Classification of Auscultation Sounds Using a Smart System

by

Zahra Kanji

B.S., Massachusetts Institute of Technology (2002)
M.P.H., Harvard School of Public Health (2017)

Submitted to the
Integrated Design and Management Program and
Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering & Management and
Masters of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2022

Author .....
Integrated Design and Management Program and
Department of Mechanical Engineering
May 18, 2022

Certified by .....
Daniel Frey, Thesis Supervisor
Professor of Mechanical Engineering

Accepted by .....
Tony Hu
Director, Integrated Design and Management Program

Accepted by .....
Nicolas Hadjiconstantinou
Chairman, Committee on Graduate Students

Classification of Auscultation Sounds Using a Smart System

by

Zahra Kanji

Submitted to the
Integrated Design and Management Program and
Department of Mechanical Engineering
on May 18, 2022, in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering & Management and
Masters of Science in Mechanical Engineering

Abstract

Respiratory diseases are a leading cause of death worldwide. Despite modern medicine, treatment of lung diseases is limited by the tools available to diagnose these disorders, especially in low resource settings. While tools such as chest x-ray and CT scans are highly accurate, their high cost provides a high barrier for many patient populations. The physical exam has been a long standing tried and true method that provides a low cost solution for for diagnosis of many common lung diseases including pneumonia. However, this method is subjective and its sensitivity is limited to the operator ability.

Lung sound classification and using a digital stethoscope can be used to provide an immediate diagnostic for respiratory-related diseases. The International Conference on Biomedical and Health Informatics (ICBHI) created a sound data base in 2017 that is annotated with a classification of the lung sound by physicians. In this thesis, artificial intelligence libraries are used in a deeo learning architecture to identify and classify the lung sounds. The data set was split into training and test data and evaluated using standard performance metrics: precision, 92.3%, accuracy, 87.3%, sensitivity (recall), 87.1%, specificity, 87.5% and F1 Score, 0.89%. Because the data set is skewed right, the best evaluation metric is the F1 Score, which is a weighted average of precision and sensitivity. The F1 score was found to be better than other comparable known attempts on this same data set.

The space for new, innovative, portable and affordable diagnostic devices that aid patients towards pulmonary health and wellness will likely push the development further of the acceptance of electronic auscultations. As telemedicine grows, this will also drive up the demands for such devices. Other holistic measures that are used in medicine will likely also be be developed as the landscape of healthtech changes what is possible.

Thesis Supervisor: Daniel Frey

Title: Professor of Mechanical Engineering

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

Respiratory diseases are a leading cause of death worldwide. Despite modern antibiotics, treatment of pneumonia and other lung diseases is limited by the tools available to diagnose these disorders especially in low resource settings [58]. While tools such as chest x-ray and CT scans are highly accurate, their high cost provides a high barrier for many patient populations [7]. Thus, physical exam remains as a core, robust method for diagnosis of many common lung diseases. Due to limited sensitivity and the range of physician (or other healthcare provider) ability, the pulmonary physical exam is often insufficient for diagnosis [20]. The motivation for this thesis research is to quantify the the art of medicine of clinical physical exam to using modern tools.

Recent technological advances in diagnostics have had some clinicians questioning the usefulness of more traditional, subjective, holistic examinations of diagnosis [52]. Counterarguments state that physical examinations remains an important aspect of medical care and requires training. Clinicians who are skilled at the bedside examination make better use of diagnostic tests and order fewer unnecessary tests [63]. Auscultations by stethoscope is considered to be an essential, low-cost tool and regarded as a diagnostic irreplaceable tool. Even in cardiology, where auscultation is considered to play a central role in examination, there too it is quickly becoming a lost art [24]. Auscultation should not be used as the sole reference for validating crackle detection algorithms. When used properly, the stethoscope are a valuable and

cost-effective clinical tool that a well-trained provider can use to make a rapid and accurate diagnosis with fewer additional tests [14].

In modern day, tech is making waves of change in all the industry it touches. Tech is entering the healthcare space and transforming the job of physicians. Some of their tasks will be taken over by artificial intelligence (AI), leaving them to have more time to work with patients with care and patience [3]. Physicians will have more access and ease of knowledge of up-to-date information in medical research. They will likely have less administrative tasks and note taking lags [17].

All physicians, in every speciality, around the globe, are all trained to use a stethoscope right from their time in medical school. They routinely listen to lung sounds during general examinations or when patients indicate distress and especially reach for it in respiratory cases [51]. Lung auscultations are an important method for physicians in decisions. Auscultation is a subjective method and improper treatment and referrals accumulate an increased time and monetary cost [24]. Training physicians in this art is a challenging task because of varying perception of sound and lack of common nomenclature to express the description of the sound [51].

1.1 Stethoscopes

Since the 1800s, the stethoscope has grown in popularity and eventually been adopted as the physician's primary medical tool. From the 1900s on-wards, stethoscopes look fairly similar to how they look today with a bin-aural design, flexible tubing, and a rigid diaphragm. Bowles and Sprague developed the combined bell and diaphragm design in 1925, then shortly following World War II, Sprague, Rappaport, and Groom experimented with the design before finding the optimal combination of the classic double-tube Rappaport-Sprague stethoscope [66].

Sound is produced by an organ in the body, and these acoustic waves cause a vibration in the stethoscope's chest piece, which acts as a resonator. This chest piece has two sides shapes, a flat, disk-like diaphragm and a hollow cup that oscillates with different frequency ranges. The acoustic vibration travels through an air-filled tube

that connects to two earpieces and relaying the sounds of the patient's body to the listener. The output of the stethoscope is designed to maximize sound pickup. While the sound is amplified, the signal sounds are still quite low for the listener's ears and it takes serious training to hear the subtle differences [10].

Lung auscultation is a diagnostic method used for checking the integrity of lung function [66]. It is a standard preliminary examination for all patients at hospitals. Health providers use stethoscopes to listen for changes in lung sounds to assess whether a patient has any obvious lung abnormalities. Despite many advances in medical equipment, the traditional analog stethoscope remains the main diagnostic tool used by physicians in lung auscultation [51].

Today, factors such as air pollution, unbalanced diets, excessive stress, erratic sleep patterns have resulted in more people suffering from respiratory system diseases. In a recent Department of Health statistics report, lung and respiratory-related diseases ranked fourth and seventh among the top ten leading causes of death. Being able to detect these subtle changes and sounds is crucial in our modern lifestyle and our quest for better health and preventive care [27].

1.2 Human Ear Limits

Studies have been conducted to test the human's ear capability to detect crackles in an auscultation signal using simulated crackles superimposed on real breath sounds [13]. The most important detection errors are due to the intensity of the respiratory signal, the type of crackles and the amplitude of crackles. The validation of automatic crackles' detection algorithms should not take auscultation as a unique reference [23].

Our understanding of the mechanics of breath sounds is imperfect. Analysis of respiratory sounds in greater detail is an opportunity for us to improve this understanding and create an objective relationship between abnormal respiratory sounds with respiratory pathology [33]. This will aid in the development of a classification system to more precisely qualify respiratory sounds [23]. Current auscultations are largely from stethoscope readings which are subjective. An objective system of mea-

sure with reproducible results is needed and can be provided with smart stethoscopes [21].

Smart stethoscopes will also be able to capture trend data from longer duration monitoring for patients at home or at hospital. It can also serve as an aid to students in medicine learning how to auscultate as it shows the association between acoustical signal and its image [21].

1.3 Human Pulmonary Auscultations

The primary purpose of the human respiratory system is to exchange carbon dioxide in our bloodstream with the oxygen from the outside environment. The lungs act as the exchange border between the atmosphere and our bloodstream, by circulating the air inside the lungs with every breath, filling them with the surrounding environment's available oxygen and expelling carbon dioxide waste [35].

The next section will cover the fundamental understanding of the human respiratory system, its basic anatomy and function, the pulmonary auscultation guidelines, and the principal characteristics of abnormal sounds and their clinical significance. Extra attention will be placed on the types of abnormal sounds that are most relevant; crackles and wheezes.

1.3.1 Anatomy

The human respiratory system is divided into two respiratory tracts, the upper respiratory tract and the lower respiratory tract, see Figure 1-1 [2]. The upper respiratory tract consists of the organs which are outside the chest cavity area, which includes the nose, pharynx and larynx. The lower respiratory tract consists of the organs which are almost entirely inside the chest cavity area, which includes the trachea, bronchi, bronchioles, alveolar ducts and alveoli [2].

