Automated Classification of Lung Sounds via Supervised Learning

Master's thesis

Valerio Radicchi
2110310014
07/06/2023

Performed at the
Fachhochschule Kärnten/Carinthia University of Applied Sciences

Degree Program Health Care IT

Supervised by

FH-Prof. Priv. Doz. Dr.

Günther Grabner

Author Radicchi, Valerio	Type of Publication Master’s Thesis	Date June 2023
	Number of pages 71	Language of publication English

Automated Classification of Lung Sounds via Supervised Learning

Healthcare digitalization and the employment of AI solutions are both new concepts that are currently developing and improving along with the always-increasing demands of the industry. The proper use of technology within healthcare bears the responsibility to improve human existence and potentially even save human lives by overcoming human limits and opening up further treatment possibilities.

In the specific field of automated pulmonary diagnosis, several researchers have already attempted the evaluation of different classification techniques. However, most of the studies, even if successful, fail to provide reproducibility of the results. In addition, there is a limited amount of publicly retrievable data for this type of analysis and there is not still a predefined standard on the acquisition protocols.

This study attempted to reproduce some of the theoretically available parts of two already successful partially reproducible kinds of research, namely Automated Lung Sound Analysis of Grønnesby (2016) and Improving the ResNet-based Respiratory Sound Classification Systems with Focal Loss of Li, Wang et al. (2022), and utilizes the best available data provided by the SPRSound: Open-Source SJTU Paediatric Respiratory Sound Database. Specifically, this study executes a pulmonary event approach by selecting predetermined time windows for the evaluation of the binary classification performance and addresses the class imbalance problem by randomly removing the excessive number of the predominant healthy class until the same number of samples is reached for both classes. The evaluation metrics (F1, AUC & PRC) are computed for each of the three deployed algorithms (Decision Tree, Deep Neural Network, and Support Vector Machine).

The results of this study suggest that pulmonary events should be evaluated on length-based principles, since, confronting signals who have too much diversity in their time length might mislead the machine learning classification algorithm. Also, this study suggests that the labeling of the most evasive adventitious events should be conducted in future databases with as much precision as possible to avoid the problem of potentially overlapping classes. However, the performance for the less evasive classes has been more than satisfactory but also a higher number of samples for the minority adventitious classes should be employed in the future to ensure the further reliability of the used methodologies.

healthcare, IT, machine learning, pulmonary, lung, support vector machine, neural network, supervised learning

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The detection of pulmonary diseases has been an ongoing challenge in the medical field for many years. As the third biggest cause of death worldwide, respiratory disorders have a huge negative impact not only on people's health, but also on the economy, and society of the burden they bring on public health systems. The development of new technologies and advances in machine learning have presented a potential solution to this problem as an optimal way to improve early diagnosis and routine monitoring of patients with respiratory disorders [1–6]. Machine learning classification for audio input is a promising approach to the automatic detection of pulmonary diseases, therefore, this research’s aim is to establish the reliability of the latter by using audio input machine learning classification for the automatic detection of pulmonary diseases.

One of the earliest methods for making a diagnosis of various pulmonary disorders is the auscultation of the respiratory system. Because it is affordable, noninvasive, safe, and simple for medical personnel to perform, it is widely used. One of the earliest diagnostic tools in medicine was the stethoscope. Its main contribution to medicine has been in establishing the basis for the important diagnostic tool for recognizing and monitoring various lung illnesses, the auscultation of the respiratory system [2,6–10].

Machine learning is one of the branches of artificial intelligence. As the name declares, it allows machines to learn from data and, therefore, enables us to identify patterns within copious amounts of information. Nowadays, as a well-established technology, it has been applied in many areas including the case of this study: the automatic detection of diseases [11]. Specifically, audio input machine learning classification is a form of machine learning that processes audio signals to detect patterns and, later, classify them into groups. [12]. Signs and symptoms are typically used to quantify and identify diseases, where a symptom is a specific sign of the patient's ailment, whereas a sign is an objective manifestation of a disease that specialists can describe. As a result, every disease has a unique set of symptoms, such fever, which can occur in a wide range of disorders but not necessarily in all of them at the same time [13].

Specifically, pulmonary diseases are a group of medical conditions that affect the lungs and the respiratory system. Consequentially, these diseases can cause difficulty in breathing, coughing, and wheezing. The early detection of these diseases is fundamental for the treatment and management of the general health condition. However, diagnosing pulmonary diseases is often difficult due to the complexity of the respiratory system and the lack of specific diagnostic tests [14]. While it is very simple to identify a single cough event, it is still challenging to evaluate cough frequency over a lengthy period for clinical and research purposes. The fact that coughing is paroxysmal, or in a less medical term “recurrent”, necessitates lengthy recording times to produce an accurate estimate and analyse its frequency, [15] therefore, this study has dedicated its focus toward the analysis of regular respiratory events.

One very valid example of morbidity that affects the pulmonary system is tuberculosis which, despite the enhanced efficacy of treatment and preventive measures, it still is the major cause of adult respiratory disease mortality in the developing world and kills many more people than it should. Other acute and chronic adult respiratory disorders, however, are becoming more prevalent globally. These diseases fall into four categories: acute diseases, such as pneumonia and influenza; chronic diseases, such as chronic obstructive pulmonary disease (COPD) and asthma; occupational lung diseases, such as byssinosis, asbestosis, and coal worker's pneumoconiosis; and other parenchymal lung diseases, such as immune-related lung diseases [14].

Fortunately, through the developments in digital healthcare technologies like artificial intelligence (AI), 3D printing, robotics, and nanotechnology, healthcare is still evolving right before our eyes. This opportunity is known as healthcare digitization and it offers a variety of chances like lowering human error rates, enhancing clinical results, or monitoring data over time. Even more recently, an increasing number of health-related sectors became heavily dependent on AI techniques such as machine learning or deep learning, for developing new diagnostic systems, storing and processing patient information, and even treating various ailments [13,16]. Within this study scope, the potential of using electronic stethoscope audio recordings for the classification of adventitious sounds is remarkably promising. With the help of machine learning algorithms, sound recordings can be processed and analyzed for identifying and classifying adventitious sounds. [3,17].

These “source signals” are sounds. Soft tissue, hard tissue (bone), and air are the three pathways by which physiological noises are transmitted [18]. Precedent research established that breathing sounds possess a long and stable duration, but without any related harmonics [15], meanwhile, adventitious sounds are extra respiratory noises that are added on top of the regular breathing sounds. A more straightforward division of adventitious lung sounds that splits them into two basic categories (continuous sounds and interrupted sounds) was proposed in 1957. High- and low-pitched wheezes were used to further categorize continuous noises, while coarse, medium, and fine crackles were used to further categorize interrupted sounds. The terminology was further streamlined by the International Lung Sound Association in 1976. Discontinuous sounds were divided into fine and coarse crackles, while continuous sounds were divided into wheezes and rhonchi. The requirement for continuous accidental sounds is to last for more than 250 milliseconds. However, within the various scientific references, the details concerning the classification of the possible adventitious sounds are somehow subject to inter-variability. In accordance to the American Thoracic Society (ATS) Committee on Pulmonary Nomenclature, **wheezes** are high-pitched continuous noises with a prominent frequency of 400 Hz or more, while rhonchi are low-pitched continuous musical sounds with a dominating frequency of roughly 200 Hz or less [19]. The following image illustrates how can pathologies reflect spectrographically, where on left the healthy signal shows a lower concentration of frequencies in comparison with the adventitious one on the right.

