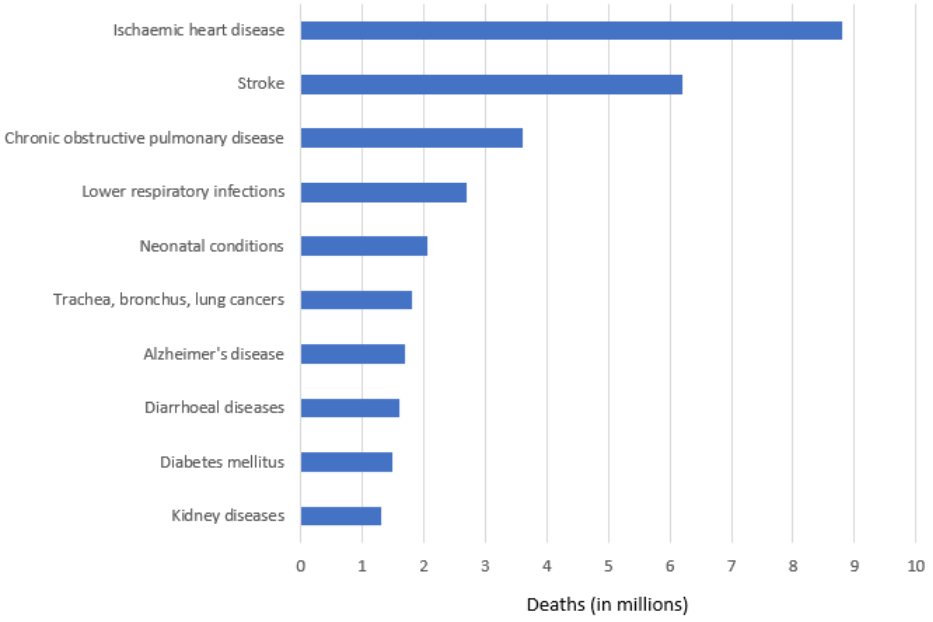

Figure 1.1: Top 10 causes of death in 2019 [1]

The best way to deal with respiratory diseases is with early diagnosis and routine monitoring and a lot of research is being projected for that.

When someone has a respiratory disease, changes in the respiratory sounds, appearing adventitious respiratory sounds, and lung ventilation can appear and these symptoms can lead to disease detection. Nowadays auscultation is the most used method for adventitious respiratory sounds detection. This is when a healthcare professional uses a stethoscope to listen to respiratory sounds coming from the chest or trachea. It is an easy, safe, non-invasive

and inexpensive method. However it has some limitations like the need to be done by an expert, it is not possible for continuous monitoring, its inherent inter-listener variability and audition limitations [2]. Spirometry measures how much air the patient can inhale and exhale as a function of time. This is a simple lung function measurement, however, it depends a lot on the effort and motivation of the patient, having great potential for error [3]. Methods that are objective, non-invasive, that enable continuous monitoring, with small margins for error are needed, and automated respiratory sound analysis could be one of them.

With this in mind, this thesis has the main goal to access the possibility to perform a differential diagnosis of respiratory pathologies, using respiratory sounds, more specifically:

- Summarize the State of the Art methods for differential diagnosis of respiratory pathologies;
- Testing different machine learning approaches to the differential diagnosis of respiratory pathologies with a specific set of features;
- Analyse the results;
- Dissertation writing.

1.2 Thesis Organization

This thesis is organized into 6 chapters:

- Chapter 1: Introduction to the document;
- Chapter 2: Definitions on respiratory sounds concepts;
- Chapter 3: Review of the State of the Art;
- Chapter 4: Materials and methods to reach the goal of this thesis;
- Chapter 5: Presentation of the obtained results and their discussion;
- Chapter 6: Conclusion of the developed work and future projects.

2 Background Concepts on Respiratory Sounds

In this chapter, concepts related to respiratory sounds are explored, which will be used throughout this dissertation.

2.1 Definitions on Respiratory Sounds

Respiratory sounds are the sounds produced when air moves through the respiratory system. Breathing sounds, adventitious sounds, coughing sounds, snoring sounds, sneezing sounds, and sounds from the respiratory muscles are all examples of respiratory sounds [4].

These sounds can give information about the physiological and pathological status of the respiratory system, so being able to distinguish normal respiratory sounds from adventitious sounds is the basis for a precise medical diagnosis.

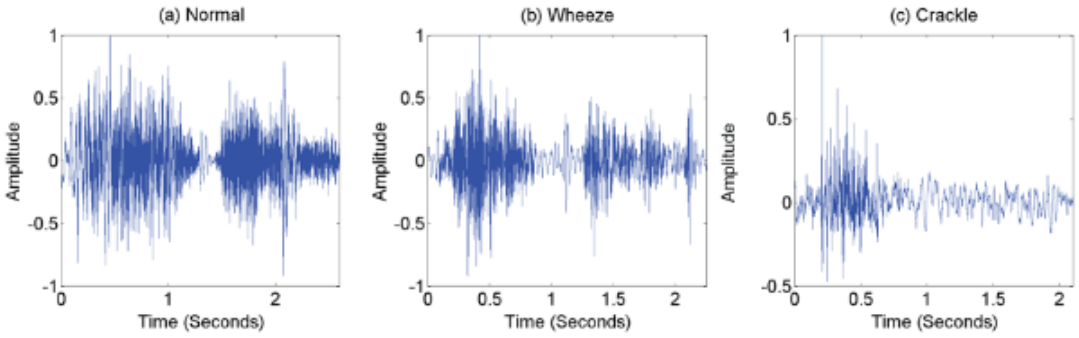

Figure 2.1: Time representation of a) normal b) wheezes c) crackles sounds [5]

2.1.1 Normal Respiratory Sounds

Normal respiratory sounds can be put into different categories, depending on the location where they are being heard or generated, and can be categorized as vesicular, bronchial,

tracheal, and mouth sounds. These sounds have a different pitch, duration, and sound quality, depending on the auscultation location [6].

Vesicular sounds are heard throughout most of the lung fields. These are soft, rustling, low pitched sounds that occur during inspiration and early parts of expiration, pausing between cycles. The frequencies of these sounds are lower, ranging from 100 to 1000 Hz, because the lung parenchyma and chest wall act as a low pass-filter [7][8].

Bronchial sounds are possible to hear in the large airways on chest near the 2^{nd} and 3^{rd} intercostal space. These sounds are loud, hollow, and high pitched, with frequencies that go from 100 to 5000 Hz. They occur both during inspiration and expiration, with a short pause between them [7][9].

Tracheal sounds are harsh, loud, high pitched, and can be heard in the suprasternal notch on the trachea. These sounds have a wide frequency range from 100 to 5000 Hz, occurring during inspiration and expiration, with a pause between them [7].

Mouth sounds are generated in the central airways and can travel upward and downward and have frequencies between 200 and 2000 Hz. Mouth sounds should be silent in a healthy individual [6].

2.1.2 Adventitious Respiratory Sounds

Adventitious sounds are respiratory sounds that overlap normal breath sounds. These sounds are generated when airway abnormalities appear in the respiratory system, normally related to some disease, meaning that the identification of adventitious sounds can be a useful tool for a correct medical diagnosis. These sounds can be divided into two categories: continuous adventitious sounds and discontinuous adventitious sounds [10].

Continuous Adventitious Sounds

Continuous adventitious sounds normally last more than 250 ms and can be divided according to their pitch, being high-pitched (Wheeze, Stridor, and Gasp) or low-pitched (Rhonchi and Squawk) [6].

Wheezes and **rhonchi** have some similar characteristics. They are both continuous sounds with a duration of over 80 ms, they are musical, sibilant, and can be heard during inspiration, but most commonly during expiration. They also have some differences. Wheezes are high-pitched, with frequencies over 400 Hz, while rhonchi are low-pitched with frequencies lower than 200 Hz. Wheezes can also be monophonic if they have one dominant

frequency, or polyphonic if they have more than one dominant frequency. Wheezes are a characteristic sign of airway obstruction, linked with asthma and Chronic Obstructive Pulmonary Disease (COPD), and ronchi usually is associated with bronchitis and COPD due to the secretions in the bronchial tree [6][8].

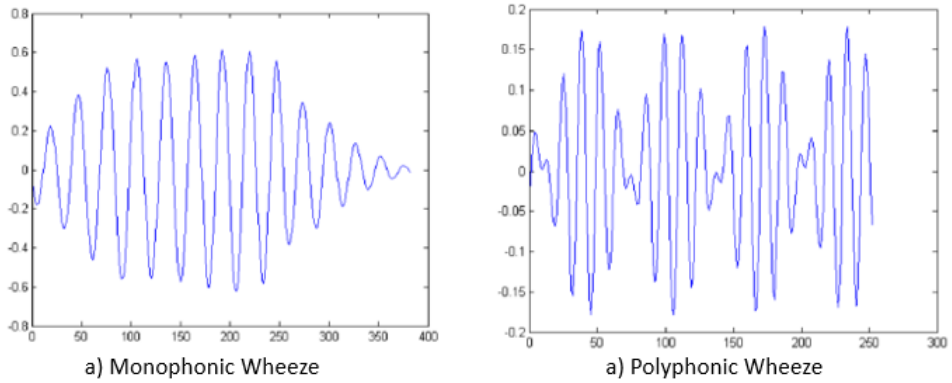

Figure 2.2: Time expanded waveform of a) monophonic b) polyphonic wheeze [11]

Stridors have more than 250 ms of duration and can mostly be heard during inspiration, but it is possible to hear them during expiration. These sounds are musical and have a high pitch, with a frequency of 500 Hz. These sounds are generated by turbulent airflow in the larynx or bronchial tree, which can be related to diseases that are related to upper airway obstruction, like epiglottitis, croup, and laryngeal edema [6].

Squawks have a duration of around 20 ms and are generated during inspiration. They are a mix of musical and non-musical sounds, have a low pitch with a frequency between 200 and 300 Hz. They are originated from oscillations of the peripheral airway and can be heard in patients with hypersensitivity pneumonia and common pneumonia [6].

Gasps are longer than 250 ms, are high pitched, and are heard during inspiration. The whoop sound of an inspiratory gasp is caused by fast moving air through the respiratory tract and is a symptom of whooping cough (pertussis) [6].

Discontinuous Adventitious Sounds

Discontinuous adventitious sounds are shorter than the continuous, lasting less than 25 ms. They can be divided into fine crackles, coarse crackles, and pleural rubs, depending on the source from where the sounds are generated [6].

Fine crackles last around 5 ms and happen during late inspiration. These sounds are explosive, non-musical, with a high pitch, having a frequency of around 650 Hz. They are generated by explosive openings of the small airways. Fine crackle sounds are usually

associated with pneumonia, congestive heart failure, and lung fibrosis [6].

Coarse crackles have a duration of around 15 ms, happening during the early phase of inspiration and expiration. They are low-pitched, with a frequency of 350 Hz. These sounds are generated by air bubbles in large bronchi and can be heard on patients with chronic bronchitis, bronchiectasis, and COPD [6].

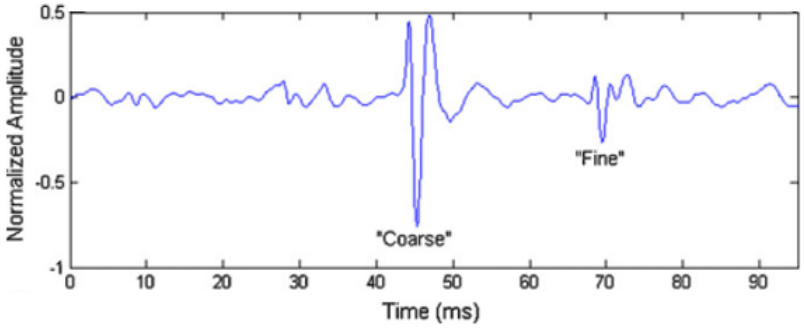

Figure 2.3: Time expanded waveform of part of a lung sound signal that includes a coarse and a fine crackle. [12]

Pleural rubs are longer than 15 ms and can be heard during inspiration and expiration. They are low-pitched, having a frequency lower than 350 Hz. These non-musical, rhythmic sounds are generated by pleural membranes rubbing against each other and can be caused by lung membrane inflammation or lung tumor [6].

2.2 Computerized Respiratory Sound Analysis

Nowadays, auscultation is the gold standard for lung sound classification. However, this method has some limitations, like the inter-listener variability. Computerized respiratory sound analysis (CORSA) has several advantages compared to the auscultation performed by the healthcare professional, namely better correlation with criteria that define respiratory sounds, reduced variability between listeners, and a more assertive and objective diagnosis of respiratory disorders. In 2000, the European Respiratory Society published a set of guidelines for research and clinical practice in the field of CORSA in order to facilitate and standardize the registering, processing, and analysis of respiratory sounds [4][3][13]. Some of the guidelines can be seen in Table 2.1.

Table 2.1: CORSA guidelines in sound acquisition [14]

Parameters	Values
Frequency Interval	100-5 kHz
Frequency Response	Flat in the frequency range of the sound; maximum deviation of 6 dB
Dynamic Range	>60 dB
Sensitivity	Independent of frequency, static pressure and sound direction
Signal-to-noise ratio	>60 dB
Directional characteristic	Omnidirectional
Sensor	Condenser microphone; Piezoeletric microphone, accelerometer
Protection against acoustic noise	Shielded microphones, protection from mechanical vibrations
Protection against electromagnetic interferences	Shielded twisted pair or coaxial cable
Microphones location	Trachea or thorax (posterior, anterior, lateral; left and right)

3 State of the Art

In this chapter, a review of the literature about classification of respiratory sounds into diseases is presented.

In [15] a comparison was made between SVM and KNN. These classifiers were used to distinguish normal, airway obstruction, and parenchymal pathologies. The sounds used are from the RALE database, which had 68 recordings, obtained from the chest wall using a contact accelerometer (EMT25C, Siemens). The sampling rate of the recordings was 10 kHz. After the sounds were obtained, they were filtered using a first-order Butterworth high-pass filter at 7.5 kHz, and an eight-order Butterworth low-pass filtered at 2.5 kHz. Then MFCC features are extracted from the recordings. These features are fed to a one-way ANOVA that showed that they were significantly different. For the SVM, a linear kernel function and a RBF kernel function was tried. With the KNN the number of neighbours tried went from 1 to 10. The KNN was able to get 98.26% accuracy, using 1 neighbour, and the SVM, using the RBF kernel function, got 92.19%.