Functionally, there are two zones, the conducting zone and the respiratory zone. The conducting zone is made up of the respiratory organs that form a path that conducts the inhaled air into the deep lung region. The respiratory zone is made up

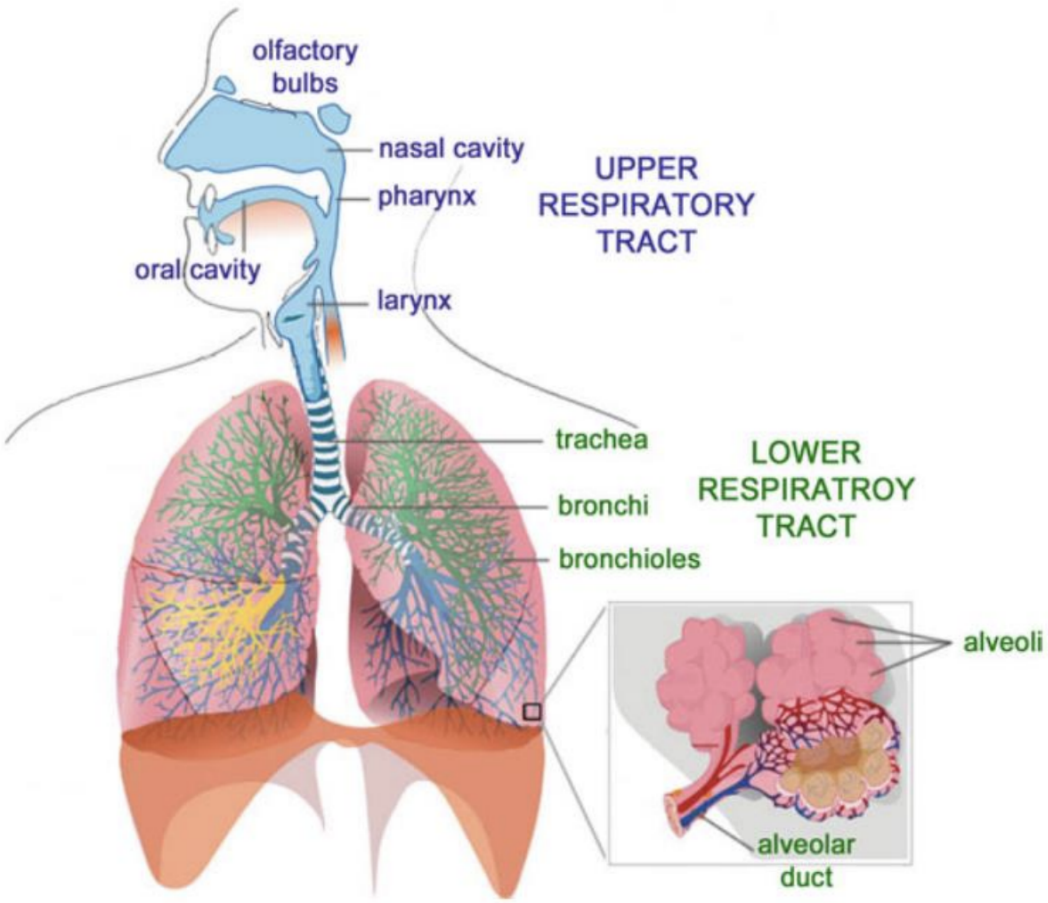

Figure 1-1: Schematic of upper and lower respiratory systems

of the alveoli and the tiny passageways that open into them where gas exchange takes place [35].

1.3.2 Physiology

Most of the respiratory tract exists primarily as a system of pipes for air to travel into alveoli, the only part of the lung that exchanges oxygen and carbon dioxide with the blood. The alveoli are a single cell membrane that allows for gas exchange to pulmonary vasculature [35]. The diaphragm and intercostal muscles help with inspiration by creating a negative pressure inside the chest cavity. The lung pressure becomes less than the atmospheric pressure causing the lungs to fill with air. The muscles help with expiration by creating a positive pressure inside the chest cavity,

where the lung pressure becomes greater than the atmospheric pressure to empty the lungs of air [35].

1.3.3 Auscultation Procedure

To auscultate the lungs properly, the physician follows a general set of steps [48]:

1. Patient is in a seated or resting position in a quiet environment.

2. Remove cloth that might interfere with the auscultation.

3. Patient to take deep breaths with an open mouth.

4. With the stethoscope's diaphragm, auscultate anteriorly at the apices, and move downward till no breath sound is heard. Listen to the back, starting at the apices and moving downward. One complete respiratory cycle should be heard at each point.

5. Compare symmetrical points on each side.

6. Listen for the quality of the breath sounds, the intensity of breath sounds, and the presence of adventitious sounds.

1.4 Physics

The human thorax is comprised of four different types of materials with significantly different acoustic properties: hard tissue (bone), soft tissue (muscle, fat, etc.), air in the major conducting airways of the bronchial tree, and parenchymal tissue that is a heterogeneous mixture of soft tissue and air found in the alveolar sacs and smaller bronchioles. The characteristics of these different components affect how sound is transmitted [2].

1.4.1 Absorption of Sound

Sound in the lumen of the lung airways experiences a frequency-dependent absorption into the airway walls and surrounding parenchymal tissue, in which high-frequency sounds propagate further within the airway branching structure, while low-frequency sounds tend to couple into the airway walls sooner. Due to the attenuation of higher frequency sounds in the surrounding parenchymal tissue, most of the signal energy of breath sounds recorded on the torso surface is concentrated at lower frequencies [59].

Analysis of sound transmission in the chest cavity indicate that the chest acts as an overall low-pass filter by absorbing higher frequencies as sound travels through it. This filtering effect is altered with the presence of different lung conditions, such as consolidation or fluid build-up, which can create large acoustic impedance mismatches with healthy parenchymal tissue and air. Alternatively, it can also couple to surrounding soft tissue of the chest wall and propagate with less attenuation, as compared to healthy parenchymal tissue [60].

1.4.2 Transmission and reflection

Transmission and reflection of sound waves is caused by interfaces between the semirigid chest wall, pleural spaces which normally contain air but can fill with fluid in disease, and the lung tissue, which is typically approximated as a homogeneous mixture of gas and tissue. Intensity of the sound is a quantitative measure of transmission [2].

1.4.3 Resonance

Resonant frequencies are frequencies at which acoustic waves are reflected back and forth constructively due to interaction with boundaries or interfaces leading to an amplified response. For the chest overall, resonance depends on several factors, including the size of the thorax. The lowest, resonant frequency of the chest for adult men is around 125 Hz, for adult women 150–175 Hz, and for children at 300–400

Hz [60]. For sound traveling in the lumen of the airways, the resonant frequencies are a strong function of the geometry and wall properties of the airways whereas direct chest stimulation will generally bypass these differences. Resonances may also occur when pathologies create trapped air cavities below the torso surface, as is in pneumoperitoneum [59].

1.5 Lung Sounds

Lung auscultation is one of the simplest, non-invasive screening methods we have for respiratory disease or diseases that affect lungs as part of the symptoms, such as congestive heart failure [2]. Using a stethoscope is a quick and cheap way of screening patients. Audible symptoms are prone to subjectivity of the investigator. Creating tools that can assist in both training and diagnostic screening using auscultation will standardize this art form and make it's capability impactful [21].

Lung sounds are difficult to define because of their inherent link to anatomy and condition severity. This also makes training challenging. The waveform from the lung can also vary by other factors including recording site, flow rate, lung volume, body position, and different breathing manoeuvres. Furthermore, sound changes with development, growth, age, environmental changes [26].

Respiratory sounds occur as the result of air flowing through the lungs and are categorized as normal or abnormal (adventitious). Normal respiratory sounds are defined as those that are in healthy airways by physiological unforced breathing. Lung sounds can be divided roughly into normal and abnormal sounds, as shown in Figure 1-2. Normal breath sounds can be divided into bronchial, vesicular-bronchial, vesicular, and tracheal sounds. Absence or deficiency of normal breath sounds or manifestation of adventitious sounds may be an indicator of pulmonary disease. Abnormal breath sounds can be divided into crackles, rhonchi, and wheezes.

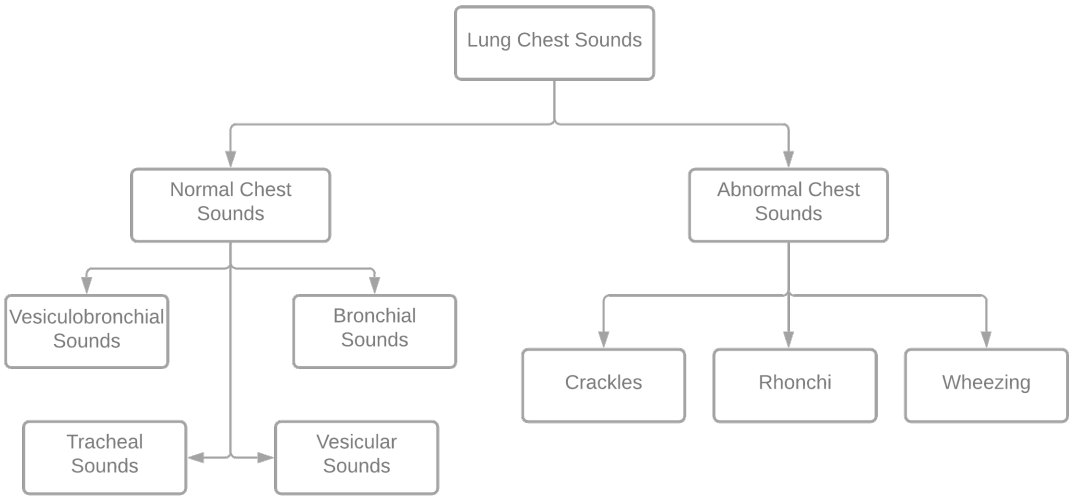

Figure 1-2: Lung sound classification

1.6 Normal Lung Sounds

1.6.1 Vesicular sounds

Vesicular murmurs can be heard during auscultation in most of the lung areas. They are easy to hear during inspiration, and can only be heard in the beginning of expiration. They have a low intensity and if the chest wall is thickened, it can appear absent [71]. Another reason they could be absent is also if the lung has collapsed due to the fluid or air pressure of the pleural cavity. In this case, no ventilation in the affected lung area, or after a pneumonectomy [38].