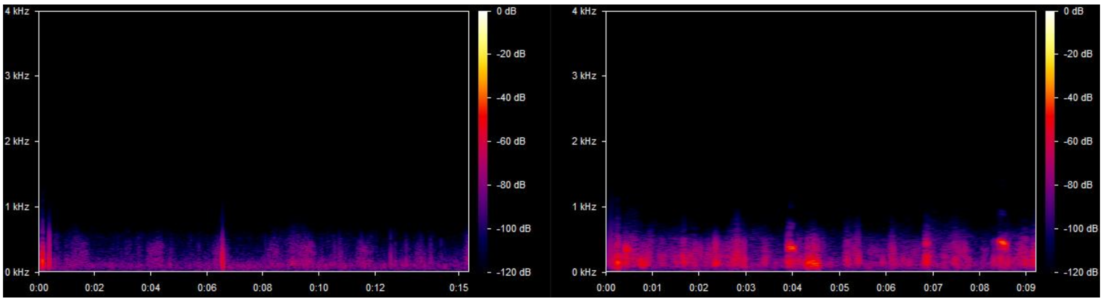

Figure 1. A spectral comparison of a “healthy” recording (left) and an “adventitious” one (right). The left signal shows a lesser concentration of frequencies in comparison to the right one, a typical behavior for both classes. Khz on the left axis, Time on the bottom axis, and decibels on the right axis.

Regardless, the general frequency range of the lung wheeze sound is estimated to be from 60 to 2,000 Hz. Meanwhile, some studies focus on the contribution of wheezing within a narrower frequency range, 600–2,000 Hz [7]. Despite the fact that, in accordance to the ATS definition of continuous sound, wheezes should last for a length of more than 250ms [1,7,19], their duration de facto has been monitored to not exceed by much the 80–100ms.

Interestingly, wheezes are typically detectable at the patient's open mouth or by auscultation across the trachea, and they are occasionally audible at a distance from the patient. They are usually louder than the underlying breath sounds [19]. In fact, wheezes are believed to be produced when air flows through airways narrowed by secretions, foreign bodies, obstructive lesions, or more generally COPD [1,7].

However, whilst wheezes can occur in a variety of situations, they are not always associated with asthma or asthma-related conditions. Scientific evidence pointed out that even in asthma patients with significant airway blockage, the wheezing sound characteristics may be completely absent. This is because there must be enough airflow for the wheezing sound to be audible. Meanwhile, the respiratory flows in acute and severe asthma cases drop basically to zero, therefore, they are unable to supply the energy needed to cause wheezes (or any sound at all). This ailment is medically known as "silent chest" and, in total respect to its nomenclature, may not cause any detectable wheezing sound. However, the relief from the pathway obstruction would consequentially increase the airflow, leading to the restoration of wheezing and, therefore, normal breathing sounds. Meanwhile, patients with asthma and chronic obstructive pulmonary disorders (COPD) usually experience extensive airway blockage, which might cause widespread wheezing within the chest. Also, focused wheezing might result from localized airway blockage brought on by a tumor, foreign body, or even a simpler mucus plug. Nevertheless, even a healthy person might experience wheezing toward the end of their expiration, especially after forcing it, like during intense physical activity. In conclusion, wheezing sounds are a nonspecific finding since even a healthy mild expiratory action could cause an apparently pathological wheeze sound to occur [9,19,20].

On the other hand, rhonchi are another typology of adventitious sounds that are best heard when the stethoscope is located above the chest wall due to their low tone. Their regular pitch tone is around 150 Hz but shares the same shape of waveform with the wheezes, which is considered a defining feature. Whilst the wheeze is typically an expiratory condition, it can also be inspiratory or even present in both phases of respiration. Specifically, inspiratory wheezes may be related to severe intrathoracic lower airway blockage or upper airway obstruction [19]. Also, healthy lung sounds emanating during inspiration and expiration have a major power spectrum in the frequency range below 100 Hz. That spectrum decreases on frequencies above 100 Hz, but remains still detectable up to 2,000 Hz [7].

Crackles are short explosive clicking or crackling noises that are produced due to the opening of the small airways. Due to their distinctive sound, they are often also called Rales or Crepitations. They are regarded as a short time event since their usual window spans between 5ms and 40ms. Naturally discontinuous and brief, they tend to be more detectable within the inspiration phase. Substantially, they are classified as either fine which is short and high pitched, or coarse that are longer and louder [7,20,21]. Furthermore, their attributes (loudness, pitch, duration, and timing) are susceptible to change since the act of coughing or even a shift in the patient's position will most likely alter the location of the mucus within the airways. The frequency range of the crackle sounds falls within the same range as the wheezing, with the major contribution being in the range of 60 Hz to 1,200 Hz [7]. Like wheezes, crackles can also be present within healthy lungs, however, an eventually continued occurrence might suggest a potential obstruction within the airways. This might be due either to an excessive amount of fluid/mucus within the airways or just a general lack of ventilation within the expiration phase. Crackles are generally prominent in patients who suffer from pneumonia, pulmonary fibrosis, and acute bronchitis. It is also important to denote that, despite already their elusive nature, crackles might be even more complicated to detect by machines since their distinctive sound could be easily reproduced by the rubbing of the stethoscope's microphone against fabric or even just chest hair [20]. Usually, the chances for a diagnosis of impaired lung functionality significantly raises when inspiratory crackles are present at two or more locations, or even the copresence of wheezes might provide an additional clue. Naturally, when such findings are present in a patient, and especially if unexpected, the immediate procedure should be to prompt additional investigation toward potential heart or lung conditions. Additionally, also age has been noted as the most significant predictor of adventitious sounds, notably crackles. [9]. Regardless of all the provided details, it is of uttermost importance within the field of respiratory sound analysis, to know how to differentiate successfully between both potentially healthy and adventitious lung sounds [7].

Indubitably, machine learning could prove to be a potential help for doctors if it could reliably differentiate between healthy and adventitious sounds and, consequentially, assist them in determining the health status of their patients [17,20,22]. Also, algorithms could further help in providing personalized care for patients since the recordings could be used to identify morbidities and monitor the progression of the condition. Even treatments could be monitored so that their effectiveness could be entirely documented and even potentially predicted. The final aim should be to deploy reliable machine learning models that could successfully classify adventitious sounds from stethoscope audio recordings with the hope that this would improve the accuracy of the conventional diagnosis methodologies by providing quick insights into the patient's condition[17].

Recently, machine learning (ML) algorithms have shown promising results in the detection of pulmonary diseases [1,20]. However, due to many factors, the reliability of these algorithms is still undetermined, leaving the medical community uncertain of their ultimate efficacy [8,13]. Due to the innate relationship between anatomy, condition and location, lung sounds classification is a tricky task. Firstly, the signal changes depending on the recording site, flow rate, lung volume, body position, and different breathing techniques are all elements that must be considered when evaluating the quality of a recording. With aging, due to cell maturation, and minor environmental disturbances such as past traumas or scarring, the sound might also be subject to change from what could be expected from a younger patient. The signal itself is so intricate and variable that it occasionally might even appear random or completely unpredictable in its behavior. However, it is more likely that the underlying anatomy and pathophysiology are reflected within the acoustic signal rather than simple chaos as justification [20]. However, it is also important to mention that despite how tempting would be the idea of furthering a complete and reliable full machine learning “diagnosis” based just on the lung sounds, its impossibility is also remarked by the fact that the adventitious sounds are not individually and exclusively correlated to specific pathologies. For example, there is no detectable difference between COVID-19 pneumonia and other types of pneumonia. Thus, this limitation highlights, even more, how fundamental it is to possess complete patient information to improve the reliability of any medical algorithm [4,8].

Perhaps, today we live on the verge of a new important stepstone for the healthcare sector, maybe even as important as the key advances that were Pasteur's germ theory of disease, Roentgen's discovery of X-rays, and Chadwick's correlation of infection and poverty [5]. Thus, this research seeks to establish the reliability of ML classification for the detection of pulmonary diseases based on audio inputs. The findings will either be used to identify the most reliable algorithms or to ascertain the inherent eventual external technical limitations of the latter so that further development would still be able to benefit from whatever accurately documented outcome. The AI commune must incorporate active best practices of principled inclusion, software growth, implementation science, and individual workstation interaction into a comprehensive best practice method for execution and safety. Since the potential ability of AI applications to improve patient outcomes is indeed remarkable [13], the aim of this research is to provide recommendations for the effective implementation of ML classification within clinical settings.