In [16] a SVM and a Gaussian mixture model (GMM) classifiers were used. In this project sounds from 20 healthy patients and 20 pathological patients were used, where 10 of the pathological patients had an obstructive disease and the other 10 a restrictive disease. The sounds were recorded using a system that is composed of 14 air coupled electret microphones (SONY ECM-44 BPT) attached on the posterior chest wall.The sounds passed through an analog amplifier-filter with a gain of 100 and a pass band of 80 to 4000 Hz. Auto-regressive model parameters can be used as the mathematical features for the classification, and here a multivariate version of the auto-regressive model is used, namely, vector auto-regressive model. The sound data are modeled using a vector auto-regressive model, and its parameters are used as features. Two types of classification are used: one where it performs classification on the three classes (healthy, obstructive and restrictive), and an hierarchical classification, where the sounds are first classified into healthy or pathological classes, and then the ones in the pathological class are classified into obstructive or restrictive. The best results were

obtained using the hierarchical GMM with a total correct classification rate of 85%.

A similar approach than the previous one was studied in [17]. The same database was used here and in [16]. A vector auto-regressive model was used to feed a SVM and classify sounds, having two types of classification: classified into healthy or pathological, or classified into healthy, or bronchiectasis or interstitial pulmonary disease. Using the binary classification, it was possible to achieve sensitivity and specificity for both classes of 85% ± 8.2%. Using three classes classification healthy recall was 95% ± 5% and precision 76% ± 8.7%, the interstitial pulmonary disease recall and precision are 100% ± 0%, and bronchiectasis recall is 30% ± 15.3% and precision 75% ± 25%.

In [18], a digital signal processor is used to design an instrument that can acquire and classify sounds into healthy or pathological sounds. The sounds are captured using an electret microphone (Sony ECM-44) attached to the posterior chest wall. The sounds then pass through a band-pass filter that goes from 80 to 2000 Hz. An auto regressive modelling of the sounds was used to feed the classifiers. The classifiers used were a KNN, with 5 neighbours and Itakura, Euclidean and city-block distance measures, and a minimum distance based classifier, with the Mahalanobis distance measure. The classification happened in two modes. First it was done offline, using the leave-one-out method with 20 healthy subjects and 20 sick patients, to have a prior idea of the performance of the instrument. With this classification, the best results were with the KNN with Itakura distance measure, getting 97.5% accuracy. When classifying online, only the best classifiers were used. They were used to classify 13 pathological subjects and 12 healthy subjects. The best results were achieved with KNN classifier with city-block distance measure, getting 96% accuracy.

In [19] a database with 296 recordings, that is divided in three classes, is used. This database has 112 recordings from healthy patients, 84 from patients with bronchitis and 100 from patients with COPD. In this study, the researchers made a search for new informative features of pathological respiratory sounds. They extracted spectral, spectrogram, wavelet, MFCC and logarithmic (mel) filterbank energies. They also use a wide variety of classifiers: decision trees,discriminant analysis, SVM, logistic regression, KNN and ensemble learning, with different customizations, having a total of 21 learners. The best results were achieved using quadratic discriminant with a combination of wavelet and logarithmic (mel) filterbank energies features, with an accuracy of 93.2%. The most accurate classifiers were the quadratic discriminant and KNN. Class COPD was the one with better results.

In [20] is used the ICBHI database, which is also used in this dissertation. They started by extracting MFCC features from the sounds. Then these features suffered convolution

with 2D kernels, becoming 2D vectors. After that they were used as input to a convolutional neural network (CNN) that had four convolutional layers, each of them followed by one max-pooling layer and three fully connected layers. They were able to get an accuracy of 90.21% and a weighted average f1 score of 0.89. The authors of this work claim that, with these results, this algorithm is ready to be used as a classifying algorithm, however that is not my opinion because they are using a very unbalanced database. Figure 3.2 shows the proposed method in this work. When the data is divided in train and test, the COPD class has 159 recordings in test, while all the other five classes all combined have 25 recordings. Some of the other classes have only 3 recordings, and one of them does not get any right prediction by this algorithm. With this, it is not possible to claim that this algorithm is ready to be used in classification, it needs to be tested in a bigger and better balanced database.

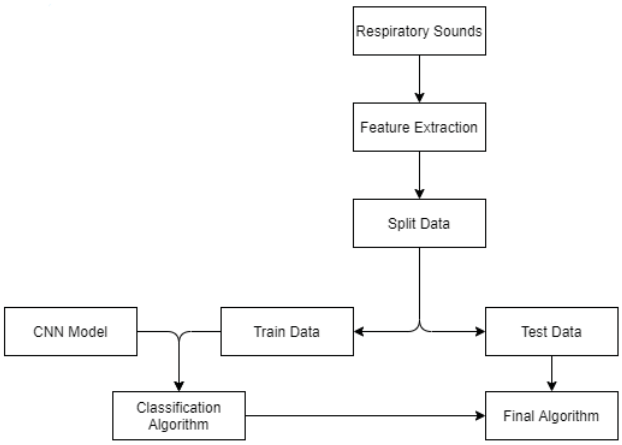

Figure 3.1: Proposed method in [20]

In [21] the ICBHI database is again used. In this experiment, the sounds go through a normalization process to remove noise from them. After that, the signals pass time stretching, pitch shifting, and dynamic range processes as a method of data augmentation. With this, the original 920 recordings become 11960. Spectrogram features are extracted from the audios and these extracted features are given as input to 2D CNN for classification. The 2D CNN architecture is composed of three convolutional layers, which are enclosed by max pool layer and finally, they are followed by two fully connected layers. With this method, they achieve an accuracy of 97%.

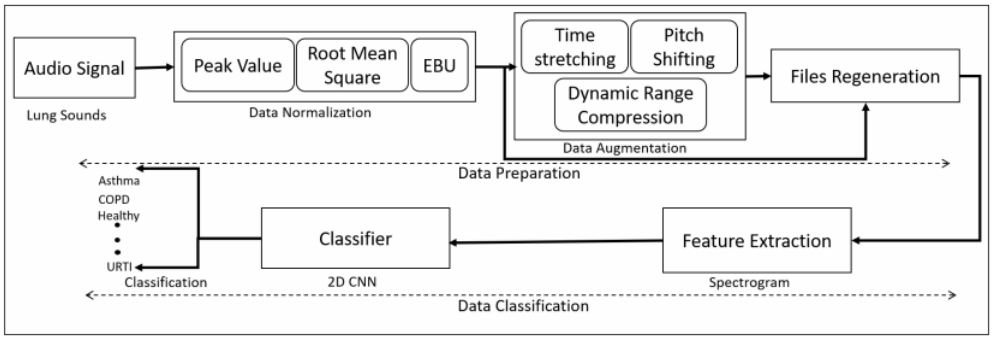

Figure 3.2: Proposed method in [21]

Similarly as in the previous study, in [22] the ICBHI database and a 2D CNN model are used. Data normalization and augmentation were used in the original audio files of the database. After that, MFCC features were extracted from the audios and given to the 2D CNN model. The CNN is composed of different convolution layers and max-pooling layers which are then followed by activation and fully connected layers. An accuracy of 92.39% was achieved.

[23] used a database composed of 30 COPD and 25 healthy subjects. A total of 42 features are extracted, of which 39 are lung sound features and 3 spirometry features. The lung sound features included temporal, spectral and spectro-temporal features. A variety of classifiers are used to classify the audios between COPD and normal, being those classifiers: SVM, KNN, decision tree, discriminant analysis and logistic regression. Different combinations of features were used. The best results were achieved using SVM using only spirometry features, and also with SVM and logistic regression using a combination of significant lung features and spirometry data, achieving an accuracy of 100%. Although these are very good results, they are obtained with a very small database.

In [24] the ICBHI database is used again. In this study, a two part method is used. First, the individual respiratory cycles are classified as having wheezes, crackles or being healthy, with a boosted decision tree, using lowlevel features, rythm features, SFX features and tonal features. The best results for this first phase are 85% accuracy. In the second part, the classifications of all the respiratory cycles from one patient are used to classify the patient as healthy or not. It is as if the features for this second part are the presence or absence of wheezes and crackles. To classify a patient as Healthy, the Healthy ratio needs to be bigger than the Sick ratio. Each ratio is calculated using Equation 3.1. With this method, it was possible to get 85% correct predictions.

$$RatioX = \frac{\#Cycles\ Predicted\ to\ be\ in\ Class\ X}{\#Total\ of\ Cycles\ Analyzed} \tag{3.1}$$

The presented studies have some limitations, like the use of small databases and databases that the sounds were recorded in very specific situations, making the results of those studies difficult to reproduce and generalize. Sometimes some studies also use very specific methods in their data, to get better results, but those methods only work for that specific data, making them useless for other studies.

In Appendix A it is possible to see a resume of all the studies seen in this chapter.

4 Materials and Methods

After reviewing the State of the Art, the methods that would be used in this dissertation were defined, using the available materials.

4.1 Databases

In this dissertation, 3 different databases were used, all of them having healthy patients and patients with some respiratory disease.

ICBHI Scientific Challenge Database

This database was created for an international competition, the first scientific challenge of the IFMBE's International Conference on Biomedical and Health Informatics (ICBHI). It includes 920 recordings from 126 participants annotated with 8 types of respiratory conditions including Healthy, Upper Respiratory Tract Infection (URTI), Asthma, Pneumonia, COPD, Lower Respiratory Tract Infection (LRTI), Bronchiectasis, and Bronchiolitis. The recordings were collected using four types of equipment: AKG C417L Microphone, 3M Littmann Classic II SE Stethoscope, 3M Littmann 3200 Electronic Stethoscope, and Welch Allyn Meditron Master Elite Electronic Stethoscope. The sounds were collected from six chest locations: left and right anterior, left and right posterior, and left and right lateral. Each recording had an annotation with the beginning and ending of respiratory cycles and the location of wheezes and crackles.

Table 4.1 summarizes the number of patients and recordings per class.

The eight conditions from this database could also be grouped into 3 classes: Healthy, Non-Chronic (URTI, LRTI, Pneumonia and Bronchiolitis) and Chronic (COPD, Bronchiectasis and Asthma). Table 4.2 summarizes the number of patients and recordings, divided into three classes.

By looking at Table 4.1 it is possible to see that there is an unbalancing problem with

Table 4.1: Summarized information about ICBHI Scientific Challenge database

Condition	# Patients	# Recordings
Healthy	26	35
URTI	14	23
Asthma	1	1
COPD	64	793
LRTI	2	2
Bronchiectasis	7	16
Pneumonia	6	37
Bronchiolitis	6	13

the condition COPD, which has a lot more patients and recording than the other conditions, having 86,2% of the total recordings of the database. This leads to an undersampling process that is explained better further ahead. Also, almost all of the patients from the Health class only had one recording, leading to few recordings from Healthy patients.

By analyzing the individual biometric information of the patients, it was possible to see that from the 26 Healthy subjects, 22 were children with ages below 16 years old.

Table 4.2: Summarized information about ICBHI Scientific Challenge database, with three classes

Condition	# Patients	# Recordings
Healthy	26	35
Chronic	72	810
Non-Chronic	28	75

In Table 4.3 it is possible to see the biometric data of the patients from the ICBHI database.

Thessaloniki Database

This database includes 613 recordings from 31 patients. These patients include a wide variety of conditions, so they were grouped into 4 classes: Healthy, Non-Chronic, Chronic/Obstructive and Interstitial. For the recording of the sounds a 3M Littmann 3200 Electronic Stethoscope, and a Welch Allyn Meditron Master Elite Electronic Stethoscope were used. The recordings

Table 4.3: ICBHI Database biometric information (NA: not available)

Number of participants	126; 77 adults, 49 children
Sex	79 male, 46 female (NA: 1)
Age	43.0 ± 32.2 years (NA: 1)
Age of adults	67.6 ± 11.6 years (NA: 1)
Age of children	4.8 ± 4.6 years
BMI of adults	27.2 ± 5.4 kg m^{-2} (NA: 2)
Weight of children	21.4 ± 17.2 kg (NA: 5)
Height of children	104.7 ± 30.8 cm (NA: 7)

Table 4.4: Summarized information about Thessaloniki database

Condition	# Patients	# Recordings
Healthy	8	178
Non-Chronic	10	156
Chronic/Obstructive	6	156
Interstitial	7	147

were collected from nine chest and trachea locations: trachea, left and right anterior, left and right posterior upper, left and right posterior bottom and left and right lateral.

In Table 4.5 it is possible to see the biometric data of the patients from the Thessaloniki database.

Table 4.5: Thessaloniki Database biometric information (NA: not available)

Number of participants	32; all adults
Sex	20 male, 12 female
Age	61.4 ± 17.3 years
BMI	29.4 ± 5.0 kg m (NA:2)

WELCOME Database

This database includes 200 recordings from 50 patients. In this database there were 3 conditions: Healthy, Asthma and Bronchitis. A 3M Littmann 3200 Electronic Stethoscope was used to record the sounds. For the auscultation, it was used the posterior inferior lobe site of the left or right lung, and the posterior middle lobe site of the left or right lung.

Table 4.6: Summarized information about the WELCOME database

Condition	# Patients	# Recordings
Healthy	46	184
Asthma	3	12
Bronchitis	1	4

In Table 4.7 it is possible to see the biometric data of the patients from the WELCOME database.

Table 4.7: WELCOME Database biometric information

Number of participants	50; 44 adults, 6 children
Sex	26 male, 24 female
Age	31.7 ± 14.8 years
BMI	24.9 ± 4.2 kg m^{-2}
Age of adults	34.1 ± 14.0 years
Age of children	13.7 ± 3.3 years

4.2 Extracted Features

In the first phase, all recordings are resampled to the same sample rate of 4000 Hz. This happened because, according to the Nyquist theorem, a signal must be sampled at more than twice the highest frequency component of the signal [25]. The maximum frequency of adventitious sounds is 2 kHz, so the sample rate must be 4 kHz. A recording has a lot of deviations across its entire length, making it difficult to be analyzed. So, each recorded is decomposed into smaller frames to make it easier to analyze. The frames have a length of 128 ms, with an overlap of 75%. Afterward, the features were extracted from each frame.

A total of 81 features were extracted from each frame and they can be divided into 3 groups: spectral, mel-frequency cepstral coefficients (MFCCs) and melodic. These features were extracted with the help of MIRToolbox from Matlab.