1.6.2 Bronchovesicular sounds

Normal bronchovesicular sounds can be heard between the scapula at the posterior chest and center part of the anterior chest [71].

1.6.3 Bronchial sounds

Bronchial sounds are audible over the chest near the second and third intercostal spaces. They are similar to tracheal sounds, high in pitch and can be heard during

both inspiration and expiration. They are more clearly heard than vesicular sounds during expiration. The sounds are high-pitched, higher than vesicular sounds, loud and tubular [71].

1.6.4 Tracheal sounds

Tracheal sounds fall in the frequency range of 100-4,000 Hz. They can be heard over the trachea, above the sternum, in the suprasternal notch. They are generated by turbulent airflow passing through the pharynx and glottis. These sounds are not filtered by the chest wall [71].

1.7 Abnormal Lung Sounds

The absence or deficiency of normal breath sounds or manifestation of adventitious sounds may be an indicator of pulmonary disease. Different abnormal lung sounds indicate different diseases [71]. Pneumonia, chronic bronchitis, bronchiectasis, congestive heart failure, and obstructive pulmonary disease produce crackles. Obstructive pulmonary disease, asthma, and bronchial stenosis produce wheezes. Pneumonia, chronic bronchitis, and congestive heart failure produce rhonchi [68]. Figure 1-3 shows the different time-domain characteristics and spectrogram of a normal, wheeze and crackle lung sound cycle. The wheeze and more so the crackle are the two hardest abnormal sounds to distinguish [19].

Abnormal breath sounds are another important component in the diagnosis of lung diseases. Different lung diseases create different lung sounds: Table 1.1 lists the most commonly known associations between abnormal lung sounds and lung diseases. Pneumonia, chronic bronchitis, bronchiectasis, congestive heart failure, and obstructive pulmonary disease produce crackles [70]. Obstructive pulmonary disease, asthma, and bronchial stenosis produce wheezes. Pneumonia, chronic bronchitis, and congestive heart failure produce rhonchi. The combined population of patients suffering from these diseases is about 30% of the global population [68].

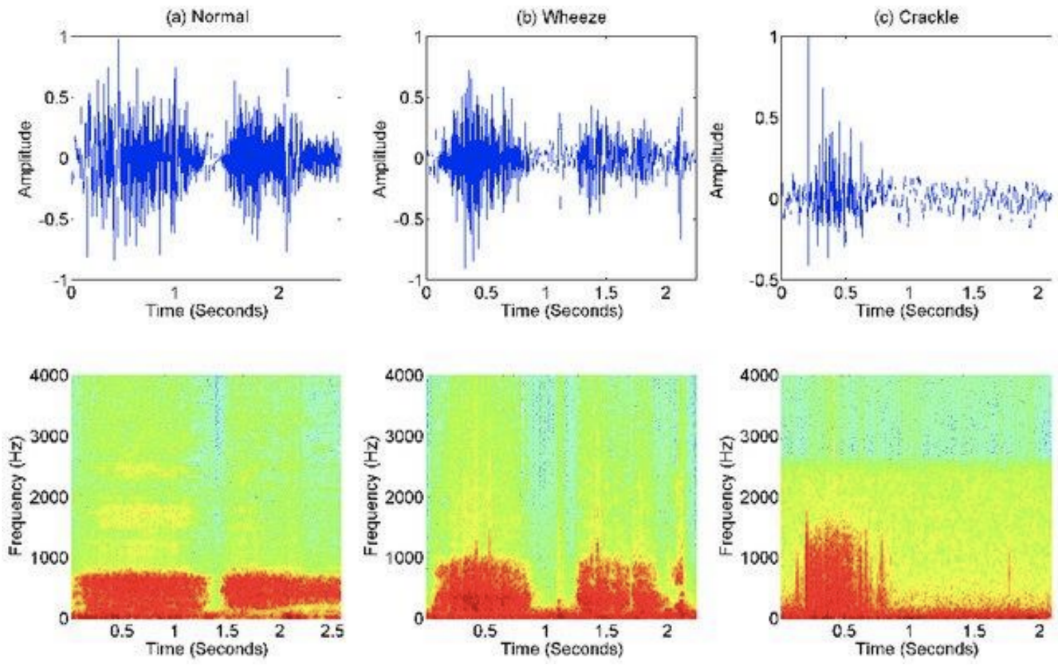

Figure 1-3: Time-domain characteristics and spectrogram of (a) normal, (b) wheeze, and (c) crackle lung sound cycle

	Crackles	Wheezes	Rhonchi
Pneumonia	X		X
Chronic bronchitise	X		X
Bronchiectasis	X		
Congestive heart failure	X		X
Obstructive pulmonary disease		X	
Asthma		X	
Bronchial stenosis		X	

Table 1.1: Abnormal lung sounds and lung diseases

1.7.1 Crackles

Crackles, also known as rales or crepitations, are short explosive clicking or crackling sounds that occur from the opening of small airways with a short duration, ranging between 5-40ms. Crackles can occur anywhere in the lung, and can be present unilaterally or bilaterally [19].

Crackles are typically divided into several main types depending on the characteristic of the sound; coarse, medium, fine, wet or dry [68]. They can be heard most often during an inspiration. Coarse crackles tend to be long, loud and low pitched,

towards the early part of the inspiration. Fine crackles are often soft and high pitched and short, occurring towards the later part of the inspiration [19].

Crackles can also occur in healthy lungs. If they are chronic, this is an indication that there are small cavities in the lungs collapsed by fluid and/or a lack of aeration during expiration. These symptoms occur in patients with pneumonia, pulmonary fibrosis, acute bronchitis and other conditions [37].

Crackles are subtle, hard to hear sounds, so a stethoscope or microphone that is rubbing over some cloth or even skin or hair can produce similar sounds [34]. These sounds can easily missed as physicians often ask patients to take deep breathes and deep breathing masks more crackles than superficial breathing [25]. Fine crackles are more readily recognizable from their waveform as the amplitude of the crackles differs more significantly from classic lung sounds. Generally, the duration of a crackle is lower than 20 ms and the frequency range is between 100 and 200 Hz [42].

1.7.2 Wheezes

Wheezes are continuous sounds, which can last up to the whole respiratory in and out breath cycle. They are caused by air being forced through small paths due to obstructions in airways, creating a whistling sound. Wheezes can be detected over the whole chest area and trachea [5]. Wheezes can vary between patients and the sound depends on the severity as well as how it was auscultated and where the stethoscope was placed [33]. Wheezes can be indications of respiratory conditions such as asthma attacks allergies that cause narrowing or obstruction of airways. They can be heard in healthy patients in intense physical exercise when airflow is increased. Their waveform is characterized by periodic waveform with a frequency usually greater than 100Hz and lasting over 100ms [5].

Chapter 2

Machine Learning and Artificial Neural Networks

In this chapter, I describe the machine learning, artificial neural networks and sound processing techniques used in the classification of lung sounds.

2.1 Artificial Neural Networks

In this section, artificial neural networks, the types of architectures, activation functions, loss functions, optimization methods and regularization methods are discussed.

2.1.1 Feed Forward Neural Networks

Feed forward neural networks (FNNs), or multi-layer perceptrons (MLPs), are the archetypes of deep learning models. The purpose of these networks is to approximate some function f by mapping an input domain to an output domain. This can be applied to solving complex problems and moving from high dimensional data to a set of labels. These networks consist of multiple layers, where the first layer is the input layer, the middle layers in the network are hidden and the last is the output layer. Each layer create an additional level of abstraction [57].

Each layer has of a number of neurons that represent activation values and that determines the width of that layer. Each neuron has a number of input weights that connect to each of the neurons of the previous layer, with the exception of the neurons in the input layer. The activation values of the input layer are propagated forward in the direction of the output layer with no feedback connections. This is where the name feed forward names is derived from [62].

The network is associated with a directed acyclic weighted graph that describes how the functions are composed together [62]. The network's parameters consists of the weights and biases between layers. The output activation values of a layer is represented as a vector, with each entry of the vector representing the activation value of a single neuron. The size of the vector corresponds to the number of neurons in that layer [8].

The weights between layers are represented as a 2D matrix. Each entry of the matrix at coordinates i, j represents the weight connecting the neuron i from layer l -1 to the neuron j in layer l. The biases between layers is represented as a vector with the same size as the number of neurons in the next layer [8].

2.1.2 Convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) are similar to feed forward neural networks. They both use the concept of neurons; each neuron receives an input and performs an operation. CNNs are different, specialized network for processing data that is more grid-like, as you find in a time-series [62]. It has three types of layers: convolutional layers, pooling layers and fully connected layers [6]. Unlike FNN, CNN uses parameters, referred to as kernals, to decrease the number of parameters needed for high-dimensional input grids. CNN always has the same number of parameters, in larger input grids. Kernels are used as detectors for local patterns in data [62].

In a convolutional layer, kernels are a set of small matrices that are applied to the input grid. The input values in each window are convolved with the weights of that kernel, added to a bias, and then processed into a nonlinear function. The output is a single output value for each input window creating a feature map [31]. By applying

multiple kernels to that same grid I get the same number of feature maps as the number of kernels in the convolutional layer. The resulting output grid has a depth of n feature maps. The remaining spatial dimension size depend on the previously defined window size, stride and input spatial dimension size [31].

2.1.3 Loss Function

There are several types of loss functions which have different purposes in machine learning. Categorical cross entropy is a loss function used for the classification of data samples. It is used for single label classification, where each data sample belongs to only one class. It compares the predicted class distribution with the true class distribution [29].