1.1 Research Question

This study tries to reveal if, Artificial Intelligence, can establish a reliable pulmonary sound classification based on digital stethoscope recordings.

Therefore, it will attempt to establish the reliability of machine learning classification for automatic lung sounds detection and determine the ideal technical requirements for this operation to ensure maximum performance in terms of precision and safety.

Additionally, since several approaches to pulmonary recordings automated classifications have been already attempted by other studies, it becomes also the aim of this study to define what could be the potential standards for future attempts by outlining the requirements that arose from this study's actual findings, in an ethical and reproducible way.

1.2 Basics

The following section will analyze in detail the techniques that have been chosen to answer the research the aim of this study, specifically: to establish the reliability of audio input machine learning classification for the automatic detection of pulmonary diseases. Within this study, two different approaches have been pursued by the researcher in the attempt to obtain the most reliable results possible. The following image illustrates different approaches toward audio signal analysis. While the segmentation approach divides the signals into windows of equal length, the record processes the signals by their complete length or a set duration. The event only studies the portions of the signal where the relevant information is occurring, stripping away all the "unnecessary" information.

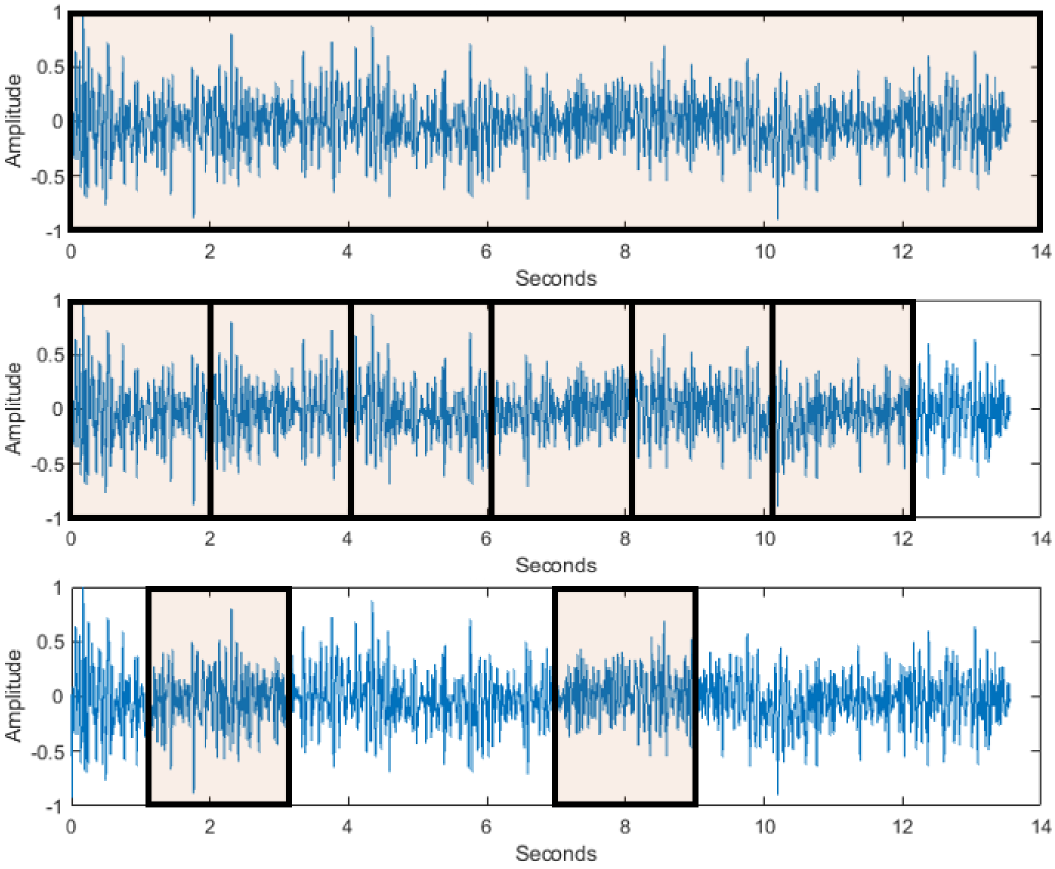

Figure 2. A visual comparison for the different audio analysis techniques in time domain. Top: Record analysis, Mid: Segment analysis, and Bottom: Event Analysis. Whilst the record processes the signals by their entire length or a fixed duration, the segmentation approach divides them in windows of equal length, and the event just analyses those parts of the signal where the relevant information is occurring cutting away all the “unnecessary” information.

Different approaches could generally be attempted when conducting respiratory analysis, namely one study could decide to focus on the whole record, whilst others might dedicate their attention to segments [20], or even single events [23]. Namely, since breathing events are usually followed by long prosodic breaks (pauses) [24], focusing on the events promotes the idea to remove the “unlabeled” segments between the breathing events and furtherly facilitates the work for the machine learning process. Purely for clarity, this section will only address the background of the methodologies directly related to the results that will be presented in the later parts of this study.

1.3 Preprocessing

Data preprocessing is the key component of any analytical process. However, despite receiving less attention than other steps like data mining, it frequently accounts for more than 50% of the entire effort put out during data analysis. The existence of missing values, noise, inconsistencies, and redundancies—all of which affect how subsequent learning processes perform—are typical characteristics of raw data. To reduce the impact of data anomalies (if any) on the effectiveness (quality and reliability) of following processing steps, a suitable preprocessing step is often necessarily carried out [25]. Specifically for this research, the preparatory part of the data has proved fundamental. Since the complex nature of biomedical signals, the preprocessing requirements for studies such as the latter have proved to be substantially higher than the conventional tabular analysis or data science task. Preprocessing could be summarized as the preparatory phase of the data for its later analysis, therefore, without a proper strategy, any further analysis can be potentially compromised by the lack of preparational effort.

1.3.1 Loudness Normalization

Loudness normalization is an acoustic signal procedure that balances the recordings up to a certain value. As a type of normalization, it is often associated with musical recordings and radio/tv streaming services since these scenarios often encounter the necessity to ensure the quality of the recording by ensuring that the different tracks will not be at different volumes [26]. A general approach toward audio normalization is to apply a predetermined constant gain to each audio recording so that, eventually, the amplitudes will match the target level. For example, if you want to reach the EBU R 128 standard of -23 LUFS, and your recordings have a variable loudness of -12/+12, then you will have to apply a specific increase or decrease of gain for each recording depending if their natural loudness is higher or lower than the desired amount [27]. Nowadays, these calculations can be fortunately automated therefore there is no need to calculate manually the gain for each “track”. However, it is also important to mention that the Loudness Normalization differs substantially from the most used Peak Normalization. The main difference is that the Peak Normalization regulates the recordings up to the highest value of the audio data, meanwhile, the Loudness Normalization normalizes the recordings over their general loudness, which includes the short-term loudness, the mid-term loudness, the global loudness, and the peak loudness [28]. The following image visualizes the diversity among the distribution of the loudness for the recordings.

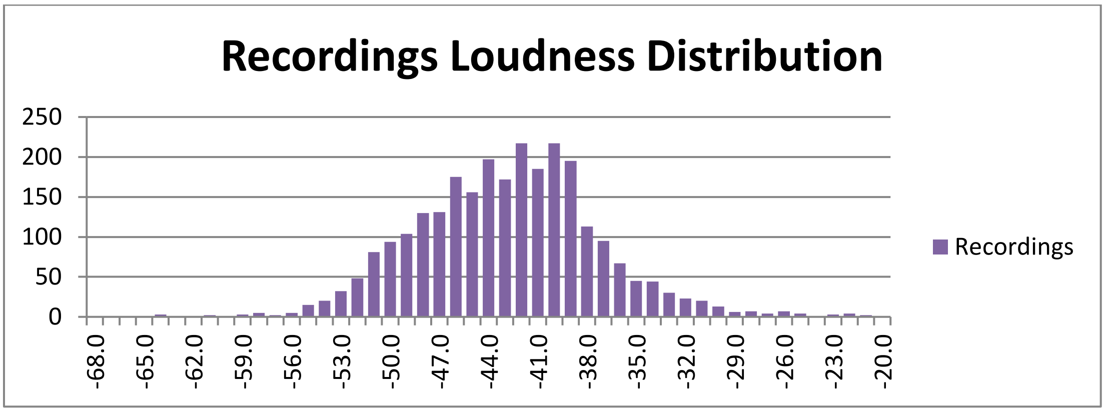

Figure 3. The distribution of the different loudness levels for the whole SJTU Paediatric Database.