- Spectral features - estimated from the spectrogram of the sound;
- MFCCs features - the most common features used to describe the spectral shape of a sound;

- Melodic features - use the pitch of the sound as a base to be calculated;

Table 4.8 shows a brief description of all the 81 features extracted.

For each recording, four statistics of each feature are calculated: mean, standard deviation, minimum value, and maximum value. Consequently, the total number of features, for each recording, given to the classifiers is 324.

Table 4.8: List of all features extracted from frames with a brief description [2]

Type	Feature	Description
Spectral	Spectral Centroid	Center of mass of the spectral distribution
	Spectral Spread	Variance of the spectral distribution
	Spectral Skewness	Skewness of the spectral distribution
	Spectral Kurtosis	Excess kurtosis of the spectral distribution
	Zero-crossing Rate	Waveform sign-change rate
	Spectral Entropy	Estimation of the complexity of the spectrum
	Spectral Flatness	Estimation of the noisiness of a spectrum
	Spectral Roughness	Estimation of the sensory dissonance
	Spectral Irregularity	Estimation of the spectral peaks' variability
	Spectral Flux	Euclidean distance between the spectrum of successive frames
	Spectral Flux Inc	Spectral flux with focus on increasing energy solely
	Spectral Flux Halfwave	Halfwave rectified spectral flux
	Spectral Flux Median	Median filtered spectral flux
	Spectral Brightness	Amount of energy above 100, 200, 400, and 800 Hz
	Brightness 400 Ratio	Ratio between spectral brightness at 400 and 100 Hz
	Brightness 800 Ratio	Ratio between spectral brightness at 800 and 100 Hz
	Spectral Rolloff	Frequency such that 95, 75, 25, and 5% of the total energy is contained below it
	Rolloff Outlier Ratio	Ratio between spectral rolloff at 5 and 95%
	Rolloff Interquartile Ratio	Ratio between spectral rolloff at 25 and 75%
MFCC	MFCC	13 Mel-frequency cepstral coefficients
	Delta-MFCC	1st-order temporal differentiation of the MFCCs
Melodic	Pitch	Fundamental frequency estimation
	Pitch Smoothing	Moving average of the pitch curve with lengths of 100, 250, 500, and 1000 ms
	Inharmonicity	Partials non-multiple of fundamental frequency
	Inharmonicity Smoothing	Moving average of the inharmonicity curve with lengths of 100, 250, 500, and 1000 ms
	Voicing	Presence of fundamental frequency
	Voicing Smoothing	Moving average of the voicing curve with lengths of 100, 250, 500, and 1000 ms

4.3 Classifiers

In this section, the classifiers used in this study to classify respiratory sounds are explained. The classifiers used are support vector machine (SVM), k-nearest neighbor (KNN) and decision tree.

SVM

SVM is a supervised learning method used for classification and regression. The main goal of SVM is, given a set of labeled data, to find the optimal hyperplane that separates classes, maximizing its margins. In the most simple cases where there are only two dimensions and the data can be linearly separable, that hyperplane can be a simple straight line. However, in more complicated cases, it is needed to map the given data to higher dimensions in order to find the optimal hyperplane. New examples given to the SVM are predicted depending on what side of the hyperplane they are.

KNN

KNN is a non-parametric classification method used for classification and regression. In the training phase, the feature vectors and labels are saved. In the classification phase, a feature vector is classified by assigning the label that is most frequent among the k training samples nearest to it, being k a positive integer, typically small.

Decision Tree

Decision tree is a non-parametric supervised learning method that can be used in machine learning for classification. It creates a predictive model that starts by splitting the root node of the tree (features) into subsets, following a specific splitting rule. The derived subsets are then split, repeating this process in a recursive manner, until splitting no longer adds value to the predictions.

4.4 Feature Selection

After classifying using all the features, an attempt to increase the performance was made by using only the best features. For this, two methods of feature selection were used: Minimum Redundancy Maximum Relevance (mRMR) and ReliefF.

mRMR

This method ranks the features taking into account the maximum relevance to classification, and with minimum redundancy with other features. This method starts by picking the most relevant feature of all. Then picks the feature with the maximum relevance, and with minimum redundancy with the feature previously picked. And keeps doing this iteration until all features are selected. The features that are chosen first are the ones with the best rank [26].

ReliefF

ReliefF gives weights to each feature according to the difference of the feature value between nearest neighbor instance pairs. A sample is chosen randomly from the data set. Then two more samples are chosen, the nearest sample with the same class and the nearest sample with another class. If the value of a feature from the randomly picked sample and the value of the same feature from the nearest sample with the same class are different, the weight of that feature is decreased. If the value of a feature from the randomly picked sample and the value of the same feature from the nearest sample with a different class are different, the weight of that feature is increased [27].

For each experiment, seven subsets of features were selected: the best 10, 30 and 75 features selected by mRMR, the best 10, 30 and 75 features selected by ReliefF, and all 324 features.

4.5 Classification

Before starting the training and testing process, the data is divided into train set and test set. For the three databases, it is needed that recordings from the same patient would be all placed in the same set, so the division is made by dividing the patients into train and test sets, not the recordings. It is also necessary that the train and test set have all of the conditions from each database.

As said in the description of the ICBHI database in the Materials and Methods chapter, there is an unbalancing problem with class COPD. Because of this, there is an undersampling process where only 10 patients with COPD, out of the total of 64, are used.

For the division of patients into train and test sets, the hold-out method is used. To

guarantee that every patient is part of the train and test set at least once, this hold-out process is repeated 100 times. In each time, 70% of the patients, from each condition, are randomly picked for training. The recordings of these patients are used to train the three different classifiers (SVM, KNN and decision tree), and then tested on the remaining 30% of patients. After each hold-out, the results are saved and the mean of the 100 iterations is presented in chapter Results. In Table 4.9 it is possible to see the distribution of patients in the train and test sets.

Table 4.9: Distribution of patients in train and data set

Database	Class	# Patients Train Set	# Patients Test Set
ICBHI	Healthy	18	8
	URTI	10	4
	LRTI	1	1
	Asthma	1	0
	COPD	7	3
	Bronchiectasis	5	2
	Pneumonia	4	2
	Bronchiolitis	4	2
Thessaloniki	Healthy	6	2
	Chronic	4	2
	Non-Chronic	7	3
	Intersticial	5	2
Lucio	Healthy	32	14
	Asthma	2	1
	Bronchitis	1	0

In each hold-out, after the train set is made, it is divided into train and validation sets, using a hold-out method, with 70% going to train and 30% going to validation. Then a parameter optimization process is started. Using a Bayesian optimization process, the classifiers are trained 100 times, each time with different parameter combinations, testing it in the validation set. At the end of those 100 times, the best parameter combination is used in the test set.

4.6 Performance Metrics

The results from every experiment are presented with the following parameters: precision, recall, specificity, f1 score and accuracy, with the first four being calculated for each class and the weighted mean of all classes total, and the last one is calculated for the total.

Precision

Corresponds to the proportion of cases that really belong to class X, from all the ones that were classified as X.

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

Recall

Corresponds to the proportion of cases correctly classified as class X, from all the ones that are from class X.

$$Recall = \frac{TP}{TP + FN} \tag{4.2}$$

Specificity

Corresponds to the proportion of cases correctly not classified as class X, from all the ones that are not from class X.

$$Specificity = \frac{TN}{TN + FP} \tag{4.3}$$

F1 Score

Measure that combines precision and recall:

$$F1Score = 2 \times Precision \times \frac{Recall}{Precision + Recall} \tag{4.4}$$

Accuracy

Corresponds to the proportion of cases correctly classified, from all the ones that are classified.

$$Accuracy = \frac{\#Correct\ Predictions}{\#Total\ Predictions} \tag{4.5}$$

5 Results and Discussion

In this chapter the results of the classification of the recordings are presented, as well as their discussion.

5.1 ICBHI Database

In this section, the results obtained using only the ICBHI Database are presented.

Healthy vs Sick

For this experiments, all of the classes that were different from the class Healthy, were converted to the class Sick. First starting by using all of the features, and then using a smaller set of features. The results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.1, 5.2, 5.3, 5.4, 5.5, 5.6 and 5.7, respectively.

Table 5.1: Results for Healthy vs Sick with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.02	0.98	0.01	0.83
	Sick	0.84	0.98	0.02	0.90	
	Weighted Average	0.71	0.83	0.17	0.76	
KNN	Healthy	0.07	0.05	0.94	0.04	0.80
	Sick	0.84	0.94	0.05	0.88	
	Weighted Average	0.72	0.80	0.19	0.75	
Decision Tree	Healthy	0.06	0.07	0.93	0.06	0.79
	Sick	0.84	0.93	0.07	0.88	
	Weighted Average	0.72	0.79	0.20	0.75	

Table 5.2: Results for Healthy vs Sick with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.00	0.99	0.01	0.85
	Sick	0.86	0.99	0.00	0.92	
	Weighted Average	0.73	0.85	0.15	0.79	
KNN	Healthy	0.08	0.09	0.92	0.06	0.80
	Sick	0.84	0.92	0.09	0.88	
	Weighted Average	0.73	0.80	0.21	0.76	
Decision Tree	Healthy	0.04	0.04	0.97	0.04	0.84
	Sick	0.86	0.97	0.04	0.91	
	Weighted Average	0.74	0.84	0.18	0.78	

Table 5.3: Results for Healthy vs Sick with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.01	0.99	0.01	0.85
	Sick	0.86	0.99	0.01	0.92	
	Weighted Average	0.74	0.85	0.15	0.79	
KNN	Healthy	0.04	0.04	0.96	0.03	0.83
	Sick	0.86	0.96	0.04	0.90	
	Weighted Average	0.74	0.83	0.17	0.78	
Decision Tree	Healthy	0.02	0.02	0.97	0.02	0.83
	Sick	0.85	0.97	0.02	0.91	
	Weighted Average	0.74	0.83	0.16	0.78	

Table 5.4: Results for Healthy vs Sick with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.00	0.00	0.99	0.00	0.85
	Sick	0.85	0.99	0.00	0.92	
	Weighted Average	0.73	0.85	0.15	0.79	
KNN	Healthy	0.07	0.05	0.96	0.04	0.83
	Sick	0.86	0.96	0.05	0.90	
	Weighted Average	0.74	0.83	0.18	0.78	
Decision Tree	Healthy	0.02	0.02	0.97	0.02	0.83
	Sick	0.85	0.97	0.02	0.91	
	Weighted Average	0.74	0.83	0.16	0.78	

Table 5.5: Results for Healthy vs Sick with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.00	1.00	0.00	0.85
	Sick	0.85	1.00	0.00	0.92	
	Weighted Average	0.73	0.85	0.15	0.78	
KNN	Healthy	0.04	0.06	0.95	0.04	0.81
	Sick	0.85	0.95	0.06	0.89	
	Weighted Average	0.73	0.81	0.19	0.77	
Decision Tree	Healthy	0.04	0.03	0.97	0.03	0.83
	Sick	0.85	0.97	0.03	0.91	
	Weighted Average	0.73	0.83	0.17	0.78	

Table 5.6: Results for Healthy vs Sick with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.07	0.05	0.96	0.04	0.83
	Sick	0.86	0.96	0.05	0.89	
	Weighted Average	0.74	0.83	0.18	0.77	
KNN	Healthy	0.07	0.09	0.90	0.07	0.78
	Sick	0.85	0.90	0.09	0.87	
	Weighted Average	0.74	0.78	0.21	0.75	
Decision Tree	Healthy	0.05	0.04	0.96	0.04	0.82
	Sick	0.85	0.96	0.04	0.90	
	Weighted Average	0.73	0.82	0.18	0.77	

Table 5.7: Results for Healthy vs Sick with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.09	0.03	0.98	0.04	0.86
	Sick	0.87	0.98	0.03	0.92	
	Weighted Average	0.78	0.86	0.15	0.81	
KNN	Healthy	0.06	0.05	0.95	0.04	0.84
	Sick	0.87	0.95	0.05	0.91	
	Weighted Average	0.77	0.84	0.16	0.80	
Decision Tree	Healthy	0.05	0.06	0.94	0.05	0.83
	Sick	0.87	0.94	0.06	0.90	
	Weighted Average	0.77	0.83	0.17	0.80	

By looking at these results is possible to see that, in every test, the results for the Healthy class are very low, meaning that it is being difficult for the algorithms to identify the Healthy class and predict almost everything to be in the Sick class. To try to counter this, a few more experiences are made. During training, the misclassification cost for wrongly predicting a Healthy recording as a Sick recording is increased. The misclassification costs used are 2, 4 and 6. The results can be seen in Tables 5.8, 5.9 and 5.10, respectively.

Table 5.8: Results for Healthy vs Sick with missclassification cost 2

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.05	0.08	0.91	0.06	0.79
	Sick	0.85	0.91	0.08	0.87	
	Weighted Average	0.73	0.79	0.21	0.75	
KNN	Healthy	0.10	0.15	0.82	0.11	0.72
	Sick	0.84	0.82	0.15	0.82	
	Weighted Average	0.73	0.72	0.25	0.72	
Decision Tree	Healthy	0.11	0.20	0.80	0.13	0.70
	Sick	0.85	0.80	0.20	0.81	
	Weighted Average	0.74	0.70	0.29	0.71	

Table 5.9: Results for Healthy vs Sick with missclassification cost 4

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.11	0.24	0.75	0.14	0.67
	Sick	0.79	0.75	0.24	0.76	
	Weighted Average	0.69	0.67	0.32	0.67	
KNN	Healthy	0.15	0.35	0.63	0.19	0.59
	Sick	0.84	0.63	0.35	0.71	
	Weighted Average	0.74	0.59	0.39	0.63	
Decision Tree	Healthy	0.17	0.37	0.66	0.22	0.62
	Sick	0.85	0.66	0.37	0.73	
	Weighted Average	0.75	0.62	0.42	0.65	

Table 5.10: Results for Healthy vs Sick with missclassification cost 6

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.14	0.57	0.42	0.20	0.44
	Sick	0.51	0.42	0.57	0.45	
	Weighted Average	0.46	0.44	0.55	0.41	
KNN	Healthy	0.14	0.55	0.43	0.21	0.45
	Sick	0.84	0.43	0.55	0.54	
	Weighted Average	0.74	0.45	0.53	0.49	
Decision Tree	Healthy	0.15	0.45	0.56	0.21	0.55
	Sick	0.86	0.56	0.45	0.66	
	Weighted Average	0.76	0.55	0.46	0.60	

In all these experiments, the algorithms are not able to identify the Healthy class, attributing almost all of the recordings to the Sick class. The biggest f1 score value for class Healthy is 0.07, seen in 5.6, with KNN and top 35 mRMR features. Even when the misclassification costs are increased, that f1 score value only achieves 0.22, seen in Table 5.9, but it lowers the Sick class metrics. Although the weighted average f1 score and accuracy levels are acceptable, that is only because the number of recordings from Sick patients is much superior to the number of Healthy patients recordings, and this imbalance in classes might be one of the problems that are causing these results. These results show that in this database, with this type of classifiers and features it is not possible to separate a recording of a Healthy subject from a recording of a Sick patient. In conclusion to this experiment, the results are not satisfactory.