- The categorical cross entropy loss is defined as [29]:

$$-\sum_{i=1}^{C}(t_i log(p_i))$$

  - C is the number of classes,
  - t_i is the true probability of class
  - i and p_i is the predicted probability of class i

- Since only one class is set to the value of 1 and the remaining values in t are 0, it is equivalent to [29]:

$$-log(p_c)$$

  - c is the class index
  - $t_c = 1$

2.1.4 Activation functions

Activation functions allow the generalization of neural networks in solving various tasks. They can improve training speed and convergence.

Rectified Linear Unit (ReLU)

The rectified linear unit (ReLU) activation function is popular in the deep learning as it has allowed deeper models to converge faster during training and achieve better results. It is defined as [43]:

$$f(x) = max(0, x)$$

ReLU is faster to compute than other activation functions and allows for simpler initialization of network parameters [29]. It induces sparse activation of the network's hidden units and it has less vanishing gradient problems when compared to the logistic sigmoid and hyperbolic tangent activation functions [43]. However, it also has some issues. Some hidden units can become stuck in inactive states regardless of the input. This means that the gradient of the unit will always be 0 and the unit will stop training and thus decreases the model's capacity to learn [29].

Leaky ReLU

The Leaky ReLU (LReLU) activation function fixes the issue of inactive neurons. It is defined as:

$$f(x) = x, \qquad if x > 0$$

$$f(x) = \alpha x, \qquad otherwise$$

The α value is a static value defined during the creation of the neural network as the slope of the negative section of the function. It allows the gradient to be different from 0, if $\alpha \neq 0$, which allows the gradient to propagate through the neuron and to train the weights.

2.1.5 Fourier Transform

The Fourier transform is an integral transformation of a signal from the time domain to the frequency domain. This allows for examination of the signal in terms of the presence and strength of the various frequencies. Looking at the frequency domain allows for methods of signal processing to be applied, including filtering,

modulation and sampling of a signal [64].

The calculation of the Fourier transform of a finite sequence of values is done with the Discrete Fourier Transform (DFT) method. The Fast Fourier Transform (FFT) is an algorithm to calculate the DFT of a signal. Applying the DFT to multiple equally spaced overlapping small windows of the signal and stacking each window's spectral creates a spectrogram that shows the evolution of the signal's frequency spectrum over time [46]

2.1.6 Power Spectral Density

The Power Spectral Density (PSD) represents which frequency variations are strong and which are weak, in units of energy per frequency. PSD is an analysis method used when a measured signal in the time domain is transformed into the frequency domain through a Fourier transform. It used to detect the frequencies and amplitudes of oscillatory signals and any periodicity in the data [49].

2.1.7 Mel Spectrogram

Studies show that humans do not perceive frequencies on a linear scale. We are much better when we decipher differences in lower frequencies than in higher frequencies. In 1937, a new scale was proposed by Stevens, Volkmann, and Newmann where a unit of pitch (equal distances in pitch) sound equally distant to the listener (as judged by humans). This is known as the mel scale. It is a way to mimic how the human ear responds to varying frequencies [39]. The frequencies are transformed to the mel scale to create a Mel Spectrogram (MS).

These steps are required to obtain the mel spectrogram [50]:

1. Separate signal to windows: Sample the input with windows of size n_fft, making hops of size hop_length each time to sample the next window.

2. Compute FFT for each window to convert from time domain to frequency domain.

3. Generate a mel scale: Take the entire frequency spectrum, and separate it into n_mels evenly spaced frequencies according to the mel distance.

4. Generate Spectrogram: For each window, decompose the magnitude of the signal into its components, corresponding to the frequencies in the mel scale.

The mel frequency scale is defined as [39]:

$$mel = 2595 \times \log_{10}(1 + hertz/700)$$

and its inverse is [39]:

$$hertz = 700 \times 10^{mel/2595} - 1$$

2.1.8 Mel Frequency Cepstral Coefficients (MFCC)

The Mel Frequency Cepstral Coefficients (MFCC) has been the standard in of almost all modern applications in speech and music. It is a tool used for feature extraction [1].

The algorithm for computing the MFCC of an audio signal uses the following steps [53]:

1. Frame the signal.

2. Compute the Discrete Fourier Transform (DFT) for each window.

3. Apply the Mel-filterbank to convert frequency to the Mel-scale.

4. Take the Log amplitude of the Mel-scaled spectrum.

5. Compute the Discrete Cosine Transform on the Mel-scaled Log amplitudes.

Framing

Audio signals change their statistical properties overtime because they are not stationary. By framing a signal into smaller chunks I can appropriate the signal to be stationary and analyze it [53].

DFT

Applying the Fourier transform to each of chunk of signal creates a spectrogram of the signal. This spectrogram denotes spectral content of the signal in the Hertz scale. At this step, I lose the phase information and keep only the absolute values [64].

Mel-Scale

Applying the Mel-filter bank converts the Hertz values into the Mel-Scale. The Mel-Scale is a perceptual scale of pitch which models pitch closer to what humans perceive rather than actual Hertz values [50].

Log Amplitude

Taking the Log of the amplitude of the Mel-scaled spectrum gives a power spectral density estimation. This denotes the energy of the different frequency bins [53].

Discrete Cosine Transform

Finally, I compute the Discrete Cosine Transform of the log power spectrum, treating it again as a signal. The resulting coefficients is the MFCC. This is a cepstral representation of the audio clip. A cepstrum contains information about the rate of change at different spectrum bands, which would be the Mel-spaced frequency bins [54].

Crackles

MFCC might not be applicable to crackles because of extremely short duration, less than 100 ms, and because of their sudden change in volume [53]. MFCCs have historically been used with sequence classifiers such as Hidden Markov Models, so its dependent on the type of classifier [55]. They were meant to model a large vocabulary in speech recognition, while crackle classification problem is more binary.

2.2 K-Means Clustering

The most widely used and commonly known method for data-segmented clustering is K-means clustering [65]. The main purpose for K-means clustering is to process a large number of high-dimension data to representative data, cluster centers. These cluster centers can be used to create data classification and compress large amounts of data. K-means clustering requires finding the number of clusters and reducing the errors in the cluster iteratively until the errors stop, or there is convergence to the final clustering results. The steps of implementing the K-means algorithm are [56]:

As an example, if the training sample is x^i

$$x^i = x^1, x^2, ..., x^m$$

1. Select K random cluster centers as u_j:

$$u_j = u_1, u_2, ..., u_k$$

2. Repeat until convergence:

   (a) For each x^i, find and assign it to the nearest cluster center

$$c^i = argmin_j \|x^i - u_j\|^2$$

   (b) For each category u_j, recalculate the mean value of the cluster and update the cluster center.

$$u_j = \frac{\sum_{i=1}^{m} \{c^i = j\} x^i}{\sum_{i=1}^{m} \{c^i = j\}}$$

2.3 K-Nearest Neighbor Algorithm

The K-nearest neighbor algorithm clusters objects in the same category closer in distance. The K-nearest neighbor algorithm is [22]:

1. Determine the number of nearest points of test data x against training data K. Use an Euclidean distance equation to compute the distance. For two points in k dimensional space, $x = [x_1, x_2, ..., x_k]$ and $y = [y_1, y_2, ..., y_k]$. The Euclidean distance between the two calculated by

$$d(x, y) = \sqrt{\sum_{i=1}^{k}(y_i - x_i)^2}$$

2. When test data x has more data-points than a certain category of K-nearest points, x is decided to be part of that category.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Materials and Methods

In this chapter, I describe the dataset that was used in this work to develop the classification methods, the signal processing methodology, the libraries and tools used to implement the methods and the experimental methodology for comparing results of different methods. I also describe the implementation challenges, the proposed solutions, the advantages and limitations, our choices and our reasoning.

3.1 Approach

I follow these steps: generating a training set, pre-processing audio files, feature selection, selection and training of a classifier.

1. A large amount of representative data is required to validate machine learning models. The more data used, the more likely it is that the data is representative of the general case. This reduces the risk of over-fitting models while making it more robust against outliers [16].

2. The audio files are labelled as either containing crackles or not, as a whole sample. I also need to look individual crackles in each file. This reduces the amount of normal data I have.

3. Pre-process the files to be classified in the same way that the training data has been pre-processed. Reduce the size of each classification and accurately

pinpoint locations of the abnormal lung sounds within a given audio file.

4. Find features that represents the data while reducing the number of dimensions that the classifier has to consider. Generalizing correctly becomes exponentially more difficult as the number of dimensions of data or feature set increases [41].

5. Validate our model using cross validation to measure specificity and sensitivity. Use the classified standard from the Tromsø study to evaluate classification accuracy on clinical data [5].

3.2 Training Set Generation

Before starting, I setup a training set for our data to use. For this, I divided the data into classes of health (absences of any abnormal sounds) and unhealthy. Unhealthy was further split into having crackles or wheezes. Audio files were typically about fifteen seconds long. Excerpts of these files containing each individual crackle or wheeze was extracted into individual files. For crackles, these excerpts were about 100 ms long. Wheezes last longer and were about three seconds long. Each original file resulted in about 4000 crackle samples, that had at least one full crackle. The excerpts were overlapped in case of any crackles that occurred at the edge of the cutoff window.