Within this study, the loudness is first measured for all the selected samples, therefore, it is subsequentially normalized up to the mean value of those samples. The researcher opted for reaching the mean value rather than a predetermined standard like the EBU R 128 since the authenticity of the files should be preserved as much as possible and normalizing too much might potentially introduce more noise than benefits. Even worse, too different loudness might not converge up to a predetermined value that is not in between them. For example, there are fewer chances for a -74 LUFS recording to reach the value of -23 LUFS than for a recording whose loudness is -35 LUFS, therefore, choosing a value in between them seems to be the better option for “normalizing” them to a certain window.

1.4 Features Extraction

Feature extraction is described as attempting to obtain useful information from the available data. The most common issue is to understand which features might be appropriate for the analysis that one wishes to carry, especially when the available data is intrinsically complex.

Any sub-field of machine learning requires meaningful features to enable accurate processing of the data from the computer side [29]. This part of the study is partially inspired by the 2016 thesis on Automated Lung Sound Analysis of Morten Grønnesby. Namely, the researcher opted to use the same five features (Variance, Range, fine SMA, coarse SMA, and Spectrum Mean), however, 3 more features were included in the attempt to improve both the results and the overall quality of the study. The data, or in this case the extracted events, enter the feature extraction part in the shape of a time-series matrix where all the amplitudes are recorded in accordance with their sampling rate. These amplitudes are divided by each event and the zero padding is tendentially kept at minimum by selecting only events that are of similar length. All the features have been necessarily scaled before entering the classification part, through via standard scaler procedure that removes the mean and scales to unit variance. The following plot is proposed to give an idea about the distribution of the feature vectors that will be theoretically presented shortly.

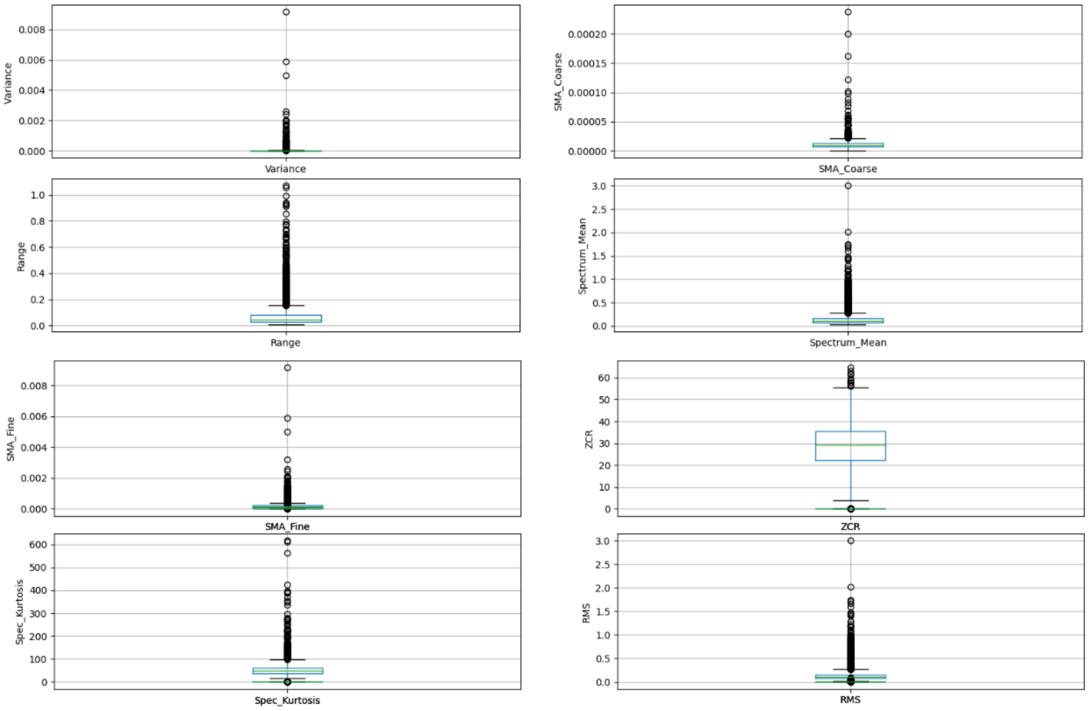

Figure 4. The distribution of the data of the 8 feature vectors before standardization.

1.4.1 Variance

Describable as the measure of the spread of a distribution [20], the variance, along with the standard deviation, is the most used tool to assess the diffusion of the data. Specifically, in this study, the variance is the distribution of the audio amplitudes over time. This means that variance, as a vector, can tell how much a specific is changing over its duration, giving potential hints to the classifier on its nature. As Grønnesby observed, adventitious windows, specifically crackles, tendentially boast a higher variance than healthy ones. To follow, the formula used for calculating the variance of each individual signal:

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

[30]

1.4.2 Range

The range is “just” the highest value of each individual signal subtracted from its minimum value [20]. Within this case, the range vector is maximum amplitude registered within each event. Again, Grønnesby’s theory is that adventitious sounds, due to their explosive nature, might result in an overall higher range value than their healthy counterparts. Therein below, is the formula used for the range of each signal:

$$|R(S) = |Max(S) - Min(S)|$$

[20]

1.4.3 Sum of Simple Moving Average (Coarse)

The sum of simple moving averages (SMA) is a feature that can help to understand how much the data is changing over time. This type of analysis is often conducted within the financial field, however, in this case, it could also help to understand the different fluctuations between healthy and adventitious sounds [20]. The "coarse" SMA is just a general mean of the whole individual signal, therefore, if the mean of the amplitude for the specific events is higher than average, the classifier might understand that there is a difference between the classes. This is resumable with the following formula:

$$SMA_{coarse}(Sig) = \sum_{n=1}^{len(Sig)} |Sig_{n-1} - Sig_n|$$

[20]

1.4.4 Sum of Simple Moving Average (Fine)

The "fine" version of the SMA executes the arithmetical mean of each individual signal but over a window size of 128 samples. This window slides forward with a step function of 16 samples until the signal is completely covered. As the main difference from Grønnesby's research, the difference between sample sizes is given by the divergent sample rates of each database. The Trømso database is declared to be recorded with a sample rate of 44100hz, meanwhile, the SPRSound database sample rate is only 8000hz. This means that the SPRSound files do not possess as many samples as the ones from the Trømso study. This principle is aligned with the Nyquist theorem, which states that to prevent information loss, the sample rate of a recording must be at least twice its frequency [31]. Therefore, it was necessary to restrict the original window size of 800 samples to a smaller window that would have been proportioned to the original proposed one.