Healthy vs Chronic vs Non-Chronic

In this section, the pathologies are combined in two categories: Chronic, which include COPD, Bronchiectasis and Asthma, and Non-Chronic, which include URTI, LRTI, Pneumonia and Bronchiolitis. Like in the previous section, the results using all features, top 10 ReliefF, top 35 ReliefF, top 75ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.11, 5.12, 5.13, 5.14, 5.15, 5.16 and 5.17, respectively.

The best results in these experiments are achieved with 0.73 weighted average f1 score, in Table 5.11, with SVM and all features, and also has the best f1 score value for each class, from these experiments. In this case, the use of a smaller set of features, chosen by feature

selection, was not useful to get better performance. From the three classes, the class Chronic is the one that is better identified, having always a better f1 score than the other two classes, and the class Healthy is the one with the worst results. From all of the classifiers, there is not any of them that stands out as the best. Even though 0.73 was achieved for the weighted average f1 score, these results are not satisfactory, because, for this purpose, the results need to be higher.

Table 5.11: Results for Healthy vs Chronic vs Non-Chronic with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.59	0.55	0.99	0.56	0.79
	Non-Chronic	0.59	0.56	0.98	0.57	
	Chronic	0.77	0.96	0.56	0.84	
	Weighted Average	0.70	0.79	0.75	0.73	
KNN	Healthy	0.33	0.19	0.97	0.22	0.65
	Non-Chronic	0.62	0.42	0.90	0.47	
	Chronic	0.65	0.88	0.41	0.74	
	Weighted Average	0.61	0.65	0.63	0.60	
Decision Tree	Healthy	0.20	0.13	0.99	0.14	0.64
	Non-Chronic	0.49	0.47	0.83	0.45	
	Chronic	0.66	0.86	0.45	0.73	
	Weighted Average	0.56	0.64	0.64	0.57	

Table 5.12: Results for Healthy vs Chronic vs Non-Chronic with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.38	0.39	0.96	0.37	0.67
	Non-Chronic	0.46	0.41	0.93	0.41	
	Chronic	0.72	0.89	0.46	0.76	
	Weighted Average	0.60	0.67	0.67	0.60	
KNN	Healthy	0.19	0.17	0.99	0.18	0.65
	Non-Chronic	0.40	0.30	0.93	0.33	
	Chronic	0.67	0.93	0.30	0.77	
	Weighted Average	0.54	0.65	0.57	0.57	
Decision Tree	Healthy	0.18	0.08	0.99	0.10	0.62
	Non-Chronic	0.36	0.21	0.96	0.25	
	Chronic	0.62	0.96	0.20	0.75	
	Weighted Average	0.49	0.62	0.53	0.52	

Table 5.13: Results for Healthy vs Chronic vs Non-Chronic with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.15	0.08	0.99	0.09	0.58
	Non-Chronic	0.39	0.33	0.87	0.33	
	Chronic	0.62	0.87	0.31	0.67	
	Weighted Average	0.48	0.58	0.59	0.48	
KNN	Healthy	0.23	0.15	0.98	0.16	0.60
	Non-Chronic	0.42	0.40	0.83	0.40	
	Chronic	0.64	0.81	0.41	0.69	
	Weighted Average	0.52	0.60	0.62	0.53	
Decision Tree	Healthy	0.19	0.14	0.99	0.15	0.61
	Non-Chronic	0.44	0.48	0.81	0.45	
	Chronic	0.66	0.84	0.47	0.72	
	Weighted Average	0.52	0.61	0.66	0.55	

Table 5.14: Results for Healthy vs Chronic vs Non-Chronic with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.20	0.14	0.99	0.15	0.61
	Non-Chronic	0.47	0.34	0.91	0.37	
	Chronic	0.61	0.93	0.33	0.72	
	Weighted Average	0.53	0.61	0.62	0.54	
KNN	Healthy	0.29	0.18	0.99	0.19	0.62
	Non-Chronic	0.54	0.39	0.88	0.41	
	Chronic	0.64	0.89	0.39	0.72	
	Weighted Average	0.58	0.62	0.64	0.56	
Decision Tree	Healthy	0.35	0.27	0.98	0.29	0.67
	Non-Chronic	0.48	0.49	0.86	0.47	
	Chronic	0.71	0.87	0.51	0.75	
	Weighted Average	0.60	0.67	0.68	0.61	

Table 5.15: Results for Healthy vs Chronic vs Non-Chronic with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.20	0.10	1.00	0.11	0.59
	Non-Chronic	0.38	0.24	0.94	0.27	
	Chronic	0.59	0.95	0.23	0.71	
	Weighted Average	0.47	0.59	0.57	0.49	
KNN	Healthy	0.07	0.07	0.98	0.06	0.57
	Non-Chronic	0.44	0.28	0.88	0.32	
	Chronic	0.59	0.88	0.28	0.69	
	Weighted Average	0.49	0.57	0.57	0.49	
Decision Tree	Healthy	0.20	0.13	0.98	0.14	0.60
	Non-Chronic	0.51	0.38	0.88	0.40	
	Chronic	0.61	0.87	0.37	0.70	
	Weighted Average	0.53	0.60	0.63	0.53	

Table 5.16: Results for Healthy vs Chronic vs Non-Chronic with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.48	0.12	1.00	0.17	0.62
	Non-Chronic	0.70	0.26	0.94	0.33	
	Chronic	0.60	0.95	0.25	0.73	
	Weighted Average	0.64	0.62	0.57	0.54	
KNN	Healthy	0.23	0.11	0.98	0.12	0.61
	Non-Chronic	0.56	0.30	0.89	0.36	
	Chronic	0.62	0.90	0.31	0.72	
	Weighted Average	0.56	0.61	0.57	0.54	
Decision Tree	Healthy	0.25	0.15	0.99	0.17	0.64
	Non-Chronic	0.44	0.40	0.87	0.40	
	Chronic	0.65	0.88	0.39	0.73	
	Weighted Average	0.55	0.64	0.61	0.57	

Table 5.17: Results for Healthy vs Chronic vs Non-Chronic with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.50	0.18	0.99	0.23	0.64
	Non-Chronic	0.63	0.30	0.95	0.37	
	Chronic	0.62	0.96	0.30	0.74	
	Weighted Average	0.62	0.64	0.60	0.56	
KNN	Healthy	0.21	0.21	0.97	0.20	0.62
	Non-Chronic	0.54	0.31	0.93	0.34	
	Chronic	0.64	0.92	0.33	0.73	
	Weighted Average	0.57	0.62	0.62	0.54	
Decision Tree	Healthy	0.24	0.14	0.99	0.16	0.63
	Non-Chronic	0.52	0.44	0.85	0.44	
	Chronic	0.65	0.88	0.44	0.73	
	Weighted Average	0.57	0.63	0.65	0.57	

Chronic vs Non-Chronic

From the previous experiments it was noticeable that the classifiers are having trouble identifying the Healthy class. With that in mind, in this section only the classes Non-Chronic and Chronic are used, leaving the class Healthy out. The experiments using all features, top 10 ReliefF, top 35 ReliefF, top 75ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.18, 5.19, 5.20, 5.21, 5.22, 5.23 and 5.24, respectively.

Using the top 75 ReliefF features and a KNN, the best result was achieved with a weighted average f1 score of 0.81. The results with all features were the worse, meaning that the feature selection methods were helpful this time. Class Chronic is once again the best class in all experiments. From the three classifiers, KNN is the one that got better results most of the time. In conclusion to this experiment, the results are satisfactory.

Table 5.18: Results for Chronic vs Non-Chronic with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.29	0.30	0.74	0.27	0.56
	Chronic	0.61	0.74	0.30	0.64	
	Weighted Average	0.50	0.56	0.47	0.51	
KNN	Non-Chronic	0.42	0.37	0.68	0.36	0.56
	Chronic	0.62	0.68	0.37	0.63	
	Weighted Average	0.56	0.56	0.48	0.54	
Decision Tree	Non-Chronic	0.35	0.36	0.63	0.33	0.52
	Chronic	0.60	0.63	0.36	0.59	
	Weighted Average	0.52	0.52	0.48	0.49	

Table 5.19: Results for Chronic vs Non-Chronic with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.42	0.28	0.91	0.32	0.67
	Chronic	0.66	0.91	0.28	0.76	
	Weighted Average	0.58	0.67	0.52	0.60	
KNN	Non-Chronic	0.64	0.49	0.87	0.52	0.74
	Chronic	0.73	0.87	0.49	0.79	
	Weighted Average	0.71	0.74	0.63	0.70	
Decision Tree	Non-Chronic	0.63	0.43	0.86	0.48	0.70
	Chronic	0.71	0.86	0.43	0.76	
	Weighted Average	0.70	0.70	0.59	0.67	

Table 5.20: Results for Chronic vs Non-Chronic with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.74	0.48	0.94	0.54	0.78
	Chronic	0.77	0.94	0.48	0.84	
	Weighted Average	0.79	0.78	0.65	0.74	
KNN	Non-Chronic	0.72	0.62	0.87	0.63	0.79
	Chronic	0.81	0.87	0.62	0.83	
	Weighted Average	0.78	0.79	0.69	0.77	
Decision Tree	Non-Chronic	0.58	0.46	0.92	0.49	0.77
	Chronic	0.76	0.92	0.46	0.82	
	Weighted Average	0.71	0.77	0.61	0.72	

Table 5.21: Results for Chronic vs Non-Chronic with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.69	0.45	0.94	0.50	0.76
	Chronic	0.75	0.94	0.45	0.83	
	Weighted Average	0.74	0.76	0.63	0.71	
KNN	Non-Chronic	0.82	0.71	0.89	0.73	0.82
	Chronic	0.85	0.89	0.71	0.85	
	Weighted Average	0.84	0.82	0.78	0.81	
Decision Tree	Non-Chronic	0.71	0.52	0.92	0.57	0.78
	Chronic	0.78	0.92	0.52	0.83	
	Weighted Average	0.77	0.78	0.66	0.75	

Table 5.22: Results for Chronic vs Non-Chronic with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.67	0.25	0.96	0.33	0.68
	Chronic	0.67	0.96	0.25	0.78	
	Weighted Average	0.69	0.68	0.53	0.61	
KNN	Non-Chronic	0.71	0.57	0.86	0.60	0.75
	Chronic	0.77	0.86	0.57	0.80	
	Weighted Average	0.76	0.75	0.68	0.73	
Decision Tree	Non-Chronic	0.59	0.47	0.85	0.49	0.71
	Chronic	0.72	0.85	0.47	0.77	
	Weighted Average	0.69	0.71	0.62	0.67	

Table 5.23: Results for Chronic vs Non-Chronic with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.79	0.41	0.95	0.49	0.75
	Chronic	0.74	0.95	0.41	0.82	
	Weighted Average	0.77	0.75	0.60	0.71	
KNN	Non-Chronic	0.70	0.47	0.90	0.53	0.75
	Chronic	0.75	0.90	0.47	0.81	
	Weighted Average	0.75	0.75	0.62	0.72	
Decision Tree	Non-Chronic	0.63	0.48	0.91	0.51	0.76
	Chronic	0.77	0.91	0.48	0.82	
	Weighted Average	0.73	0.76	0.63	0.72	

Table 5.24: Results for Chronic vs Non-Chronic with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.72	0.45	0.95	0.50	0.77
	Chronic	0.76	0.95	0.45	0.83	
	Weighted Average	0.75	0.77	0.63	0.72	
KNN	Non-Chronic	0.69	0.46	0.88	0.50	0.73
	Chronic	0.74	0.88	0.46	0.79	
	Weighted Average	0.74	0.73	0.61	0.70	
Decision Tree	Non-Chronic	0.64	0.55	0.85	0.57	0.76
	Chronic	0.78	0.85	0.55	0.81	
	Weighted Average	0.74	0.76	0.65	0.74	

Healthy vs COPD vs URTI vs Bronchiectasis vs LRTI vs Pneumonia vs Bronchiolitis

In this section, all the pathologies from this database are used in classification. The experiments using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.25, 5.26, 5.27, 5.28, 5.29, 5.30 and 5.31, respectively.

The best result is 0.25 weighted average f1 score, using SVM and top 75 mRMR features, seen in Table 5.31. The best result is using a feature selection method, however, it is not much better than using all the features. The class that has better results is COPD, achieving 0.43 f1 score. The rest of the classes were difficult to identify, being Bronchiolitis and Bronchiectasis the ones with the worst metrics, in general. With these results it is possible to acknowledge that it is not possible to classify recordings in these different pathologies, using this database, features and classifiers. In conclusion to these experiments, the results are not satisfactory.