3.3 Dataset

3.3.1 Background About the Dataset

The International Conference on Biomedical and Health Informatics (ICBHI) 2017 respiratory sound database was part of an organized scientific challenge to test and compare the robustness of state of the art techniques for lung sound processing and classification. The creation of this dataset was motivated by the lack of large publicly available datasets that could be used to develop and compare different lung sound processing methods. Funding was provided by the International federation of Medical

and Biological Engineering. This dataset contains various events (e.g., noise, cough, wheezes, crackles) collected from both healthy subjects and patients with different respiratory disorders and each sound is annotated by health professionals [36].

This dataset was chosen because other private datasets were smaller and came without environmental noise, making them incomparable to actual clinical practice. The dataset consists of a set of respiratory sound recordings and their respective annotation files. The audio samples were collected independently by two research teams: "Respiratory Research and Rehabilitation Laboratory of the School of Health Sciences, University of Aveiro" (Lab3R) and "Aristotle University of Thessaloniki" (AUTH) in two different countries, over several years. The dataset contains 920 annotated audio recordings which were collected from 126 participants [23].

The ICBHI 2017 classified of each individual respiratory cycle into one of four classes: Normal, Crackle, Wheeze, Both collected from healthy subjects and patients with different respiratory disorders, annotated by health professionals [36]. Figure 3-1 shows an example of an annotated sound recording.

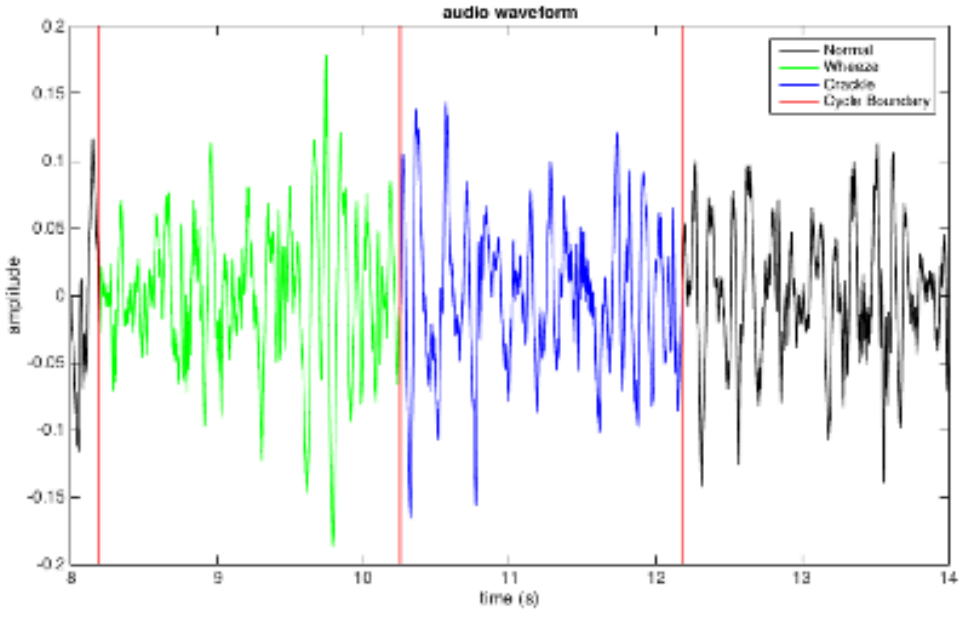

Figure 3-1: Three respiratory cycles: wheezes (green), crackles (blue), and normal sounds (black). Vertical lines separate the respiratory cycle boundaries (red)

3.3.2 Data Collection

Each audio recording was obtained using multi-channel or single-channel acquisition method, with each channel representing an auscultation point of the participant and each channel is stored in a separate file. The auscultation points are: Anterior left (Al), Anterior right (Ar), Lateral left (Ll), Lateral right (Lr), Posterior left (Pl), Posterior right (Pr) and Trachea (Tc) [36].

The types of equipment used by Lab3R to collect the lung sounds were [23]:

- Welch Allyn Meditron Master Elite Plus Stethoscope Model 5079-400 digital stethoscope
- 3M Littmann Classic II SE stethoscopes with a microphone in the main tube.
- Air coupled electret microphones (C 417 PP, AKG Acoustics) located in capsules made of teflon.

The types of equipment used by AUTH were [23]:

- WelchAllyn Meditron Master Elite Plus Stethoscope Model 5079-400 digital stethoscope
- 3M Littmann 3200 digital stethoscope.

Due to different equipment types used for the capture of the lung sounds, the sampling rate of each audio recording differs based on which was used.

3.3.3 Annotation

Each audio recording was annotated manually into individual respiratory cycles, where each cycle is given a starting timestamp, an ending timestamp, a binary number to indicate if the cycle contains a crackle and a binary number to indicate if the cycle contains a wheeze. The annotation process was done by respiratory health professionals [23].

Sound files originating from the Lab3R database were annotated by only one expert whereas the AUTH dataset was annotated by three experienced physicians, two specialized pulmonologists, and one cardiologist [36].

3.3.4 Challenge Dataset

The dataset was split into training, 60%, and testing sets, 40%. 2063 respiratory cycles from 539 recordings derived from 79 participants were included in the training set, while 1579 respiration cycles from 381 recordings derived from 49 patients were included in the testing set.

- N is the total number of normal sounds,
- Nn is the number of correctly classified normal sounds
- C is the total number of crackle sounds
- Cc is the number of correctly classified crackle sounds
- W is the total number of wheeze sounds
- Ww is the number of correctly classified wheezes
- B is the total number of sounds that contain both crackle and wheeze sounds
- Bb is the number of correctly classified sounds that contain both adventitious sounds.

3.3.5 Dataset Statistics

I performed some preliminary statistics on the dataset before I started to get a sense of the data as shown in Table 3.1.

	Cycle Count	Patient Count	Max Dur.	Min Dur.	Avg Dur.
Normal	3,642	124	16.16sec	0.2sec	2.6sec
Crackle	1,864	74	8.74sec	0.37sec	2.79sec
Wheeze	886	63	9.22sec	0.23sec	2.7sec
Both	506	35	8.59sec	0.57sec	3.1sec

Table 3.1: Statistics for each of the cycle classes.

3.3.6 Dataset Observations

The number of class samples in this dataset, Table 3.1 shows it is skewed towards normal sounds. The duration of each recording session and individual respiratory cycle has high variability with some patients lack a recording sample for at least one of their auscultation points or having too many samples for some auscultation points.

Furthermore, some recordings have extremely large respiratory cycles, which is caused from the placement of the device. Smaller duration respiratory cycles are often the ending or starting cycles of a recording and have been cut off. Finally, the properties of the equipment creates different sampling rates in the recordings. All these factors including noise artifacts and patient demographics make it difficult to apply a simple method for classification.

3.4 Libraries

The project was implemented using the Python programming language and Google Colab environment. The module used to load and down-sample the audio files was the Librosa module. Scipy module was used to filter the audio files with a butter worth band-pass filter. The input waveform was time-sampled by applying a window function and then a discrete Fourier transform (DFT). The Tensorflow library was used as the calculation method for Mel-frequency Cepstral Coefficients (MFCCs) of an audio signal, while the kapre module was used to calculate the Power Spectral Density (PSD) and Mel Spectrogram (MS) of the signal. The Keras library was used to implement the various neural network models, train the models and test them.

3.4.1 Signal processing methodology

In order to solve the issues of the various sampling rates, the audio recordings were down-sampled to 4000 Hz, changing the frequency range of the signal from 0 to 2000 Hz. The frequency range needed to detect crackles and wheezes is within 0 to 2000 Hz. Next, I remove noise artifacts by applying a 12th order butter-worth band pass filter using cutoff frequencies of 120 Hz to 1800 Hz. I choose this filter and frequency by following what was chosen from the method described in the best results of the official ICBHI 2017 challenge dataset paper.

Next, the signal was normalized. There were varying amplitudes from the signals coming from the different auscultation points and participant demographics. Respiratory cycles individually normalized, so cycles with differing duration do not influence the resulting distribution. I calculate the PSD, MS and MFCCs of the signal at runtime, during cycle classification. The PSD and MS of the signal are converted to the decibel scale from 0 to -80 once the values are normalized to the range from 0 to 1. The MFCCs are normalized with respect to the mean standard deviation of the coefficient values of the whole respiratory cycle signal.

3.5 Experimental methodology

In this section, I present the methods used to solve challenges that occurred during the project.

3.5.1 Batch size

In order to fit multiple audio signals in the same mini-batch, I need to mask or pad the signals. This was because each respiratory cycle varies in length of duration. This was the training process required when using Keras; the batch must be a static tensor [32]. This is a typical requirement for machine learning libraries [15]. Padding the signals to fit the same size is easier to implement compared to masking and can be done as a batch process. This does leave the possibility open that the model

learns the padding in the signal which would result in over-fitting the actual signal and an overall decreased generalization. It is possible to check for over-fitting with our evaluation metric if this was the case.

An alternative way to both masking or padding is to use the mini-batch gradient descent method by calculating the gradient of the model for each input signal individually and then averaging the gradients. The Keras library does not currently allow for this type of method of mini-batch gradient calculation [15]. This process could be implemented manually, which poses the same problem of complexity and time intensive as masking does. Thus, a mini-batch size of one was used. Training a model using one sample at a time has its own set of possible problems including optimizer stability. A work around for this is to use the stochastic gradient descent optimization method. This has been known to work better than other optimizes and have less over-fitting issues [4].