This would have led to a window size of 145 samples, but the researcher thought that it would have been better to restrict it to the closer binary number of 128 for the same principle of binary compatibility described within the incipit of this section. Consequentially, also the original sliding function size was 100 samples, it could have been proportioned down to 18 samples, but it has been restrained furtherly to 16 samples only. The following formula resumes what is described so far:

$$SMA_{fine}(Sig) = Max(SMA_{coarse}(win_1), SMA_{coarse}(win_2), \ldots, SMA_{coarse}(win_n))$$

[20]

1.4.5 Spectrum Mean

This value is calculated as the mean value of the Direct Cosine Transform computed on each individual signal. The Direct Cosine Transform is a solution for the problem that arises from using the standard Fast Fourier Transform to get the spectrum of the signal. This is because the resulting numbers from the Fast Fourier Transform are imaginary, therefore, machine learning algorithms struggle to categorize them especially when in combination with integers. Again, Grønnesby's theory states that the adventitious signals should have higher values than the healthy ones also for this feature. To follow the formula used for the calculation of the spectrum:

$$X_k = \frac{1}{2}(x_0 + (-1)^k x_{N-1}) + \sum_{n=1}^{N-2} x_n \cos\left[\frac{\pi}{N-1} n\, k\right] \qquad \text{for } k = 0, \ \ldots\ N-1\,.$$

[32]

1.4.6 Spectral Kurtosis

The analysis of Spectral Kurtosis is popular among the various signal-processing techniques. It is often adopted within the frequency domain to find harmonics, transients, and repetitive impulses. Specifically, the Kurtosis is the fourth central statistical moment divided by the square of the variance. [33]. As a statistical tool, it can identify nonstationary and/or non-Gaussian behavior within the frequency domain. Despite being commonly used to analyze and assess mechanical systems [34], the author of this study believes that it could prove beneficial as an added feature for the assessment of biomechanical systems such as the respiratory apparatus. Specifically, kurtosis is calculated from the already extracted Discrete Cosine Transform spectrum. Interestingly, it still follows the same principle of distribution of the other features, by exerting a much higher standard deviation for the adventitious sounds than the healthy ones. To follow, the formula used for the calculation of the kurtosis of the spectrum:

$$S(t, f) = \int_{-\infty}^{+\infty} x(t)w(t-\tau)e^{-2\pi ft}dt,$$

[34]

1.4.7 Zero-Crossing Rate

The Zero-Crossing Rate (ZCR) is a renowned feature within the field of acoustic signal analysis. It has been extensively used in voice detection tasks, sound classification, optics, biomedical engineering, and even hydraulics and radars. When conventionally adopted on the time domain, its measurements result susceptible to non-stationary noise. It possesses a very low computational cost since it basically tells how many times a signal is "crossing" through zero [35]. In accordance with Grønnesby's theory, the distribution of the Zero Crossing Rate for the healthy signals is noticeably higher than the adventitious ones. This might be explained by the fact that normal sounds tendentially feature lower amplitudes, therefore their "crossing" toward the zero happens more frequently than the adventitious ones since their amplitudes prevent them from reaching the zero in the same amount of time. To follow, the formula for the calculation of the ZCR:

$$\mathrm{ZCR(n)} = \frac{1}{2\mathrm{N}}\Sigma_{\mathrm{m}=-\infty}^{\infty}|\mathrm{sgn}[\mathrm{s(m)}] - \mathrm{sgn}[\mathrm{s(m-1)}]|\mathrm{w(n-m)}$$

[35]

1.4.8 Root Mean Square

The Root Mean Square (RMS) is a mathematical operation commonly used in energetic engineering. It is often employed successfully for evaluating short term signals. As a statistical tool, it relates for its similarity to the standard deviation [36]. However, the difference between the RMS and the Standard Deviation is in the fact that with RMS the data points are squared, whilst with the standard deviation the difference between each data point is squared along with the mean [37]. Given the short, timed nature of breathing cycles, the author of this research believes that the RMS could provide additional hints to the machine for the classification task. However, despite featuring a higher standard deviation for the healthy sounds, the maximum value for the healthy sounds is still tendentially lower than the adventitious ones. To follow, the basic formula for the calculation of the RMS:

$$x_{\mathrm{RMS}} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}$$

[38]

1.5 Supervised Learning Classification

Supervised learning is the umbrella term for AI algorithms that work under human direct instruction. Direct instruction, or supervision, is given in the form of “labeling”. Labeling is the act of pre-determining a set of information with class inheritance so that the machine can learn how to differentiate between different classes of information according to the hints given by its human counterpart. Whilst supervised learning techniques necessarily require labeled data to work, unsupervised learning techniques do not. Unsupervised learning algorithms are more or less capable of “understanding” the difference between the data by themselves [39]. However, since this case study is based on medical data, the researcher decided that a supervised approach would be more ethically reliable since machines are not yet capable of medical diagnosis.

Additionally, the deployment of unsupervised learning algorithms, which might be hard to comprehend due to their complexity, furtherly increases the black box effect from which, even the creator of the algorithm, could not be able to explain entirely what happens behind the curtains. The algorithm explainability must always be a priority when classification models are deployed for medical reasons since their decisions could inevitably affect humans’ lives. Another important limitation of the black box effect is not only inherent to the classification algorithm but also to the data that has been acquired by a third party. Meanwhile, a supervised learning approach might enable the researchers to further inquire also about the quality of the provided data. Given the complexity of unsupervised learning models, one might even think that their performance would naturally surpass the ones of supervised ones. However, this is a general tendency of believing that unexplainable things might guarantee better performance since several scientific studies revealed that the performances of more deep, unsupervised, and unexplainable methodologies gave the approximately the same results as their simpler yet explainable counterparts. This would not be guaranteed by the deployment of an unsupervised learning approach which might provide apparent good results without understanding what exactly were the criteria behind it [40].

1.5.1 Decision Tree

Decision Trees (DT) are machine learning supervised algorithms that were initially invented in 1963 [41]. They quickly became popular not only thanks to their easiness of deployment, but also for their wide range of applications which includes decision making criterias and general predictions. DTs can simplify complex databases by identifying their main features in a simple yet purposeful graphical visualization [42]. These classifiers are characterized by a recursive structure that can express classification rules, or even prediction, in the same way as a natural tree would grow. For classification tasks, starting from the root, the data is assigned to its leaves in accordance to the relevance criteria, where the most important information is tendentially situated at the inner stages of the branches whilst the farthest branches host the elements that have less compatibility with their labels [43]. The main advantage of DTs is the computational efficiency from which they could be deployed even without any knowledge of the domain and without any idea of which parameters to set [44]. This makes them the optimal choice for any early-stage classification analysis since they can provide first feedback on the structure of the data with very little time cost for deployment. On the other hand, DTs are highly susceptible to data manipulation, therefore, even small changes in the features would potentially alter the structure of the tree in an unrecognizable way [44,45]. Also, despite being regarded as efficient within their accuracy capabilities, other classifiers could be developed in a more specific way so that they better address both the data and the classification task. The criterion for the splitting node of the branches is calculated via the following formula:

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\}$$
$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta)$$

[46]

The data node "m" is represented by "Q_m" with "n_m" samples where each of the candidates for the branching are split θ = (j, t_m), where "j" is a feature whilst "t_m" is the threshold that partitions the data into the respective left and right subset of "Q". Meanwhile, the quality of the candidate is determined by the following formula which computes the impurity function (H) depending also on which kind of task, regression or classification, the algorithm must accomplish.

$$G(Q_m, \theta) = \frac{n_m^{left}}{n_m} H(Q_m^{left}(\theta)) + \frac{n_m^{right}}{n_m} H(Q_m^{right}(\theta))$$

[46]

Lastly, θ selects the parameters that minimize the impurity (loss).

$$\theta^* = \operatorname{argmin}_\theta G(Q_m, \theta)$$

[46]

For this study, the loss function is calculated via entropy methodology whilst the tree is branched to a depth of eights which is equal to the number of features.

1.5.2 Support Vector Machine

Support vector machines (SVM) are machine learning classifiers that were originally invented in 1962. Despite being originally proposed only for pattern recognition tasks, nowadays their use has been extended also to text/image classification, biological data analysis, and data mining. Their task is to divide the data points present in the hyperplane, in accordance with any chosen criteria. This criterion is named Kernel and it can be of different shapes such as linear, radial, sigmoid, polynomial, etc. As illustrated in the picture below, each kernel might be suited for solving a specific data distribution or classification/regression task.