Table 5.25: Results with all classes with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.07	0.08	0.93	0.07	0.31
	URTI	0.04	0.08	0.95	0.05	
	COPD	0.36	0.70	0.29	0.42	
	Bronchiectasis	0.02	0.05	0.96	0.03	
	Pneumonia	0.06	0.11	0.89	0.07	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.19	0.31	0.70	0.21	
KNN	Healthy	0.14	0.20	0.80	0.15	0.29
	URTI	0.09	0.09	0.94	0.08	
	COPD	0.36	0.56	0.45	0.40	
	Bronchiectasis	0.05	0.05	0.96	0.05	
	Pneumonia	0.19	0.14	0.87	0.12	
	Bronchiolitis	0.01	0.01	0.99	0.01	
	Weighted Average	0.24	0.29	0.72	0.24	
Decision Tree	Healthy	0.14	0.22	0.80	0.15	0.24
	URTI	0.09	0.10	0.93	0.08	
	COPD	0.35	0.43	0.56	0.34	
	Bronchiectasis	0.03	0.08	0.92	0.03	
	Pneumonia	0.11	0.18	0.81	0.12	
	Bronchiolitis	0.03	0.04	0.98	0.03	
	Weighted Average	0.22	0.24	0.76	0.20	

Table 5.26: Results with all classes with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.07	0.12	0.86	0.09	0.30
	URTI	0.01	0.01	0.99	0.01	
	COPD	0.34	0.74	0.27	0.42	
	Bronchiectasis	0.00	0.00	0.99	0.00	
	Pneumonia	0.06	0.11	0.89	0.07	
	Bronchiolitis	0.00	0.01	1.00	0.01	
	Weighted Average	0.18	0.30	0.70	0.21	
KNN	Healthy	0.15	0.24	0.79	0.16	0.28
	URTI	0.05	0.07	0.94	0.05	
	COPD	0.38	0.59	0.43	0.40	
	Bronchiectasis	0.01	0.01	0.99	0.01	
	Pneumonia	0.15	0.12	0.86	0.09	
	Bronchiolitis	0.01	0.01	0.99	0.01	
	Weighted Average	0.23	0.28	0.72	0.22	
Decision Tree	Healthy	0.14	0.17	0.83	0.12	0.27
	URTI	0.02	0.03	0.97	0.02	
	COPD	0.31	0.59	0.40	0.36	
	Bronchiectasis	0.00	0.00	1.00	0.00	
	Pneumonia	0.10	0.18	0.82	0.12	
	Bronchiolitis	0.03	0.03	0.98	0.03	
	Weighted Average	0.20	0.27	0.73	0.20	

Table 5.27: Results with all classes with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.08	0.09	0.85	0.07	0.25
	URTI	0.10	0.10	0.94	0.08	
	COPD	0.39	0.59	0.46	0.38	
	Bronchiectasis	0.08	0.09	0.96	0.08	
	Pneumonia	0.12	0.14	0.83	0.10	
	Bronchiolitis	0.03	0.03	0.97	0.03	
	Weighted Average	0.23	0.25	0.76	0.20	
KNN	Healthy	0.16	0.18	0.82	0.14	0.27
	URTI	0.09	0.08	0.94	0.08	
	COPD	0.39	0.62	0.46	0.42	
	Bronchiectasis	0.02	0.05	0.97	0.03	
	Pneumonia	0.16	0.15	0.86	0.14	
	Bronchiolitis	0.04	0.03	0.98	0.03	
	Weighted Average	0.24	0.27	0.75	0.22	
Decision Tree	Healthy	0.13	0.14	0.85	0.13	0.25
	URTI	0.11	0.17	0.89	0.12	
	COPD	0.35	0.53	0.52	0.37	
	Bronchiectasis	0.12	0.16	0.94	0.12	
	Pneumonia	0.13	0.12	0.83	0.10	
	Bronchiolitis	0.03	0.02	0.98	0.02	
	Weighted Average	0.22	0.25	0.77	0.21	

Table 5.28: Results with all classes with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.14	0.13	0.89	0.12	0.30
	URTI	0.16	0.17	0.95	0.15	
	COPD	0.36	0.66	0.39	0.40	
	Bronchiectasis	0.11	0.10	0.96	0.09	
	Pneumonia	0.28	0.19	0.90	0.19	
	Bronchiolitis	0.04	0.03	0.96	0.03	
	Weighted Average	0.28	0.30	0.75	0.25	
KNN	Healthy	0.13	0.16	0.83	0.13	0.24
	URTI	0.08	0.10	0.92	0.08	
	COPD	0.33	0.57	0.44	0.36	
	Bronchiectasis	0.04	0.04	0.96	0.03	
	Pneumonia	0.23	0.14	0.88	0.15	
	Bronchiolitis	0.05	0.03	0.99	0.03	
	Weighted Average	0.23	0.24	0.77	0.20	
Decision Tree	Healthy	0.15	0.21	0.83	0.16	0.27
	URTI	0.07	0.12	0.93	0.08	
	COPD	0.33	0.60	0.43	0.37	
	Bronchiectasis	0.03	0.03	0.98	0.03	
	Pneumonia	0.12	0.14	0.87	0.11	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.20	0.27	0.76	0.20	

Table 5.29: Results with all classes with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.11	0.12	0.87	0.09	0.29
	URTI	0.05	0.05	0.97	0.04	
	COPD	0.36	0.65	0.35	0.42	
	Bronchiectasis	0.06	0.02	0.98	0.03	
	Pneumonia	0.08	0.16	0.84	0.09	
	Bronchiolitis	0.03	0.03	0.99	0.03	
	Weighted Average	0.20	0.29	0.71	0.22	
KNN	Healthy	0.22	0.25	0.79	0.19	0.26
	URTI	0.04	0.05	0.94	0.04	
	COPD	0.39	0.51	0.47	0.37	
	Bronchiectasis	0.05	0.05	0.95	0.04	
	Pneumonia	0.13	0.12	0.88	0.11	
	Bronchiolitis	0.04	0.07	0.97	0.05	
	Weighted Average	0.24	0.26	0.75	0.22	
Decision Tree	Healthy	0.10	0.19	0.84	0.12	0.27
	URTI	0.04	0.06	0.92	0.05	
	COPD	0.38	0.56	0.47	0.40	
	Bronchiectasis	0.03	0.04	0.95	0.03	
	Pneumonia	0.10	0.16	0.83	0.11	
	Bronchiolitis	0.03	0.01	0.99	0.01	
	Weighted Average	0.22	0.27	0.74	0.22	

Table 5.30: Results with all classes with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.09	0.05	0.93	0.06	0.30
	URTI	0.04	0.04	0.96	0.04	
	COPD	0.36	0.78	0.25	0.45	
	Bronchiectasis	0.02	0.04	0.98	0.02	
	Pneumonia	0.07	0.06	0.90	0.05	
	Bronchiolitis	0.00	0.01	0.97	0.00	
	Weighted Average	0.18	0.30	0.70	0.20	
KNN	Healthy	0.15	0.16	0.82	0.13	0.28
	URTI	0.09	0.05	0.94	0.06	
	COPD	0.36	0.67	0.37	0.42	
	Bronchiectasis	0.10	0.07	0.96	0.06	
	Pneumonia	0.13	0.07	0.92	0.08	
	Bronchiolitis	0.01	0.02	0.99	0.02	
	Weighted Average	0.22	0.28	0.73	0.21	
Decision Tree	Healthy	0.10	0.13	0.89	0.10	0.30
	URTI	0.13	0.16	0.93	0.12	
	COPD	0.38	0.63	0.43	0.44	
	Bronchiectasis	0.10	0.05	0.97	0.05	
	Pneumonia	0.11	0.18	0.82	0.12	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.21	0.30	0.73	0.22	

Table 5.31: Results with all classes with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.12	0.11	0.90	0.09	0.34
	URTI	0.07	0.06	0.97	0.07	
	COPD	0.40	0.72	0.29	0.46	
	Bronchiectasis	0.01	0.02	0.97	0.02	
	Pneumonia	0.07	0.08	0.91	0.07	
	Bronchiolitis	0.02	0.04	0.97	0.03	
	Weighted Average	0.24	0.34	0.67	0.25	
KNN	Healthy	0.15	0.18	0.81	0.14	0.28
	URTI	0.12	0.11	0.94	0.10	
	COPD	0.39	0.58	0.41	0.42	
	Bronchiectasis	0.01	0.04	0.95	0.02	
	Pneumonia	0.13	0.07	0.89	0.06	
	Bronchiolitis	0.01	0.01	0.99	0.01	
	Weighted Average	0.25	0.28	0.71	0.23	
Decision Tree	Healthy	0.08	0.12	0.85	0.10	0.28
	URTI	0.10	0.16	0.93	0.12	
	COPD	0.40	0.55	0.46	0.41	
	Bronchiectasis	0.05	0.10	0.95	0.05	
	Pneumonia	0.12	0.14	0.83	0.12	
	Bronchiolitis	0.02	0.03	0.99	0.02	
	Weighted Average	0.24	0.28	0.73	0.23	

COPD vs URTI vs Bronchiectasis vs Pneumonia vs Bronchiolitis

As seen before the difficulty in identifying class Healthy, in here experiments done in the previous section are repeated, but without that class. The experiments using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.32, 5.33, 5.34, 5.35, 5.36, 5.37 and 5.38 respectively.

The best results in these experiments are with 0.42 weighted average f1 score, using SVM and top 35 mRMR features. The feature selection methods do not give a much better performance than with all features, giving very similar results. These results are very similar to the ones in the previous section, with COPD being the best class and Bronchiectasis and Bronchiolitis the worst, in general. SVM is the classifier that almost always has the best f1 score, however not by much. In conclusion to these experiments, the results are not satisfactory.

Table 5.32: Results with all classes, except Healthy, with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.02	0.03	0.97	0.02	0.51
	COPD	0.56	0.88	0.13	0.66	
	Bronchiectasis	0.03	0.04	0.98	0.03	
	Pneumonia	0.03	0.06	0.95	0.04	
	Bronchiolitis	0.01	0.02	0.98	0.01	
	Weighted Average	0.35	0.51	0.50	0.40	
KNN	URTI	0.11	0.11	0.93	0.09	0.45
	COPD	0.57	0.77	0.25	0.62	
	Bronchiectasis	0.04	0.07	0.96	0.05	
	Pneumonia	0.10	0.11	0.88	0.08	
	Bronchiolitis	0.02	0.02	0.99	0.02	
	Weighted Average	0.38	0.45	0.57	0.38	
Decision Tree	URTI	0.03	0.06	0.95	0.04	0.43
	COPD	0.54	0.74	0.23	0.58	
	Bronchiectasis	0.02	0.03	0.98	0.02	
	Pneumonia	0.10	0.15	0.84	0.09	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.35	0.43	0.56	0.36	

Table 5.33: Results with all classes, except Healthy, with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.00	0.00	0.98	0.00	0.48
	COPD	0.52	0.90	0.11	0.64	
	Bronchiectasis	0.00	0.00	1.00	0.00	
	Pneumonia	0.05	0.08	0.92	0.05	
	Bronchiolitis	0.01	0.01	1.00	0.01	
	Weighted Average	0.30	0.48	0.52	0.36	
KNN	URTI	0.05	0.08	0.93	0.05	0.45
	COPD	0.52	0.79	0.22	0.60	
	Bronchiectasis	0.01	0.01	0.99	0.01	
	Pneumonia	0.22	0.15	0.88	0.15	
	Bronchiolitis	0.00	0.01	0.99	0.01	
	Weighted Average	0.34	0.45	0.57	0.36	
Decision Tree	URTI	0.05	0.04	0.98	0.03	0.47
	COPD	0.53	0.85	0.16	0.63	
	Bronchiectasis	0.00	0.00	1.00	0.00	
	Pneumonia	0.11	0.13	0.89	0.11	
	Bronchiolitis	0.00	0.01	0.98	0.01	
	Weighted Average	0.32	0.47	0.54	0.37	

Table 5.34: Results with all classes, except Healthy, with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.02	0.03	0.96	0.02	0.47
	COPD	0.53	0.83	0.15	0.62	
	Bronchiectasis	0.06	0.03	0.97	0.03	
	Pneumonia	0.08	0.07	0.92	0.07	
	Bronchiolitis	0.02	0.04	0.98	0.02	
	Weighted Average	0.33	0.47	0.52	0.37	
KNN	URTI	0.05	0.05	0.94	0.03	0.45
	COPD	0.52	0.81	0.17	0.61	
	Bronchiectasis	0.04	0.02	0.97	0.02	
	Pneumonia	0.09	0.05	0.91	0.05	
	Bronchiolitis	0.00	0.01	1.00	0.01	
	Weighted Average	0.32	0.45	0.53	0.36	
Decision Tree	URTI	0.05	0.04	0.96	0.03	0.45
	COPD	0.53	0.79	0.18	0.61	
	Bronchiectasis	0.02	0.03	0.96	0.02	
	Pneumonia	0.13	0.10	0.89	0.08	
	Bronchiolitis	0.01	0.01	0.99	0.01	
	Weighted Average	0.34	0.45	0.54	0.36	

Table 5.35: Results with all classes, except Healthy, with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.03	0.05	0.96	0.04	0.50
	COPD	0.56	0.85	0.14	0.65	
	Bronchiectasis	0.01	0.03	0.98	0.02	
	Pneumonia	0.10	0.06	0.94	0.06	
	Bronchiolitis	0.03	0.03	0.98	0.03	
	Weighted Average	0.36	0.50	0.51	0.39	
KNN	URTI	0.08	0.13	0.93	0.09	0.45
	COPD	0.56	0.76	0.25	0.62	
	Bronchiectasis	0.04	0.05	0.95	0.04	
	Pneumonia	0.13	0.09	0.90	0.10	
	Bronchiolitis	0.02	0.02	0.98	0.02	
	Weighted Average	0.37	0.45	0.55	0.39	
Decision Tree	URTI	0.08	0.14	0.94	0.10	0.44
	COPD	0.55	0.70	0.28	0.60	
	Bronchiectasis	0.09	0.16	0.93	0.10	
	Pneumonia	0.10	0.11	0.88	0.09	
	Bronchiolitis	0.02	0.03	0.97	0.02	
	Weighted Average	0.36	0.44	0.55	0.38	

Table 5.36: Results with all classes, except Healthy, with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.01	0.00	1.00	0.01	0.52
	COPD	0.54	0.93	0.09	0.66	
	Bronchiectasis	0.04	0.03	0.99	0.02	
	Pneumonia	0.03	0.07	0.94	0.04	
	Bronchiolitis	0.05	0.03	0.99	0.03	
	Weighted Average	0.33	0.52	0.48	0.39	
KNN	URTI	0.02	0.05	0.96	0.02	0.47
	COPD	0.54	0.85	0.15	0.63	
	Bronchiectasis	0.04	0.06	0.97	0.03	
	Pneumonia	0.07	0.04	0.93	0.04	
	Bronchiolitis	0.03	0.03	0.99	0.03	
	Weighted Average	0.34	0.47	0.52	0.37	
Decision Tree	URTI	0.02	0.06	0.93	0.03	0.44
	COPD	0.52	0.79	0.19	0.60	
	Bronchiectasis	0.00	0.02	0.96	0.01	
	Pneumonia	0.10	0.10	0.92	0.09	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.33	0.44	0.54	0.35	

Table 5.37: Results with all classes, except Healthy, with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.07	0.02	0.99	0.03	0.55
	COPD	0.57	0.96	0.05	0.71	
	Bronchiectasis	0.01	0.01	1.00	0.01	
	Pneumonia	0.01	0.01	0.98	0.01	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.35	0.55	0.45	0.42	
KNN	URTI	0.09	0.13	0.93	0.10	0.49
	COPD	0.57	0.80	0.19	0.65	
	Bronchiectasis	0.04	0.06	0.97	0.04	
	Pneumonia	0.16	0.09	0.94	0.11	
	Bronchiolitis	0.00	0.00	0.98	0.00	
	Weighted Average	0.38	0.49	0.52	0.41	
Decision Tree	URTI	0.03	0.05	0.96	0.04	0.48
	COPD	0.56	0.79	0.18	0.63	
	Bronchiectasis	0.04	0.08	0.96	0.04	
	Pneumonia	0.08	0.09	0.90	0.07	
	Bronchiolitis	0.01	0.02	0.99	0.01	
	Weighted Average	0.36	0.48	0.51	0.39	

Table 5.38: Results with all classes, except Healthy, with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	URTI	0.01	0.02	0.98	0.02	0.51
	COPD	0.56	0.88	0.09	0.67	
	Bronchiectasis	0.02	0.02	0.98	0.02	
	Pneumonia	0.03	0.04	0.94	0.03	
	Bronchiolitis	0.03	0.02	0.99	0.02	
	Weighted Average	0.34	0.51	0.47	0.40	
KNN	URTI	0.12	0.14	0.91	0.12	0.40
	COPD	0.55	0.64	0.34	0.57	
	Bronchiectasis	0.10	0.12	0.94	0.09	
	Pneumonia	0.16	0.15	0.83	0.14	
	Bronchiolitis	0.00	0.01	0.98	0.01	
	Weighted Average	0.38	0.40	0.59	0.37	
Decision Tree	URTI	0.08	0.11	0.93	0.08	0.46
	COPD	0.56	0.73	0.25	0.61	
	Bronchiectasis	0.03	0.07	0.97	0.04	
	Pneumonia	0.10	0.14	0.85	0.10	
	Bronchiolitis	0.00	0.00	0.99	0.00	
	Weighted Average	0.36	0.46	0.54	0.39	

Comparing the results obtained for this database with the results from the State of the Art that also used this database, it is easily seen that the ones obtained here have a worst performance. However, those studies use more complicated methods, like data augmentation and deep learning. For those studies it is expected that they have better results, like in [22] and [21].