This phenomenon from stochastic gradient optimizations is understood by the following few properties [12].

1. Sharp local min are associated with poorer generalization and produce larger gradients. Two, by applying quick gradient updates to the model, the model lands on local sharp min. This nudges the model to leave that the region.

2. With repeated updates, the model reaches a region that has local min that correlates with better generalization properties.

3. By decreasing the learning rate slowly, the model becomes less sensitive to local min, which allows it to decrease model loss.

3.5.2 Under sampling

Under sampling on all methods to account for the class imbalance present in the dataset. Under sampling is a technique used specifically in cases where there is a need to balance the number of samples per class. It gives a more balanced estimate on the loss and statistics during the training of machine learning models and on the

loss and statistics during model training [61]. This helps prevents the model from memorizing the minority classes first, which gives a false positive accuracy rating as the resulting model would not be able to be applied to other cases. The maximum amount of samples from each class is defined as the number of samples in the minority class. Under sampling is applied to both training and test datasets [32].

I implemented this by doing the following: for N number of training epochs, repeat

- Sample random X amount of samples from each class
- Shuffle samples
- Train model on the samples

3.5.3 Method evaluation and comparison

The challenge has defined metrics that were used to evaluate and compare the different methods I tested. These metrics are classification accuracy and the classification confusion matrix to assess and troubleshoot potential pitfalls in the training process. Final methods were evaluated and compared using the five-fold cross validation method, see Figure 3-2. Five fold cross valuation is where the data set is split into a five sections. Each section is used as a testing set at some point. For example, in the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set [67]

3.5.4 Model hyper-parameter search method

I look at different signal features as input for the networks:

- Raw filtered audio signal (1-dimensional)
- PSD of the signal (2-dimensional)

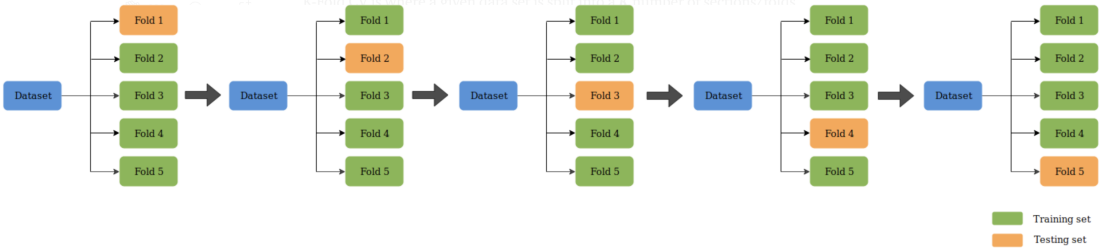

Figure 3-2: Five fold cross valuation method

- MS of the signal (2-dimensional)
- MFCCs of the signal (2-dimensional)

The hyper-parameters for the networks are:

- PSD parameters:
  - PSDNDFT (DFT window size)
  - PSDNHOP (DFT window stride)
  - PSDFMIN (spectrum y-axis min cutoff)
  - PSDFMAX (spectrum y-axis max cutoff)
- MS Parameters
  - MSNDFT (DFT window size)
  - MSNHOP (DFT window stride)
  - MSFMIN (spectrum y-axis min cutoff)
  - MSFMAX (spectrum y-axis max cutoff)
  - MSNMELS (number of mel conversion kernals to convert the spectrum to mel scale)
- MFCC parameters:
  - MFCCNDFT (DFT window size)

  - MFCCNHOP (DFT window stride)
  - MFCCFMIN (resulting spectrum y-axis min cutoff)
  - MFCCMAX (resulting spectrum y-axis max cutoff)
  - MFCCNMELS (number of mel conversion kernals to convert the spectrum to mel scale)
  - MFCCNMFCCS (number of DCT coefficients to keep)

- Number of convolution layers
- Number of fully-connected layers
- Number of kernels per convolution layer
- Kernel size and stride

The minimum duration of a respiratory cycle in the dataset was found to be is 0.2 seconds. The signal has a sampling rate of 4000 Hz which equals to 800 signal samples. The signal size has to be larger than the window size to apply DFT so 128ms (512) was selected as the as the window size for the DFT of the PSD, MS and MFCC. A larger window size increases the spectrum's frequency resolution and also increases the size of the spectrum grid on the frequency axis. Different values were tried and experimented with for the number of mels and kernals in the convolution layers. The best results was found from using 64. Three was used as the convolution layers to balance between over-fitting behavior and having too few layers which hinders the models ability to learn the signal patterns. The number of fully-connected layers was kept at one to also limit over-fitting behavior.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Results

In this section the results are presented and discussed to gauge how well the algorithm performed.

4.1 Noise Reduction

After clipping the signals to the appropriate length, the signal was processed first through the denoising method, Figure 4-1. Looking at the signal in an expanding time view, Figure 4-2 it is clearer to see that the denoising worked and the signal is smoother.

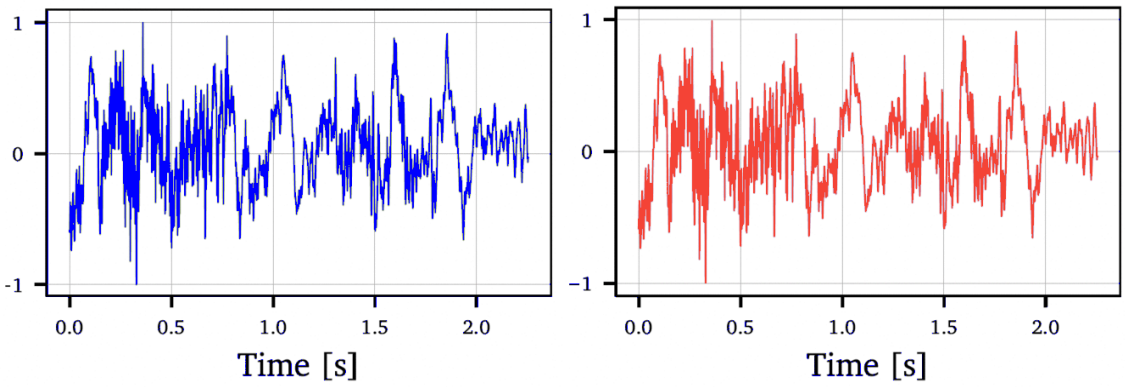

Figure 4-1: Original Signal (Blue), on the left vs the Denoised Signal (Red) on the right

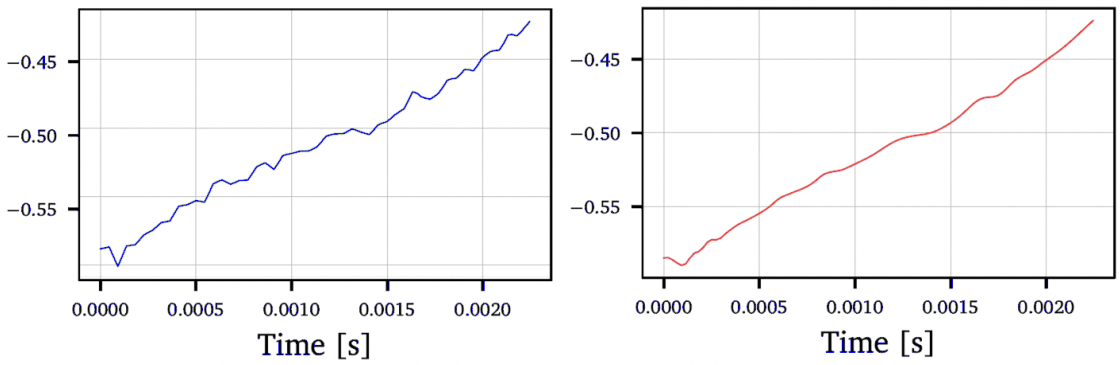

Figure 4-2: Original Signal (Blue), on the left vs the Denoised Signal (Red) on the right

4.2 Evaluation Metrics

The standard performance metrics used include precision, accuracy, sensitivity (recall), specificity, and F1 Score [47]. Comparing our results to the conclusive decision by the pulmonologists, I am able to determine

- True Positives (TP), an outcome where our model correctly predicts the positive case.
- True Negatives (TN), an outcome where our model correctly predicts the negative case.
- False Positives (FP), an outcome where our model incorrectly predicts a positive case, when it is really negative.
- False Negatives (FN), an outcome where our model incorrectly predicts a negative case, when it is really positive.

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

Accuracy is the most intuitive measure and is simply a ratio of correctly predicted observation to the total observations. Accuracy works best in symmetric datasets where the number of false positive and false negatives are almost the same [30].

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

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations [30].

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

Sensitivity is the ratio of correctly predicted positive observations to the all observations in that class [30].

$$Specificity = \frac{TN}{TP + FP}$$

Specificity measures the model's ability to predict true negatives [30].

$$F1\ Score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$$

F1 score is the weighted average of Precision and Sensitivity, taking into account both false positives and false negatives. The F1 score is often more useful than accuracy, especially in an uneven class distribution, as is in the distribution used here [28].

4.3 Classification

In considering what the best algorithm combination would be, several combinations were tried as described in the Methods section. The final classification selected and used was evaluated with the evaluation metrics named above, paying more attention to the F1 Score.

4.4 Evaluation Results

The results above in Table 4.1 indicate fairly accurate results. There were five more misclassifications in the case of the healthy cases, those with no abnormal lung sounds and eight fewer misclassifications for the non-healthy cases, those with abnormal lung sounds of any kind.