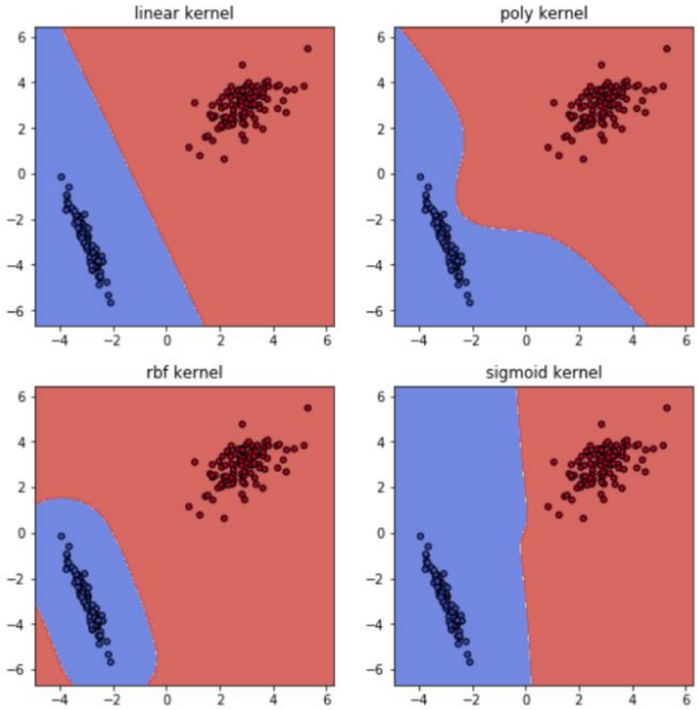

Figure 5. A visual representation of different SVM kernels in action against the same datapoints [47].

The support vector machines are renowned for their ability to successfully differentiate between linear classes, although, the kernel trick might be used to solve also non-linear problems [48]. A support vector machine that attempts to separate, linearly, data points within the hyperplane could be summarized in the following formula:

$$f(x) = w^T x + w_0 = 0 \qquad f(x) \begin{cases} c_1, & if\ w^T x > 0 \\ c_2, & if\ w^T x < 0 \end{cases}$$

[20]

However, it is also important to mention that SVM has a high computational cost since they require to calculate their predictions of each vector of data that has been provided. Therefore, SVM should usually be restricted to small batches of data otherwise their calculation time might become simply unreasonable [48].

1.5.3 Deep Artificial Neural Network

A Deep learning classification algorithm, or in this case, a deep neural network, is an algorithm that can be defined as a multi-faceted representation of a learning process [49]. Neural Networks are however a confusing term since, due to their structure, they are not limited to supervised learning procedures. Neural networks are in fact capable of also addressing dimensionality reduction tasks and even clustering the data in a purely unsupervised manner [50]. That process progresses through a non-linearity principle that could effectively convolve the raw input into a higher degree of abstract representation. The outcome of these transformations is a complex learning function that can, potentially, answer either classification or regression research questions [49]. As illustrated below, the generic structure of an artificial neural network neuron processes the input with its weights into the bias so that an activation function could potentially "trigger" in an output.

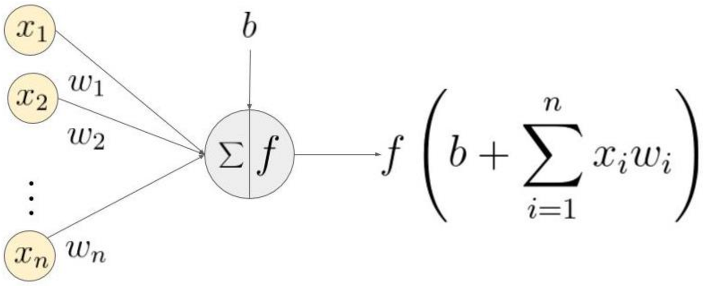

Figure 6. An artificial neural network neuron example where "x" is the input, "w" is the weights, "b" is the bias and f the activation function that gets applied to the weighted sum of the inputs [51].

In this case, the complex learning function translates to learning the structure of the features' data points, where the network is used to reduce the error rate and, therefore, increase the chance of correctly differentiating between the samples, always with respect to their labeling.

1.6 Metrics

The following section will cover the basic notions behind the metrics that have been chosen for evaluating the results of this study. Namely, the F1-Score, the Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and the Precision-Recall Area Under the Curve (PR-AUC).

1.6.1 F1-Score

For asserting the classification, the F1 Score is the most reliable metric for binary classification since it is the harmonic mean of both precision and recall, as explained in the formula below:

$$2 * \frac{\text{Precision * Recall}}{\text{Precision + Recall}}$$

[52]

Specifically, precision is the metric used for calculating the amount of correct positive predictions:

$$\frac{\text{True Positives}}{\text{True Positives + False Positives}} = \frac{\text{N. of Correctly Predicted Positive Instances}}{\text{N. of Total Positive Predictions you Made}} = \frac{\text{N. of Correctly Predicted People with Cancer}}{\text{N. of People you Predicted to have Cancer}}$$

[52]

Meanwhile, the recall, also known as sensitivity, is an index of how “sensible” the classifier is toward the identification of positive samples:

$$\frac{\text{True Positives}}{\text{True Positives + False Negatives}} = \frac{\text{N. of Correctly Predicted Positive Instances}}{\text{N. of Total Positive Instances in the Dataset}} = \frac{\text{N. of Correctly Predicted People with Cancer}}{\text{N. of People with Cancer in the Dataset}}$$

[52]

The following table demonstrates that there is a complete tendency for the Support Vector Machine to prefer non-normalized data, meanwhile the Deep Neural Network seems to eventually benefit from the Loudness Normalization, and the Decision Tree follows the same trend but even to an higher extent. To increase reproducibility, both the DT and the SVM have been assigned to random state variables meanwhile the DNN has been tested three times for each class and the scores are the mean value between the three instances.

1.6.2 ROC-AUC & PR-AUC Score

The AUC is the Area under the curve of the Receiver operating characteristic (ROC). It is a popular metric in machine learning and extremely useful to evaluate binary classification tasks because, unlike the accuracy, both the ROC-AUC and PR-AUC (or just PRC) include in the evaluation all the data points of the chosen model. The values of the curves are averaged using the Riemann Sum of the confusion matrix variables which are later discretized by a set of thresholds that confront both the pair values of recall and precision. Finally, the AUC value is calculated via the height of the recall values and the false positive rates, whilst the PRC is calculated via the height of the precision and the recall. The values are then approximated for both metrics on a scale from 0 to 1 so that they could also be expressed in percentages [53]. Also, for these metrics the DNN has been tested and averaged three times whilst the SVM and DT have been kept reproducible by random state. The following figure shows an example of both performance curves extracted directly from the DNN performance on the wheezes class.

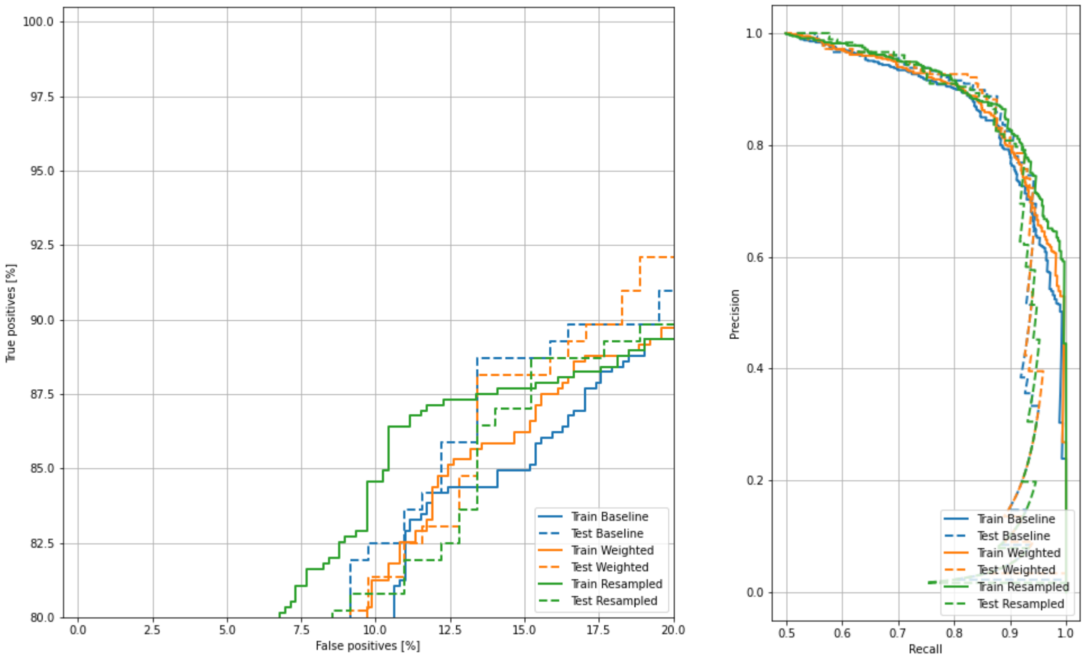

Figure 7. The AU-Curve and the PR-Curve of the class wheezes, both computed on the Deep Neural Network.