In general, with this database is difficult for the classifiers to identify the class Healthy. This might be because there are few recordings from healthy subjects, almost all of the healthy subjects only have one recording, making it more difficult for the classifiers to learn how to identify healthy recordings. Another possible cause for this is that almost all of the healthy subjects are children, from the 26 healthy subjects, 22 are children that go from 2 to 16 years old. This might influence the classifier's learning process since children's healthy respiratory sounds are different from adults' healthy respiratory sounds.

The best results with this database are with Chronic vs Non-Chronic, which follows what is said before, seen that classifiers are having trouble classifying the Healthy class, and in this case it is left out.

Classifying using the different pathologies was not successful, probably because of the low number of patients and recordings that some classes have, like Bronchiolitis and Bronchiectasis. However, the COPD class was the one with better results in every experiment. Even when only three classes are being used, class Chronic, which includes COPD, has the best results. This is a consequence of existing more COPD recordings, even after undersampling, because COPD patients have more recordings than patients from other classes. However, in [19] COPD class is also the one that, in general, has better results. This can mean that COPD recordings have distinctive properties which are easily learned by classifiers.

Another possible cause for these results not being better might be the features used here not being the best for the purpose of diagnosing pathologies.

5.2 Thessaloniki Database

In this part, the same experiments done with ICBHI database are be repeated to the Thessaloniki database.

Healthy vs Sick

For this experiments, all of the classes that were different from the class Healthy, were converted to the class Sick. First starting by using all of the features, and then using a smaller set of features. The results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.39, 5.40, 5.41, 5.42, 5.43, 5.44 and 5.45, respectively.

The best results for this experiment are 0.64 weighted average f1 score, which is achieved in five experiments. All of the seven experiments have very similar results, with feature selection having no influence. The type of classifier used also did not influence the results. The Sick class was the better, reaching 0.85 f1 score, however, the Healthy class was very difficult to be identified by the classifiers. Almost all of the recordings are being classified as Sick, which might be caused by the number of recordings from the Sick class being superior to the number of Healthy recordings, increasing the accuracy and weighted average f1 score.

Table 5.39: Results with Healthy vs Sick with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.06	0.04	0.94	0.05	0.71
	Sick	0.75	0.94	0.04	0.83	
	Weighted Average	0.57	0.71	0.27	0.63	
KNN	Healthy	0.12	0.07	0.90	0.07	0.69
	Sick	0.74	0.90	0.07	0.81	
	Weighted Average	0.59	0.69	0.28	0.63	
Decision Tree	Healthy	0.06	0.05	0.95	0.05	0.72
	Sick	0.75	0.95	0.05	0.83	
	Weighted Average	0.58	0.72	0.27	0.64	

With these results, it is possible to see that using this database and these methods, it is not possible to distinguish a recording from class Healthy from one of class Sick. In conclusion to these experiments, the results are not satisfactory.

Table 5.40: Results with Healthy vs Sick with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.00	0.00	0.99	0.00	0.74
	Sick	0.75	0.99	0.00	0.85	
	Weighted Average	0.56	0.74	0.25	0.64	
KNN	Healthy	0.04	0.01	0.97	0.02	0.73
	Sick	0.74	0.97	0.01	0.84	
	Weighted Average	0.57	0.73	0.26	0.63	
Decision Tree	Healthy	0.04	0.02	0.97	0.02	0.73
	Sick	0.74	0.97	0.02	0.84	
	Weighted Average	0.57	0.73	0.26	0.63	

Table 5.41: Results with Healthy vs Sick with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.00	0.99	0.00	0.74
	Sick	0.75	0.99	0.00	0.85	
	Weighted Average	0.56	0.74	0.25	0.64	
KNN	Healthy	0.18	0.07	0.91	0.08	0.70
	Sick	0.74	0.91	0.07	0.81	
	Weighted Average	0.60	0.70	0.28	0.63	
Decision Tree	Healthy	0.07	0.05	0.93	0.06	0.71
	Sick	0.74	0.93	0.05	0.82	
	Weighted Average	0.57	0.71	0.27	0.63	

Table 5.42: Results with Healthy vs Sick with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.03	0.01	0.98	0.01	0.73
	Sick	0.75	0.98	0.01	0.85	
	Weighted Average	0.57	0.73	0.25	0.64	
KNN	Healthy	0.18	0.10	0.90	0.11	0.69
	Sick	0.75	0.90	0.10	0.81	
	Weighted Average	0.61	0.69	0.30	0.63	
Decision Tree	Healthy	0.08	0.05	0.94	0.06	0.72
	Sick	0.75	0.94	0.05	0.83	
	Weighted Average	0.58	0.72	0.27	0.64	

Table 5.43: Results with Healthy vs Sick with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.00	0.00	1.00	0.00	0.74
	Sick	0.74	1.00	0.00	0.85	
	Weighted Average	0.55	0.74	0.26	0.63	
KNN	Healthy	0.08	0.07	0.92	0.06	0.70
	Sick	0.72	0.92	0.07	0.80	
	Weighted Average	0.56	0.70	0.29	0.61	
Decision Tree	Healthy	0.05	0.03	0.97	0.03	0.73
	Sick	0.74	0.97	0.03	0.84	
	Weighted Average	0.56	0.73	0.27	0.63	

Table 5.44: Results with Healthy vs Sick with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.00	0.99	0.00	0.74
	Sick	0.74	0.99	0.00	0.85	
	Weighted Average	0.56	0.74	0.25	0.63	
KNN	Healthy	0.12	0.09	0.90	0.09	0.69
	Sick	0.74	0.90	0.09	0.81	
	Weighted Average	0.58	0.69	0.30	0.63	
Decision Tree	Healthy	0.06	0.04	0.94	0.04	0.71
	Sick	0.74	0.94	0.04	0.82	
	Weighted Average	0.57	0.71	0.27	0.62	

Table 5.45: Results with Healthy vs Sick with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.01	0.00	0.99	0.00	0.74
	Sick	0.75	0.99	0.00	0.85	
	Weighted Average	0.57	0.74	0.25	0.64	
KNN	Healthy	0.16	0.13	0.84	0.13	0.66
	Sick	0.74	0.84	0.13	0.78	
	Weighted Average	0.60	0.66	0.31	0.62	
Decision Tree	Healthy	0.05	0.02	0.95	0.02	0.72
	Sick	0.74	0.95	0.02	0.83	
	Weighted Average	0.57	0.72	0.25	0.63	

Table 5.46: Results with all classes with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.22	0.32	0.63	0.25	0.23
	Non-Chronic	0.20	0.22	0.73	0.21	
	Chronic	0.25	0.17	0.85	0.20	
	Interstitial	0.23	0.23	0.77	0.22	
	Weighted Average	0.23	0.23	0.75	0.22	
KNN	Healthy	0.22	0.35	0.60	0.27	0.23
	Non-Chronic	0.21	0.26	0.67	0.23	
	Chronic	0.24	0.14	0.87	0.17	
	Interstitial	0.24	0.18	0.83	0.20	
	Weighted Average	0.23	0.23	0.75	0.22	
Decision Tree	Healthy	0.23	0.32	0.62	0.26	0.23
	Non-Chronic	0.23	0.24	0.74	0.22	
	Chronic	0.21	0.16	0.83	0.17	
	Interstitial	0.23	0.21	0.78	0.21	
	Weighted Average	0.23	0.23	0.74	0.22	

Healthy vs Non-Chronic vs Chronic vs Interstitial

In this section, the classification is made using 4 classes: Healthy, Non-Chronic, Chronic and Interstitial. Like in the previous section, the results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.46, 5.47, 5.48, 5.49, 5.50, 5.51 and 5.52, respectively.

Using KNN and top 75 ReliefF features, it was possible to get 0.28 weighted average f1 score. The use of feature selection methods was not influential here. From the four classes, the class Healthy is the one that got better results, reaching 0.36 f1 score. KNN is the classifier that gets the better results, in general. With these results it is possible to see that with these methods, it is not possible to classify recordings from this database in these four classes. In conclusion to these experiments, the results are not satisfactory.

Table 5.47: Results with all classes with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.24	0.50	0.46	0.32	0.26
	Non-Chronic	0.29	0.26	0.80	0.25	
	Chronic	0.26	0.11	0.92	0.15	
	Interstitial	0.26	0.19	0.84	0.21	
	Weighted Average	0.27	0.26	0.76	0.23	
KNN	Healthy	0.25	0.40	0.58	0.30	0.27
	Non-Chronic	0.26	0.29	0.74	0.26	
	Chronic	0.34	0.21	0.86	0.25	
	Interstitial	0.26	0.18	0.84	0.20	
	Weighted Average	0.28	0.27	0.76	0.25	
Decision Tree	Healthy	0.25	0.31	0.66	0.26	0.25
	Non-Chronic	0.24	0.30	0.73	0.25	
	Chronic	0.26	0.21	0.84	0.23	
	Interstitial	0.28	0.21	0.79	0.22	
	Weighted Average	0.26	0.25	0.76	0.24	

Table 5.48: Results with all classes with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.23	0.46	0.47	0.30	0.23
	Non-Chronic	0.21	0.20	0.76	0.19	
	Chronic	0.23	0.09	0.91	0.12	
	Interstitial	0.21	0.16	0.83	0.17	
	Weighted Average	0.22	0.23	0.74	0.20	
KNN	Healthy	0.27	0.42	0.58	0.32	0.28
	Non-Chronic	0.27	0.26	0.76	0.26	
	Chronic	0.38	0.24	0.87	0.28	
	Interstitial	0.24	0.19	0.82	0.20	
	Weighted Average	0.29	0.28	0.76	0.27	
Decision Tree	Healthy	0.24	0.35	0.63	0.28	0.26
	Non-Chronic	0.27	0.26	0.75	0.24	
	Chronic	0.34	0.22	0.84	0.25	
	Interstitial	0.26	0.21	0.80	0.22	
	Weighted Average	0.28	0.26	0.76	0.25	

Table 5.49: Results with all classes with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.21	0.35	0.59	0.26	0.23
	Non-Chronic	0.23	0.21	0.76	0.21	
	Chronic	0.21	0.15	0.84	0.17	
	Interstitial	0.23	0.21	0.79	0.21	
	Weighted Average	0.23	0.23	0.75	0.21	
KNN	Healthy	0.27	0.37	0.65	0.30	0.28
	Non-Chronic	0.29	0.33	0.71	0.30	
	Chronic	0.30	0.21	0.85	0.24	
	Interstitial	0.31	0.24	0.84	0.25	
	Weighted Average	0.29	0.28	0.76	0.28	
Decision Tree	Healthy	0.22	0.31	0.64	0.25	0.25
	Non-Chronic	0.25	0.28	0.74	0.25	
	Chronic	0.29	0.19	0.85	0.22	
	Interstitial	0.26	0.25	0.78	0.24	
	Weighted Average	0.26	0.25	0.76	0.24	

Table 5.50: Results with all classes with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.24	0.80	0.18	0.36	0.24
	Non-Chronic	0.07	0.11	0.88	0.07	
	Chronic	0.10	0.03	0.98	0.05	
	Interstitial	0.13	0.05	0.96	0.06	
	Weighted Average	0.13	0.24	0.76	0.13	
KNN	Healthy	0.24	0.43	0.54	0.29	0.24
	Non-Chronic	0.21	0.23	0.71	0.20	
	Chronic	0.29	0.16	0.88	0.19	
	Interstitial	0.29	0.17	0.86	0.21	
	Weighted Average	0.26	0.24	0.75	0.23	
Decision Tree	Healthy	0.26	0.46	0.55	0.30	0.25
	Non-Chronic	0.16	0.20	0.76	0.16	
	Chronic	0.22	0.14	0.87	0.16	
	Interstitial	0.23	0.19	0.81	0.20	
	Weighted Average	0.22	0.25	0.75	0.21	

Table 5.51: Results with all classes with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.22	0.55	0.36	0.29	0.22
	Non-Chronic	0.11	0.14	0.81	0.11	
	Chronic	0.17	0.08	0.93	0.10	
	Interstitial	0.15	0.13	0.86	0.13	
	Weighted Average	0.16	0.22	0.75	0.16	
KNN	Healthy	0.24	0.36	0.60	0.28	0.24
	Non-Chronic	0.23	0.29	0.69	0.25	
	Chronic	0.26	0.18	0.85	0.20	
	Interstitial	0.25	0.17	0.85	0.19	
	Weighted Average	0.25	0.24	0.75	0.23	
Decision Tree	Healthy	0.22	0.34	0.61	0.25	0.24
	Non-Chronic	0.22	0.23	0.75	0.21	
	Chronic	0.31	0.21	0.82	0.24	
	Interstitial	0.22	0.20	0.81	0.20	
	Weighted Average	0.25	0.24	0.75	0.23	

Table 5.52: Results with all classes with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.21	0.42	0.51	0.26	0.23
	Non-Chronic	0.22	0.23	0.78	0.22	
	Chronic	0.23	0.14	0.86	0.17	
	Interstitial	0.18	0.15	0.82	0.16	
	Weighted Average	0.21	0.23	0.75	0.20	
KNN	Healthy	0.24	0.36	0.62	0.29	0.27
	Non-Chronic	0.28	0.31	0.73	0.29	
	Chronic	0.31	0.19	0.84	0.22	
	Interstitial	0.29	0.20	0.84	0.22	
	Weighted Average	0.28	0.27	0.76	0.26	
Decision Tree	Healthy	0.23	0.36	0.59	0.27	0.23
	Non-Chronic	0.18	0.21	0.76	0.19	
	Chronic	0.23	0.16	0.85	0.18	
	Interstitial	0.26	0.22	0.78	0.22	
	Weighted Average	0.23	0.23	0.75	0.21	

Non-Chronic vs Chronic vs Interstitial

In this section, the classification is made like in the previous section, but without class Healthy. Like in the previous section, the results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.53, 5.54, 5.55, 5.56, 5.57, 5.58 and 5.59, respectively.