Total No. of Subjects	296
True Positives (TP)	81
False Positives (FP)	9
True Negatives (TN)	63
False Negatives (FN)	12
Accuracy	87.27%
Precision	92.28%
Sensitivity	87.10%
Specificity	87.50%
F1 Score	0.886

Table 4.1: Performance Metric Results

Of the data points used, there was some statistical overlap with the highest degree of separation between healthy and unhealthy recordings. These are were selected as possible inputs in a neural network for automated categorizing of the data into its healthy and unhealthy groups. Network optimization included a cross-validation approach across four different training algorithms of which final network error evaluation proved the top fourteen input measurements needed to be used.

The final neural network system used comprised of twelve different input parameters and yielded an overall accuracy of 87.27% and an F1 Score of 0.89.

Chapter 5

Conclusion

Since the development of the traditional stethoscope we have made significant developments that can capture, record, study and recognize auscultations beyond using the a physician's judgement [9]. These new capabilities can be a transformative aid to medical practitioners. The sounds can be stored as a part of the patient's medical history. Over time, this raw data combined with the consultation notes and medical images can provide a more complete picture of where the patient was from one visit to the next as well as how a pathology is progressing or regressing [45].

Tech is entering the healthcare space at a rapid pace and transforming the job of physicians. Some of their tasks will be taken over by AI, leaving them to have more time to work with patients with care and patience [3]. Physicians will have more access and ease of knowledge of up-to-date information in medical research. They will likely have less administrative tasks and note taking lags. Computers and sensors are developing beyond the capabilities of the human ear and can captures and analyse a larger range of data [45]. With artificial intelligence and more technology integrated in the medical system and record keeping, we will have better and better understanding of the nuances and subtleties of pulmonary auscultations which will lead to new clinical discoveries. Furthermore, these tools are becoming more portable and able to do more real-time data collection at the bedside with the integration of the internet of things (IoT) devices [69].

Technological advances in diagnostics have caused some clinicians to question

the usefulness of more traditional, subjective, holistic examinations of diagnosis [52]. Others argue that the physical examinations remains an important aspect of medical care and requires training [18]. Clinicians who are skilled at the bedside examination make better use of diagnostic tests and order fewer unnecessary tests. Even in cardiac, where auscultation is considered to be a central player in examination, despite its usefulness and value as a low-cost diagnostic irreplaceable tool, it too is quick becoming a lost art. When used properly, the stethoscope are a valuable and cost-effective clinical tool that a well-trained provider can use to make a rapid and accurate diagnosis with fewer additional tests [14].

5.1 Comparison to Other Work

A few other groups have used this dataset to also work on the classification problem as presented here. Their work and results can be found on Kaggle.com. The resulting F1 scores were lower than what was found in this attempt, ranging from 0.6-0.75. This is likely because of the different methods used in this approach that were different from what was tried before. I have not been able to directly compare our results to previous approaches since the source code and datasets used are not available. Our approach and results differ in a few distinct ways from reviewing the methods used in other attempts. I included data from healthy patients that were from the general population. Prior studies data was from patients in hospitals and thus had far fewer controls. Looking at the results on on specificity and sensitivity, these numbers ranged widely from under 30% to close to 100%. This is often what happens when there is overfitting and an imbalance in the training and test dataset. I modified the methods to avoid overfitting and set aside a training dataset that was different but a good control for our algorithm.

The methods used consisted of simulated crackles superimposed on real breathing sounds. The results indicated that the most significant detection errors are owed to the following two reasons. One, intensity or strength of of the respiratory signal. Deep breathing masks more crackles than superficial breathing. Second, the type of

crackles found. Fine crackles are easily recognizable as their waveform differs more significantly from that of classic lung sounds. The amplitude of the crackles also changes how difficult they are to distinguish. Crackles are clearly the most difficult type of lung sound to be able to be detected and distinguished accurately. Thus, I do not recommend that auscultation alone should not be the sole reference for validating crackle detection.

5.2 Future Development

Electronic auscultation is still rather nascent and requires a lot more acceptance and use in the hospital systems [44]. It might still take quite some time before digital auscultation tools replace a clinician and his stethoscope. The space for new, innovative, portable and affordable diagnostic device that aid patients towards pulmonary health and wellness will likely push the development further of this work. Telemedicine, and its need especially in rural places, developing nations or even during a pandemic might also drive up the demands for such devices [11]. For example, tuberculosis (TB), not included in this thesis, is a bacterial infection of the lungs that can cause a range of symptoms, including chest pain, breathlessness, and severe coughing. It is life-threatening without treatment and active patients spread the bacteria through the air. Many infected individuals with TB bacteria go through a latent period where do not feel sick or experience any symptoms but can spread the disease. The World Health Organization (WHO) has indicated this is a attention worthy disease [40]. TB cannot be currently be diagnosed through auscultation, but it is plausible that new advances and instrumentation can detect the unique pulmonary pattern and metrics of TB in the future.

THIS PAGE INTENTIONALLY LEFT BLANK

References

[1] Asith Abeysinghe, Mohammad Fard, Reza Jazar, Fabio Zambetta, and John Davy. Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise. *The Journal of the Acoustical Society of America*, 150(1):193–201, 2021.

[2] Kaveh Ahookhosh, Oveis Pourmehran, Habib Aminfar, Mousa Mohammadpourfard, Mohammad Mohsen Sarafraz, and Hamed Hamishehkar. Development of human respiratory airway models: A review. *European Journal of Pharmaceutical Sciences*, 145:105233, 2020.

[3] Abhimanyu S Ahuja. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, 7:e7702, 2019.

[4] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your maml. *arXiv preprint arXiv:1810.09502*, 2018.

[5] Juan Carlos Aviles-Solis, Cristina Jacome, A Davidsen, R Einarsen, Sophie Vanbelle, Hans Pasterkamp, and Hasse Melbye. Prevalence and clinical associations of wheezes and crackles in the general population: the tromsø study. *BMC pulmonary medicine*, 19(1):1–11, 2019.

[6] Imon Banerjee, Yuan Ling, Matthew C Chen, Sadid A Hasan, Curtis P Langlotz, Nathaniel Moradzadeh, Brian Chapman, Timothy Amrhein, David Mong, Daniel L Rubin, et al. Comparative effectiveness of convolutional neural network (cnn) and recurrent neural network (rnn) architectures for radiology text report classification. *Artificial intelligence in medicine*, 97:79–88, 2019.

[7] Rolf Behling and Florian Grüner. Diagnostic x-ray sources—present and future. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 878:50–57, 2018.

[8] Yougang Bian, Yang Zheng, Shengbo Eben Li, Qing Xu, Jianqiang Wang, and Keqiang Li. Behavioral cooperation of multiple connected vehicles with directed acyclic interactions using feedforward-feedback control. In *14th international symposium on advanced vehicle control (AVEC), Beijing, China*, pages 16–20, 2018.

[9] Karin Bijsterveld. *Sonic Skills: Listening for Knowledge in Science, Medicine and Engineering (1920s-Present)*. Springer Nature, 2019.

[10] PJ Bishop. Evolution of the stethoscope. *Journal of the Royal Society of Medicine*, 73(6):448–456, 1980.

[11] Joaquin A Blaya, Hamish SF Fraser, and Brian Holt. E-health technologies show promise in developing countries. *Health Affairs*, 29(2):244–251, 2010.

[12] Léon Bottou. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages 421–436. Springer, 2012.

[13] Michael Cao, Roy S Gardner, Ramesh Hariharan, Devi G Nair, Christopher Schulze, Qi An, Pramodsingh H Thakur, Brian Kwan, Yi Zhang, and John P Boehmer. Ambulatory monitoring of heart sounds via an implanted device is superior to auscultation for prediction of heart failure events. *Journal of cardiac failure*, 26(2):151–159, 2020.

[14] Michael A Chizner. Cardiac auscultation: rediscovering the lost art. *Current problems in cardiology*, 33(7):326–408, 2008.

[15] Young-Bok Cho. Keras based cnn model for disease extraction in ultrasound image. *Journal of Digital Contents Society*, 19(10):1975–1980, 2018.

[16] Ryan J Delahanty, JoAnn Alvarez, Lisa M Flynn, Robert L Sherwin, and Spencer S Jones. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Annals of emergency medicine*, 73(4):334–344, 2019.

[17] William Diprose and Nicholas Buist. Artificial intelligence in medicine: humans need not apply? *The New Zealand Medical Journal (Online)*, 129(1434):73, 2016.

[18] Avedis Donabedian. Evaluating the quality of medical care. *The Milbank Quarterly*, 83(4):691, 2005.

[19] Konstantinos Douros, Vasilis Grammeniatis, and Ioanna Loukou. Crackles and other lung sounds. *Breath Sounds*, pages 225–236, 2018.

[20] TA Dronova, CC Aires, SA Oliveira, and JL Silas. Problems of terminology differences of auscultatory sounds for correct diagnosis in pulmonology. . . , page 440, 2015.

[21] Mounya Elhilali and James E West. The stethoscope gets smart: Engineers from johns hopkins are giving the humble stethoscope an ai upgrade. *IEEE Spectrum*, 56(2):36–41, 2019.

[22] Guo-Feng Fan, Yan-Hui Guo, Jia-Mei Zheng, and Wei-Chiang Hong. Application of the weighted k-nearest neighbor algorithm for short-term load forecasting. *Energies*, 12(5):916, 2019.