1.7 Auditory System Limitations

Nowadays, traditional auscultation with a stethoscope does not match the criteria for a diagnostic test anymore mostly because of the auditory system in the human ear's limitations. The ears can detect deterministic noises in the time or frequency domains, but they are much less proficient at recognizing, identifying, and categorizing noises than machines could be. Specifically, the low signal-to-noise ratio is one of the main factors that contribute to humans' shortcomings when attempting the auscultation of lung sounds. In comparison to the background noise of heart and ribcage muscle sounds, the loudness of the pulmonary sounds is relatively modest. Many doctors do not even longer use any more auscultation as a diagnostic technique since it lacks objectivity and the qualitative nature of lung sounds might be too much susceptible to variations [10].

Another main drawback of conventional auscultation methodologies is that continuous monitoring cannot be provided since physicians are usually required to operate within entire wards and are still not able to be in multiple places at the same time [54]. Also, due to the complex nature of the medical diagnosis, only a qualified expert should perform auscultation, especially when it is opportune to locate and identify the respective adventitious noises. This is also a very limiting factor since, for instance, asthma wheezing is a common nighttime symptom, and due to human biological limitations such as the circadian rhythm, the hospital personnel tends to be tendentially reduced at nighttime [55]. Ideally, auscultation should be performed quietly, ideally when the patient is resting in a motionless position, and under extremely stringent conditions such as avoiding producing artifacts through not-so-careful handling of the stethoscope. Also, there are only a finite number of people who are certified to perform auscultation that could be later used as a diagnostic tool, especially when there is a need to identify the many types of adventitious sounds and, consequentially, determine how this information could aid in diagnosis or monitoring of the patient. An auscultation expert must have a great deal of expertise since there is always a risk that patients and even doctors might overlook symptoms or downplay their seriousness. This would prevent the correct care from being delivered, therefore, it is correct to concur that conventional auscultation can likewise be limited by the human auditory system and its efficiency [21,54,56].

In fact, according to the scientific literature, conventional auscultation should be used carefully when as a reference for automatic pulmonary sound studies [56,57]. For example, the adventitious sounds could be covered by the irregular loudness of the healthy respiratory sounds, resulting in partial or even total overlapping. Most importantly, the human ear could miss some instances where the intensity might be too low to be heard due to the inter-variability of the adventitious sounds. Consequently, the very usefulness of conventional auscultation, purely as a method of symptom monitoring and management, is ineluctably hampered by the mentioned shortcomings. Ideally, these restrictions could be overcome by a reliable automated lung sound analysis, that could accurately classify adventitious sounds. [56]. Additionally, given the elusiveness of the adventitious sounds, it would prove really complicated for a human to accurately pinpoint the location of the adventitious window within the time domain since the time window could be the size of even a few milliseconds.

Additionally, one of the most prevalent sensory decreases in the aging population is hearing loss, also known as Presbycusis. Presbycusis is characterized by a loss in speech perception as well as a worsening of the processing of temporal sound characteristics, suggesting a potential central component [58]. Presbycusis is a bilateral hearing loss that shifts with aging from high to low frequencies. The variety in hearing level is only sporadically correlated with age, and the pace of hearing decline is highly variable and nonlinear for each subject [59].

Hearing is the ability to perceive sounds. The ear is a technological marvel because its sensory receptors can convert sound vibrations with amplitudes as small as 0.3 nm (the size of an atom of gold) into electrical signals 1000 times more quickly than photoreceptors can do the same for light. The receptors for the equilibrium, the sense that aids in balance maintenance and spatial orientation awareness, are all elements that are found within the ear. Anatomically, the ear is divided into three main areas: the internal ear, which contains the hearing and balance receptors, the middle ear, which transmits sound vibrations to the oval window, and the external ear, which gathers and routes sound waves inward [60].

Unlike the cochlea, which is significantly impacted by aging, the external and middle ear anatomical changes associated with aging do not appear to have a negative impact on audiometric function or the capacity to perceive speech. The inner ear's functional components alter with age due to a variety of pathophysiological events. Meanwhile, the aging process is determined by genetic elements, but it is also influenced by environmental factors such as lifestyle. [59]. The following image illustrates the internal structure of the cochlea and its related frequencies. This illustration of the inner ear's cochlea, shows how and where various frequencies are processed. Presbycusis patients may lose the ability to perceive these frequencies effectively since the inner section of the cochlea is the first to degrade with ageing.

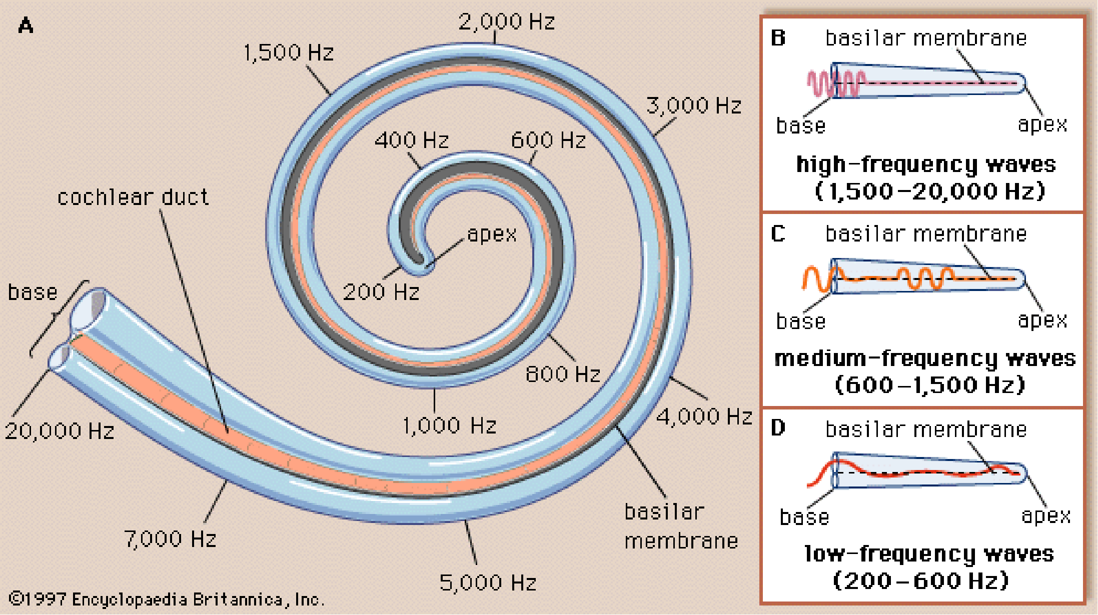

Figure 8. A visual representation of the structure of the cochlea (inner ear) and where the different frequencies are processed. The inner part of the cochlea is the first to degenerate due to the aging process, therefore, those affected by Presbycusis might lose the ability to perceive effectively such frequencies [61].