The best result is achieved with 0.39 weighted average f1 score, with SVM classifier and top 10 ReliefF features. With the feature selection methods is possible to get better results than with all features. The best class is Non-Chronic, reaching 0.43 f1 score, and Chronic being the worst, in general. None of the different classifiers stands out from the others in terms of performance. In conclusion to these experiments, the results are not satisfactory.

Table 5.53: Results with Non-Chronic vs Chronic vs Interstitial with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.32	0.36	0.58	0.33	0.31
	Chronic	0.32	0.24	0.74	0.27	
	Interstitial	0.32	0.35	0.65	0.33	
	Weighted Average	0.32	0.31	0.66	0.31	
KNN	Non-Chronic	0.34	0.48	0.50	0.39	0.33
	Chronic	0.31	0.24	0.76	0.26	
	Interstitial	0.32	0.27	0.74	0.28	
	Weighted Average	0.33	0.33	0.66	0.31	
Decision Tree	Non-Chronic	0.36	0.39	0.63	0.36	0.34
	Chronic	0.35	0.31	0.70	0.32	
	Interstitial	0.30	0.31	0.68	0.30	
	Weighted Average	0.34	0.34	0.67	0.33	

Table 5.54: Results with Non-Chronic vs Chronic vs Interstitial with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.44	0.45	0.67	0.43	0.39
	Chronic	0.44	0.33	0.77	0.36	
	Interstitial	0.35	0.42	0.65	0.37	
	Weighted Average	0.42	0.39	0.71	0.39	
KNN	Non-Chronic	0.38	0.49	0.57	0.42	0.38
	Chronic	0.42	0.32	0.77	0.34	
	Interstitial	0.36	0.33	0.73	0.32	
	Weighted Average	0.40	0.38	0.69	0.36	
Decision Tree	Non-Chronic	0.33	0.34	0.65	0.33	0.33
	Chronic	0.34	0.28	0.73	0.29	
	Interstitial	0.32	0.38	0.62	0.34	
	Weighted Average	0.34	0.33	0.67	0.32	

Table 5.55: Results with Non-Chronic vs Chronic vs Interstitial with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.35	0.38	0.63	0.33	0.34
	Chronic	0.32	0.22	0.81	0.25	
	Interstitial	0.33	0.45	0.59	0.37	
	Weighted Average	0.34	0.34	0.69	0.32	
KNN	Non-Chronic	0.39	0.54	0.55	0.43	0.38
	Chronic	0.42	0.29	0.81	0.33	
	Interstitial	0.37	0.33	0.72	0.32	
	Weighted Average	0.40	0.38	0.70	0.36	
Decision Tree	Non-Chronic	0.35	0.38	0.66	0.35	0.37
	Chronic	0.43	0.33	0.79	0.36	
	Interstitial	0.34	0.43	0.62	0.36	
	Weighted Average	0.38	0.37	0.70	0.35	

Table 5.56: Results with Non-Chronic vs Chronic vs Interstitial with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.34	0.34	0.61	0.31	0.31
	Chronic	0.32	0.25	0.74	0.27	
	Interstitial	0.32	0.38	0.64	0.33	
	Weighted Average	0.34	0.31	0.67	0.30	
KNN	Non-Chronic	0.40	0.48	0.59	0.42	0.37
	Chronic	0.41	0.30	0.78	0.33	
	Interstitial	0.31	0.33	0.69	0.31	
	Weighted Average	0.38	0.37	0.69	0.36	
Decision Tree	Non-Chronic	0.39	0.37	0.65	0.37	0.35
	Chronic	0.39	0.33	0.76	0.34	
	Interstitial	0.30	0.35	0.62	0.31	
	Weighted Average	0.37	0.35	0.68	0.35	

Table 5.57: Results with Non-Chronic vs Chronic vs Interstitial with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.32	0.55	0.42	0.36	0.32
	Chronic	0.21	0.11	0.89	0.14	
	Interstitial	0.26	0.32	0.67	0.24	
	Weighted Average	0.26	0.32	0.67	0.24	
KNN	Non-Chronic	0.37	0.49	0.50	0.40	0.33
	Chronic	0.27	0.18	0.79	0.21	
	Interstitial	0.31	0.31	0.70	0.29	
	Weighted Average	0.32	0.33	0.66	0.31	
Decision Tree	Non-Chronic	0.29	0.28	0.69	0.28	0.32
	Chronic	0.37	0.33	0.72	0.33	
	Interstitial	0.29	0.38	0.59	0.31	
	Weighted Average	0.32	0.32	0.67	0.31	

Table 5.58: Results with Non-Chronic vs Chronic vs Interstitial with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.31	0.49	0.45	0.35	0.32
	Chronic	0.30	0.21	0.79	0.22	
	Interstitial	0.30	0.30	0.75	0.28	
	Weighted Average	0.31	0.32	0.67	0.29	
KNN	Non-Chronic	0.41	0.51	0.59	0.44	0.38
	Chronic	0.40	0.28	0.77	0.32	
	Interstitial	0.36	0.35	0.71	0.33	
	Weighted Average	0.39	0.38	0.69	0.37	
Decision Tree	Non-Chronic	0.33	0.32	0.65	0.30	0.33
	Chronic	0.34	0.27	0.76	0.29	
	Interstitial	0.32	0.42	0.60	0.35	
	Weighted Average	0.34	0.33	0.67	0.31	

Table 5.59: Results with Non-Chronic vs Chronic vs Interstitial with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.30	0.40	0.58	0.33	0.33
	Chronic	0.34	0.23	0.78	0.27	
	Interstitial	0.35	0.40	0.65	0.35	
	Weighted Average	0.33	0.33	0.68	0.31	
KNN	Non-Chronic	0.36	0.46	0.58	0.40	0.36
	Chronic	0.38	0.27	0.78	0.30	
	Interstitial	0.37	0.37	0.69	0.36	
	Weighted Average	0.38	0.36	0.69	0.35	
Decision Tree	Non-Chronic	0.32	0.35	0.63	0.32	0.33
	Chronic	0.33	0.28	0.75	0.30	
	Interstitial	0.34	0.39	0.63	0.34	
	Weighted Average	0.33	0.33	0.68	0.32	

5.3 Mixed Databases

In this section a mix of the databases is used, depending on what is the classification.

Healthy vs Sick

For this experiments the three databases are used. All of the classes that are different from the class Healthy, are converted to the class Sick. First starting by using all of the features, and then using a smaller set of features. The results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.60, 5.61, 5.62, 5.63, 5.64, 5.65 and 5.66, respectively.

The best result is with 0.8 in various experiments. The feature selection methods give almost the same results as with all the features, not being very influential in these experiments. The classifier also is not very influential, giving the three classifiers similar results. Class Sick has the best results, reaching 0.85 f1 score, with Healthy class only reaching 0.74.

The results in this experiment give much higher results than using only one database, where normally the classifiers can not identify the Healthy class. Because this seemed a little strange, some analysis is made trying to understand how this happened. Looking at every correct prediction and from which database it came from, it is possible to see the accuracy from each class, from each database, and the accuracy from all the databases. Table 5.66 shows those accuracies for one of the best results, SVM with top 35 ReliefF features. Analyzing that table is possible to see that the classifiers still can not identify the Healthy recordings from ICBHI and Thessaloniki databases, predicting almost all of the recordings from those databases as Sick. However, with the WELCOME database it happens the opposite, it predicts well the Healthy recordings, but it has more difficulty predicting Sick recordings. Although the WELCOME database has very few recordings from the Sick class, only 8% of the recordings, which might be the cause of not being able to identify the Sick class.

Looking at just the results from Tables 5.60-5.66 it would be possible to conclude that, with these methods, these are very satisfactory results. However seeing that the classifiers are not predicting uniformly in all databases, these results can not be valid.

Table 5.60: Results with Healthy vs Sick with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.82	0.65	0.88	0.72	0.79
	Sick	0.77	0.88	0.65	0.82	
	Weighted Average	0.79	0.79	0.75	0.78	
KNN	Healthy	0.84	0.64	0.90	0.72	0.79
	Sick	0.77	0.90	0.64	0.83	
	Weighted Average	0.80	0.79	0.75	0.78	
Decision Tree	Healthy	0.89	0.60	0.94	0.72	0.80
	Sick	0.76	0.94	0.60	0.84	
	Weighted Average	0.82	0.80	0.75	0.79	

Table 5.61: Results with Healthy vs Sick with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.77	0.63	0.85	0.69	0.76
	Sick	0.75	0.85	0.63	0.80	
	Weighted Average	0.76	0.76	0.73	0.75	
KNN	Healthy	0.84	0.62	0.90	0.71	0.78
	Sick	0.76	0.90	0.62	0.82	
	Weighted Average	0.79	0.78	0.74	0.77	
Decision Tree	Healthy	0.91	0.60	0.95	0.72	0.80
	Sick	0.75	0.95	0.60	0.84	
	Weighted Average	0.82	0.80	0.75	0.79	

Table 5.62: Results with Healthy vs Sick with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.92	0.61	0.96	0.73	0.81
	Sick	0.76	0.96	0.61	0.85	
	Weighted Average	0.83	0.81	0.76	0.80	
KNN	Healthy	0.85	0.61	0.92	0.71	0.78
	Sick	0.76	0.92	0.61	0.83	
	Weighted Average	0.80	0.78	0.75	0.78	
Decision Tree	Healthy	0.90	0.62	0.95	0.74	0.81
	Sick	0.77	0.95	0.62	0.85	
	Weighted Average	0.83	0.81	0.76	0.80	

Table 5.63: Results with Healthy vs Sick with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.91	0.62	0.95	0.73	0.81
	Sick	0.76	0.95	0.62	0.85	
	Weighted Average	0.83	0.81	0.76	0.80	
KNN	Healthy	0.83	0.64	0.89	0.72	0.78
	Sick	0.76	0.89	0.64	0.82	
	Weighted Average	0.79	0.78	0.75	0.78	
Decision Tree	Healthy	0.87	0.63	0.92	0.73	0.79
	Sick	0.76	0.92	0.63	0.83	
	Weighted Average	0.81	0.79	0.75	0.79	

Table 5.64: Results with Healthy vs Sick with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.90	0.63	0.95	0.74	0.81
	Sick	0.77	0.95	0.63	0.85	
	Weighted Average	0.83	0.81	0.77	0.80	
KNN	Healthy	0.82	0.64	0.88	0.72	0.78
	Sick	0.77	0.88	0.64	0.82	
	Weighted Average	0.79	0.78	0.74	0.78	
Decision Tree	Healthy	0.88	0.62	0.93	0.72	0.80
	Sick	0.77	0.93	0.62	0.84	
	Weighted Average	0.81	0.80	0.75	0.79	

Table 5.65: Results with Healthy vs Sick with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.91	0.61	0.96	0.73	0.81
	Sick	0.77	0.96	0.61	0.85	
	Weighted Average	0.83	0.81	0.76	0.80	
KNN	Healthy	0.86	0.63	0.91	0.72	0.79
	Sick	0.77	0.91	0.63	0.83	
	Weighted Average	0.81	0.79	0.75	0.79	
Decision Tree	Healthy	0.85	0.61	0.91	0.71	0.79
	Sick	0.76	0.91	0.61	0.83	
	Weighted Average	0.80	0.79	0.74	0.78	

Table 5.66: Results with Healthy vs Sick with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Healthy	0.90	0.62	0.95	0.74	0.81
	Sick	0.77	0.95	0.62	0.85	
	Weighted Average	0.83	0.81	0.76	0.80	
KNN	Healthy	0.86	0.63	0.91	0.72	0.79
	Sick	0.76	0.91	0.63	0.83	
	Weighted Average	0.80	0.79	0.75	0.78	
Decision Tree	Healthy	0.89	0.61	0.94	0.72	0.80
	Sick	0.76	0.94	0.61	0.84	
	Weighted Average	0.82	0.80	0.75	0.79	

Table 5.67: Accuracy for each class from each database, and for all the database

Thessaloniki			ICBHI			WELCOME		
Healthy	Sick	Total	Healthy	Sick	Total	Healthy	Sick	Total
0.09	0.91	0.66	0.00	0.99	0.95	0.99	0.28	0.93

Non-Chronic vs Chronic

In this experiment, the recordings from ICBHI and Thessaloniki databases, that are part of the classes No-Chronic and Chronic are used. First starting by using all of the features, and then using a smaller set of features. The results using all features, top 10 ReliefF, top 35 ReliefF, top 75 ReliefF, top 10 mRMR, top 35 mRMR, top 75 mRMR are presented in Tables 5.68, 5.69, 5.70, 5.71, 5.72, 5.73 and 5.74, respectively.

The best results are achieved with SVM and top 10 ReliefF features, getting 0.52 weighted average f1 score. The different classifiers did not influence much the classification, as so the feature selection methods. Also, none of the classes had a better performance than the other, having similar results. In conclusion to this experiment, the results are not satisfactory.