[23] Pedro Faustino, Jorge Oliveira, and Miguel Coimbra. Crackle and wheeze detection in lung sound signals using convolutional neural networks. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 345–348. IEEE, 2021.

[24] Christopher A Feddock. The lost art of clinical skills. *The American journal of medicine*, 120(4):374–378, 2007.

[25] Sotirios Fouzas, Michael B Anthracopoulos, and Abraham Bohadana. Clinical usefulness of breath sounds. In *Breath Sounds*, pages 33–52. Springer, 2018.

[26] Noam Gavriely and David W Cugell. *Breath sounds methodology*. cRc PrEss, 2019.

[27] Renata E Howland, Nicholas R Cowan, Scarlett S Wang, Mitchell L Moss, and Sherry Glied. Public transportation and transmission of viral respiratory disease: Evidence from influenza deaths in 121 cities in the united states. *PloS one*, 15(12):e0242990, 2020.

[28] Hao Huang, Haihua Xu, Xianhui Wang, and Wushour Silamu. Maximum f1-score discriminative training criterion for automatic mispronunciation detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4):787–797, 2015.

[29] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE, 2020.

[30] Brendan Juba and Hai S Le. Precision-recall versus accuracy and the role of large data sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4039–4048, 2019.

[31] Saeed Khaki, Hieu Pham, Ye Han, Andy Kuhl, Wade Kent, and Lizhi Wang. Convolutional neural networks for image-based corn kernel detection and counting. *Sensors*, 20(9):2721, 2020.

[32] Aditya Khamparia, Deepak Gupta, Nhu Gia Nguyen, Ashish Khanna, Babita Pandey, and Prayag Tiwari. Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access*, 7:7717–7727, 2019.

[33] Yoonjoo Kim, YunKyong Hyon, Sung Soo Jung, Sunju Lee, Geon Yoo, Chaeuk Chung, and Taeyoung Ha. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific Reports*, 11(1):1–11, 2021.

[34] Robert Lethbridge and Mark L Everard. The stethoscope: historical considerations. In *Breath Sounds*, pages 15–31. Springer, 2018.

[35] Michael G Levitzky. *Pulmonary physiology*. Number 1. : McGraw-Hill Education„ 2018.

[36] Nicos Maglaveras. *International Conference on Biomedical and Health Informatics: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017*. Springer, 2017.

[37] N Abdul Malik, W Idris, TS Gunawan, RF Olanrewaju, and S Noorjannah Ibrahim. Classification of normal and crackles respiratory sounds into healthy and lung cancer groups. *International Journal of Electrical and Computer Engineering*, 8(3):1530, 2018.

[38] Salvatore Mangione and Linda Z Nieman. Pulmonary auscultatory skills during training in internal medicine and family practice. *American journal of respiratory and critical care medicine*, 159(4):1119–1124, 1999.

[39] Jash Mehta, Deep Gandhi, Govind Thakur, and Pratik Kanani. Music genre classification using transfer learning on log-based mel spectrogram. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1101–1107. IEEE, 2021.

[40] Jaime Melendez, Clara I Sánchez, Rick HHM Philipsen, Pragnya Maduskar, Rodney Dawson, Grant Theron, Keertan Dheda, and Bram Van Ginneken. An automated tuberculosis screening strategy combining x-ray-based computer-aided detection and clinical information. *Scientific reports*, 6(1):1–8, 2016.

[41] Reza Moradi, Reza Berangi, and Behrouz Minaei. A survey of regularization strategies for deep models. *Artificial Intelligence Review*, 53(6):3947–3986, 2020.

[42] Onofre Moran-Mendoza, Thomas Ritchie, and Sharina Aldhaheri. Fine crackles on chest auscultation in the early diagnosis of idiopathic pulmonary fibrosis: a prospective cohort study. *BMJ open respiratory research*, 8(1):e000815, 2021.

[43] Bekhzod Olimov, Sanjar Karshiev, Eungyeong Jang, Sadia Din, Anand Paul, and Jeonghong Kim. Weight initialization based-rectified linear unit activation function to improve the performance of a convolutional neural network model. *Concurrency and Computation: Practice and Experience*, 33(22):e6143, 2021.

[44] Ana L Padilla-Ortiz and David Ibarra. Lung and heart sounds analysis: state-of-the-art and future trends. *Critical Reviews™ in Biomedical Engineering*, 46(1), 2018.

[45] Hüseyin Polat and İnan Güler. A simple computer-based measurement and analysis system of pulmonary auscultation sounds. *Journal of medical systems*, 28(6):665–672, 2004.

[46] Olga Ponomareva, Alexey Ponomarev, and Vladimir Ponomarev. Evolution of forward and inverse discrete fourier transform. In *2018 IEEE East-West Design & Test Symposium (EWDTS)*, pages 1–5. IEEE, 2018.

[47] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.

[48] J Proctor and E Rickards. How to perform chest auscultation and interpret the findings. *Nursing Times*, pages 23–26, 2020.

[49] Rosita Rahma and Jatmika Nurhadi. Can power spectral density (psd) be used to measure reading concentration? 2018.

[50] Kalpana Rangra and Monit Kapoor. Exploring the mel scale features using supervised learning classifiers for emotion classification. *International Journal of Applied Pattern Recognition*, 6(3):232–253, 2021.

[51] Tom Rice. 'the hallmark of a doctor': the stethoscope and the making of medical identity. *Journal of Material Culture*, 15(3):287–301, 2010.

[52] Erich W Russi and Erich W Russi. Lung auscultation-a useless ritual? *Swiss medical weekly*, 135(35-36):513–514, 2005.

[53] Tinir Mohamed Sadi and Raini Hassan. Development of classification methods for wheeze and crackle using mel frequency cepstral coefficient (mfcc): A deep learning approach. *International Journal on Perceptive and Cognitive Computing*, 6(2):107–114, 2020.

[54] Garima Sharma, Kartikeyan Umapathy, and Sridhar Krishnan. Trends in audio signal feature extraction methods. *Applied Acoustics*, 158:107020, 2020.

[55] Lin Shi, Ishtiaq Ahmad, YuJing He, and KyungHi Chang. Hidden markov model based drone sound recognition using mfcc technique in practical noisy environments. *Journal of Communications and Networks*, 20(5):509–518, 2018.

[56] Kristina P Sinaga and Miin-Shen Yang. Unsupervised k-means clustering algorithm. *IEEE Access*, 8:80716–80727, 2020.

[57] Hugo Siqueira and Ivette Luna. Performance comparison of feedforward neural networks applied to streamflow series forecasting. *Mathematics in Engineering, Science & Aerospace (MESA)*, 10(1), 2019.

[58] Joan B Soriano, Parkes J Kendrick, Katherine R Paulson, Vinay Gupta, Elissa M Abrams, Rufus Adesoji Adedoyin, Tara Ballav Adhikari, Shailesh M Advani, Anurag Agrawal, Elham Ahmadian, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet Respiratory Medicine*, 8(6):585–596, 2020.

[59] Sankararaman Sreejyothi, Ammini Renjini, Vimal Raj, Mohanachandran Nair Sindhu Swapna, and Sankaranarayana Iyer Sankararaman. Unwrapping the phase portrait features of adventitious crackle for auscultation and classification: a machine learning approach. *Journal of Biological Physics*, pages 1–13, 2021.

[60] Thomas Sühn, Moritz Spiller, Rutuja Salvi, Stefan Hellwig, Axel Boese, Alfredo Illanes, and Michael Friebe. Auscultation system for acquisition of vascular sounds–towards sound-based monitoring of the carotid artery. *Medical Devices (Auckland, NZ)*, 13:349, 2020.

[61] G Ganesh Sundarkumar and Vadlamani Ravi. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence*, 37:368–377, 2015.

[62] M Talaat, MA Farahat, Noura Mansour, and AY Hatata. Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. *Energy*, 196:117087, 2020.

[63] Abraham Verghese and Ralph I Horwitz. In praise of the physical examination, 2009.

[64] Aparna Vyas, Soohwan Yu, and Joonki Paik. Fundamentals of digital image processing. In *Multiscale Transforms with Application to Image Processing*, pages 3–11. Springer, 2018.

[65] Xueqiong Wei and Yuanjun Wang. Inferior alveolar canal segmentation based on cone-beam computed tomography. *Medical Physics*, 48(11):7074–7088, 2021.

[66] Fred Weinberg. The history of the stethoscope. *Canadian Family Physician*, 39:2223, 1993.

[67] Trevor S Wiens, Brenda C Dale, Mark S Boyce, and G Peter Kershaw. Three way k-fold cross-validation of resource selection functions. *Ecological Modelling*, 212(3-4):244–255, 2008.

[68] Yu-Sheng Wu, Chia-Hung Liao, and Shyan-Ming Yuan. Automatic auscultation classification of abnormal lung sounds in critical patients through deep learning models. In *202020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, pages 9–11. IEEE, 2020.

[69] Ravi Teja Yarlagadda. Internet of things & artificial intelligence in modern society. *International Journal of Creative Research Thoughts (IJCRT), ISSN*, pages 2320–2882, 2018.

[70] EG Zaitseva, MV Chernetsky, and NA Shevel. About possibility of remote diagnostics of the respiratory system by auscultation. , 11(2), 2020.

[71] Barret Zimmerman and Donna Williams. Lung sounds. *StatPearls [Internet]*, 2020.