Table 5.68: Results with Chronic vs Non-Chronic with all features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.43	0.50	0.49	0.46	0.49
	Chronic	0.52	0.49	0.50	0.49	
	Weighted Average	0.49	0.49	0.50	0.48	
KNN	Non-Chronic	0.46	0.58	0.39	0.50	0.48
	Chronic	0.50	0.39	0.58	0.43	
	Weighted Average	0.49	0.48	0.48	0.47	
Decision Tree	Non-Chronic	0.49	0.51	0.52	0.49	0.52
	Chronic	0.54	0.52	0.51	0.52	
	Weighted Average	0.53	0.52	0.51	0.52	

Table 5.69: Results with Chronic vs Non-Chronic with top 10 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.52	0.59	0.48	0.54	0.53
	Chronic	0.54	0.48	0.59	0.49	
	Weighted Average	0.54	0.53	0.54	0.52	
KNN	Non-Chronic	0.52	0.60	0.45	0.54	0.52
	Chronic	0.51	0.45	0.60	0.46	
	Weighted Average	0.53	0.52	0.53	0.51	
Decision Tree	Non-Chronic	0.50	0.54	0.47	0.51	0.50
	Chronic	0.51	0.47	0.54	0.47	
	Weighted Average	0.51	0.50	0.51	0.49	

Table 5.70: Results with Chronic vs Non-Chronic with top 35 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.45	0.75	0.25	0.55	0.47
	Chronic	0.49	0.25	0.75	0.31	
	Weighted Average	0.48	0.47	0.53	0.42	
KNN	Non-Chronic	0.46	0.67	0.36	0.53	0.49
	Chronic	0.54	0.36	0.67	0.41	
	Weighted Average	0.51	0.49	0.53	0.47	
Decision Tree	Non-Chronic	0.44	0.53	0.46	0.47	0.49
	Chronic	0.55	0.46	0.53	0.49	
	Weighted Average	0.51	0.49	0.50	0.49	

Table 5.71: Results with Chronic vs Non-Chronic with top 75 ReliefF features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.46	0.69	0.27	0.53	0.47
	Chronic	0.35	0.27	0.69	0.29	
	Weighted Average	0.41	0.47	0.50	0.41	
KNN	Non-Chronic	0.48	0.54	0.47	0.50	0.50
	Chronic	0.53	0.47	0.54	0.49	
	Weighted Average	0.51	0.50	0.50	0.50	
Decision Tree	Non-Chronic	0.46	0.50	0.48	0.48	0.49
	Chronic	0.51	0.48	0.50	0.49	
	Weighted Average	0.50	0.49	0.49	0.49	

Table 5.72: Results with Chronic vs Non-Chronic with top 10 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.46	0.64	0.36	0.50	0.47
	Chronic	0.46	0.36	0.64	0.37	
	Weighted Average	0.47	0.47	0.53	0.43	
KNN	Non-Chronic	0.46	0.65	0.37	0.51	0.49
	Chronic	0.54	0.37	0.65	0.40	
	Weighted Average	0.52	0.49	0.52	0.46	
Decision Tree	Non-Chronic	0.48	0.53	0.51	0.49	0.51
	Chronic	0.56	0.51	0.53	0.53	
	Weighted Average	0.54	0.51	0.53	0.51	

Table 5.73: Results with Chronic vs Non-Chronic with top 35 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.47	0.68	0.30	0.53	0.47
	Chronic	0.36	0.30	0.68	0.32	
	Weighted Average	0.41	0.47	0.51	0.42	
KNN	Non-Chronic	0.47	0.65	0.34	0.54	0.49
	Chronic	0.49	0.34	0.65	0.40	
	Weighted Average	0.49	0.49	0.50	0.47	
Decision Tree	Non-Chronic	0.48	0.57	0.44	0.52	0.50
	Chronic	0.53	0.44	0.57	0.47	
	Weighted Average	0.51	0.50	0.51	0.50	

Table 5.74: Results with Chronic vs Non-Chronic with top 75 mRMR features

Classifier	Class	Precision	Recall	Specificity	F1 Score	Accuracy
SVM	Non-Chronic	0.44	0.48	0.49	0.41	0.47
	Chronic	0.41	0.49	0.48	0.42	
	Weighted Average	0.43	0.47	0.50	0.41	
KNN	Non-Chronic	0.49	0.59	0.38	0.52	0.48
	Chronic	0.49	0.38	0.59	0.41	
	Weighted Average	0.49	0.48	0.49	0.47	
Decision Tree	Non-Chronic	0.50	0.49	0.51	0.48	0.49
	Chronic	0.50	0.51	0.49	0.50	
	Weighted Average	0.51	0.49	0.51	0.49	

Final Discussion

Looking at all the experiments, from the three classifiers, SVM is the one that most of the experiments got the better results and decision tree was never the best in any experiment. The feature selection method that got most of the best results is ReliefF, however, most of the experiments it did not have a much better performance than using all features, having similar results.

When looking at the results with ICBHI, the classifiers can not identify the Healthy class. This could be because of the few recordings that exist of that class, creating an unbalancing problem, or that most of the healthy patients are children and the features used are not suitable to use with children recordings. With the Thessaloniki database, the classifiers are also not able to identify the Healthy class in Healthy vs Sick experiments, probably also due to unbalanced classes, however, in Healthy vs Non-Chronic vs Chronic vs Interstitial, the Healthy class is the one with the best results and here the classes are more balanced. This could suggest that the problem for identifying the Healthy class is the fact that there are fewer of those recordings.

Other possible causes for the classifiers not being able to identify Healthy class could be the classifiers themselves and the features. Looking at other studies with the same purpose as this, like [15], it was possible to achieve good performance, using similar classifiers as the ones used in this dissertation, like 98.26% with KNN and 92.19% with SVM. In [17] and [18], a SVM and a KNN, respectively, were also used and good performances were achieved. This might lead to conclude that the problem is not with the classifiers, but could be with

the features used.

Besides the low performance with the Healthy class, the other experiments are also not successful, with the only satisfactory results being with the ICBHI database, classifying Chronic vs Non-Chronic. This is a sign that these features might not be adequate for this type of classification problem.

6 Conclusion and Future Work

This dissertation had the objective to access the possibility to perform a differential diagnosis of respiratory pathologies, using respiratory sounds. It is possible to conclude that using these specific materials and methods, it is not possible to perform differential diagnosis of respiratory pathologies, using respiratory sounds.

Next are some propositions for possible future work:

- Use of a different set of features;
- Use of bigger databases, with less unbalancing problems between classes;
- With the use of a bigger database, some deep learning techniques can be implemented;
- Investigate if for children and adults the same methods and features should be used;
- Analyse if other parameters might influence classification, like local of respiratory sound acquisition and type of equipment used.

7 Bibliography

[1] World health organization global health estimates 2020. `https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death`. Accessed: 2021-10-15.

[2] Bruno Machado Rocha, Diogo Pessoa, Alda Marques, Paulo Carvalho, and Rui Pedro Paiva. Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem? *Sensors*, 21(1):57, December 2020.

[3] Alda Marques, Ana Oliveira, and Cristina Jácome. Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: A systematic review. *Respiratory Care*, 59(5):765–776, 2014.

[4] Anssi Sovijärvi, F. Dalmasso, J. Vanderschoot, Leo Malmberg, G. Righini, and S.A.T. Stoneman. Definition of terms for applications of respiratory sounds. *Eur. Respir. Rev.*, 10:597–610, 2000.

[5] Nandini Sengupta, Md Sahidullah, and Goutam Saha. Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, 75:118–129, 2016.

[6] Renard Xaviero Adhi Pramono, Stuart Bowyer, and Esther Rodriguez-Villegas. Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*, 12(5):1–43, 05 2017.

[7] Malay Sarkar, Irappa Madabhavi, Narasimhalu Niranjan, and Megha Dogra. Auscultation of the respiratory system. *Annals of Thoracic Medicine*, 10(3):158–168, September 2015.

[8] Yukio Nagasaka. Lung sounds in bronchial asthma. *Allergology International*, 61(3):353–363, 2012.

[9] Abraham Bohadana, Gabriel Izbicki, and Steve S. Kraman. Fundamentals of lung auscultation. *New England Journal of Medicine*, 370(8):744–751, 2014. PMID: 24552321.

[10] Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès. Analysis of respiratory sounds: State of the art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2:CCRPM.S530, 2008. PMID: 21157521.

[11] Sezer Ulukaya, Ipek Sen, and Yasemin P. Kahya. Feature extraction using time-frequency analysis for monophonic-polyphonic wheeze discrimination. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5412–5415, Milan, August 2015. IEEE.

[12] Mete Yeginer and Yasemin P. Kahya. Elimination of vesicular sounds from pulmonary crackle waveforms. *Computer Methods and Programs in Biomedicine*, 89(1):1–13, January 2008.

[13] Arati Gurung, Carolyn G. Scrafford, James M. Tielsch, Orin S. Levine, and William Checkley. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis. *Respiratory Medicine*, 105(9):1396–1403, September 2011.

[14] Laura Vannuccini, John Earis, Panu Helistö, Barry Cheetham, M. Rossi, Anssi Sovijärvi, and J. Vanderschoot. Capturing and preprocessing of respiratory sounds. *Eur Respir Rev*, 10:616–620, 01 2000.

[15] Rajkumar Palaniappan, Kenneth Sundaraj, and Sebastian Sundaraj. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. *BMC Bioinformatics*, 15(1):223, December 2014.

[16] Ipek Sen, Murat Saraclar, and Yasemin P. Kahya. A Comparison of SVM and GMM-Based Classifier Configurations for Diagnostic Classification of Pulmonary Sounds. *IEEE Transactions on Biomedical Engineering*, 62(7):1768–1776, July 2015.

[17] I. Sen, M. Saraclar, and Y. P. Kahya. Computerized Diagnosis of Respira tory Disorders: SVM Based Classification of VAR Model Parameters of Respiratory Sounds. *Methods of Information in Medicine*, 53(04):291–295, 2014.

[18] Sameer Alsmadi and Yasemin P. Kahya. Design of a DSP-based instrument for real-time classification of pulmonary sounds. *Computers in Biology and Medicine*, 38(1):53–61, January 2008.

[19] H. S. Porieva, K. O. Ivanko, C. I. Semkiv, and V. I. Vaityshyn. Investigation of lung sounds features for detection of bronchitis and copd using machine learning methods. *Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia*, (84):78–87, Mar. 2021.

[20] Elena-Anca Paraschiv and Cătălina-Maria Rotaru. Machine learning approaches based on wearable devices for respiratory diseases diagnosis. In *2020 International Conference on e-Health and Bioengineering (EHB)*, pages 1–4, 2020.

[21] Zeenat Tariq, Sayed Khushal Shah, and Yugyung Lee. Lung disease classification using deep convolutional neural network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 732–735, 2019.

[22] Reetodeep Hazra and Sudhan Majhi. Detecting respiratory diseases from recorded lung sounds by 2d cnn. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, pages 1–6, 2020.

[23] Nishi Shahnaj Haider, Bikesh Kumar Singh, R. Periyasamy, and Ajoy K. Behera. Respiratory Sound Based Classification of Chronic Obstructive Pulmonary Disease: a Risk Stratification Approach in Machine Learning Paradigm. *Journal of Medical Systems*, 43(8):255, August 2019.

[24] Gaëtan Chambres, Pierre Hanna, and Myriam Desainte-Catherine. Automatic detection of patient with respiratory diseases using lung sound analysis. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, 2018.

[25] David Skalicky, Vaclav Koucky, Daniel Hadraba, Martin Viteznik, Martin Dub, and Frantisek Lopot. Detection of Respiratory Phases in a Breath Sound and Their Subsequent Utilization in a Diagnosis. *Applied Sciences*, 11(14):6535, July 2021.

[26] Chris Ding and Hanchuan Peng. MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA. *Journal of Bioinformatics and Computational Biology*, 03(02):185–205, April 2005.

[27] Francisca Rosario and K Thangadurai. RELIEF: Feature Selection Approach. *ISSN 2278 – 0211 (Online)*, 4(11):218 – 224, October 2015.

Appendix A

Resume of studies reviewed in State of the Art

Table A.1: Resume of studies reviewed in State of the Art

Reference	Population	Classes	Features	Classifier	Results
[15]	17 normal pathology, 26 airway obstruction pathology, 25 parenchymal pathology	Normal, Airway obstruction, and Parenchymal pathologies	MFCC	SVM, KNN	Accuracy 98.26% with KNN and 92.19% with SVM
[16]	20 healthy patients, 10 with obstructive disease and 10 with a restrictive disease	Healthy, Obstructive disease and Restrictive disease	Vector Auto-Regressive Model Parameters	SVM, GMM	Total Correct Classification Rate of 85%.
[18]	Offline: 20 healthy subjects, 20 sick patients, Online: 12 healthy subjects, 13 pathological subjects	Healthy, Pathological	Auto Regressive Model	Minimum Distance Classifier, KNN	Offline: Accuracy 97.5% , Online: Accuracy 96%
[19]	112 healthy recordings, 84 bronchitis and 100 COPD	Healhy, Bronchitis and COPD	Spectral, Spectrogram, Savelet, MFCC and Logarithmic (Mel) Filterbank Energies	Decision Trees, Discriminant Analysis, SVM, Logistic Regression, KNN and Ensemble Learning	Accuracy 93.2% with Quadratic Discriminant and Combination of Wavelet and Logarithmic (Mel) Filterbank Energies features

Reference	Population	Classes	Features	Classifier	Results
[20]	ICBHI database	COPD, URTI, Bronchiectasis, Bronchiolitis, Pneumonia and Healthy	MFCC	CNN	Accuracy 90.21% and Weighted Average f1 score 0.89
[21]	ICBHI database	COPD, URTI, Bronchiectasis, Bronchiolitis, Pneumonia and Healthy	Spectrogram features	CNN	Accuracy 97%
[22]	ICBHI database	COPD, URTI, Bronchiectasis, Bronchiolitis, Pneumonia and Healthy	MFCC	CNN	Accuracy 92.39%
[23]	30 COPD and 25 healthy subjects	COPD and Healthy	Temporal, Spectral, Spectro-Temporal and Spirometry features	SVM, KNN, Decision Tree, Discriminant Analysis and Logistic Regression	Accuracy 100%
[24]	ICBHI database	Healthy, Sick	Presence or absence of wheezes and crackles	Class Ratio	Correct Predictions 85%