Different frequencies of sound waves cause different parts of the basilar membrane to vibrate more strongly than other parts. Each of the segments of the basilar membrane is specifically "tuned" for a certain pitch. High-frequency (high-pitched) noises create the most vibrations towards the base of the cochlea (closer to the oval window), where the membrane is smaller and stiffer. The basilar membrane is broader and more flexible near the cochlea's apex, where low-frequency (low-pitched) noises cause the basilar membrane to vibrate the most. The strength of the sound waves determines how loud something is. Larger basilar membrane vibrations brought on by high-intensity sound waves result in greater frequencies of nerve impulses reaching the brain. Additionally, louder sounds could activate more hair cells [60].

With this knowledge, one could argue that biological factors could eventually limit the reliability of the already established medical sole auscultation diagnosis [54]. However, it is not the aim of this research to define the limits of the actual methods, nor an attempt at replacing them since, AI, should never act undetermined when in direct connection with human health. Instead, AI could and should be employed as an augmentative asset that could enhance human performance and ensure, in this case, the best possible outcome for the patients.

1.8 Cognitive Limitations

When focusing on a difficult activity, attention can function as a set of blinders, allowing significant inputs to go right by our eyes unnoticed. A peculiar study from 1999 involved their participants passing a ball to each other in movement while a person dressed as a gorilla wandered around the "game". The game was recorded and later shown to other participants of the study that did not participate in the activity nor were prepared for the detection of the gorilla. Remarkably, most participants did not notice at all the gorilla, therefore, this study became famous for demonstrating the existence of the phenomena known as "sustained inattentional blindness". This process happens because typically no one would expect a gorilla to dance in such a scenario. However, the following question becomes: Does inattentional blindness (IB) persist when the observers are trained professionals? Some evidence suggests that competence might further lessen the impact of IB. Unfortunately, on the other hand, multiple studies revealed that the rate of IB is still relevantly present in all fields. Most worryingly, this phenomenon is also present in radiology, a field where relevant professional figures are involved. Specifically, within the healthcare field, the image of the dancing gorilla (scaled to be 48 times bigger than the average nodule) was integrated into a random amount of CT scan images. Despite all the preparation and experience, 60% of the radiologists failed to detect it [62]. The following image shows in detail the gorilla used for this test at different opacities.

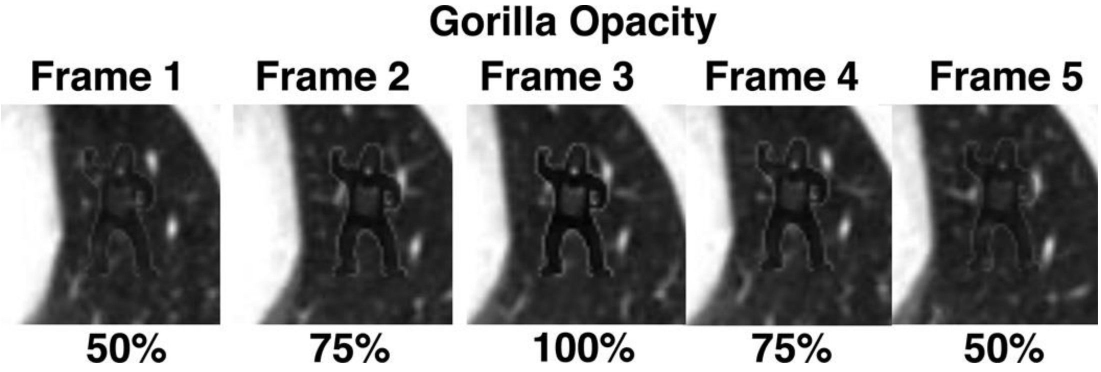

Figure 9. The dancing gorilla inserted into Lung Computer Tomography images at different opacities [62].

The scientific evidence supports the idea that it is typically more difficult to keep attention in tedious, intellectually undemanding settings rather than in fascinating ones since they could naturally be perceived as cognitively demanding. Within the psychology sector, experts are aware of this ostensibly counterintuitive paradox where it is harder to maintain attention on tasks that are regarded as easy, therefore trivial, than keeping the focus on complex ones.

Easier tasks could be related as basic or even repetitive, therefore requiring an unnatural demand for higher attention. This higher demand has been frequently observed to be associated with an increased stress response and, therefore, the energy expenditure becomes less efficient for the subject. Considering the actual distinction between cognitively easy and more complex tasks, it is also crucial to define the term "vigilant attention" (VA) to describe the ability to sustain attention during these repetitive and intellectually undemanding tasks. Specifically, VA is used to refer to the process of maintaining effective conscious sensory processing for lengths longer than around 10 seconds and up to many minutes. The VA process involves the detection and discrimination of the stimuli, including basic cognitive-motoric responses. However, it excludes the higher executive or attentional functions such as spatial orientation, resolving interference, dividing attention, or choosing between several overt responses [63]. To this extent, it would be important to ask if the simple action of auscultation, which might prove trivially undemanding to experienced medical practitioners, could trigger the same scenario where the VA is lowered by the increasingly higher perception of expertise against a repetitive and trivial task. Also, could a prolonged emergency like a pandemic further lower the accuracy of the analysis since it would force the medical personnel not only to work extended hours but also to repeat the same operation many additional times?

Another concern toward the attention span deficit is offered by the evidence that listeners frequently struggle to keep their attention on pertinent audio streams while weeding out distracting ones in complex auditory environments such as a medical ward. The listeners necessarily must employ auditory information to separate the target stream from the distractions in order to solve this difficulty [64]. Within the pulmonary field, clinicians might not always operate in optimal conditions, especially when addressing crowded scenarios such as the ones who have been already mentioned (pandemic outbreaks, busy hospital wards, etc). These conditions might trigger the demands of an auditory “cocktail party", which means that the clinicians are required to focus on the target of the signal whilst trying to track both the spectral and temporal properties of all the other auditory streams. This situation is also somehow relatable to the source separation problem that is frequently addressed in signal analysis [65]. In this case, the human brain, like its electronic counterparts, must perform an independent component analysis to diversify between the multi-channel observations to which is exposed to. It is consequential to concur that this process is indeed energy-demanding since it directly relates to multi-tasking, therefore, other than increasing stress it also lowers overall brain efficiency [66].

Additionally, it is widely accepted that the medical profession can be stressful, particularly during the internship years. Due to their long work hours and variable schedules, clinicians are prone to weariness and chronic sleep deprivation [67]. Unfortunately, as demonstrated by the precedent studies, not even a high level of expertise can completely immunize from the inherent limitations of the human perception and attention span [62], thus, when even external factors contribute to lowering the awareness and overall energy of the practitioner, the reliability of conventional auscultation diagnosis becomes ineluctably compromised [13]. From all of this, the research question of this study arises: Could machines provide reliable help to overcome their own creators' limits?

1.9 Piezoelectricity & Stethoscopes

In the field of medicine, stethoscopes are of the most essential diagnostic tools. The stethoscope can be utilized to listen to lung sounds as well as heart sounds to diagnose lung abnormalities and heart-related disorders. Auscultation is the medical term used for describing these actions. [57]. The invention of the stethoscope is originally attributed to Rene-Theophile-Hyacinthe Laennec in the 1819 [17,68]. Since then, the stethoscope became a staple symbol in medicine to the point that is nearly associated with the meaning of medicine itself. After conducting in-depth research, Laennec presented his first prototype in 1816, which was nothing more than a rudimentary tightly coiled paper tube. This primitive design later changed to a thicker pasteboard cylinder and then to a more durable stage of cellulose, a wooden tube. Despite their former inverse durability evolution phases, the stethoscopes still underwent several changes in the successive years, until they finally became completely synthetic like the ones we use in our modern era. Perhaps, the first attempt to a binaural stethoscope was attempted by Doctor George Cammann in 1855. As the following figure shows, this early model had a tube for each of the earpieces and it extended until the single bell-shaped chest piece, exactly like the ones that we are familiar with [17,19]. The following figure shows the first prototype of the binaural stethoscope ever made.

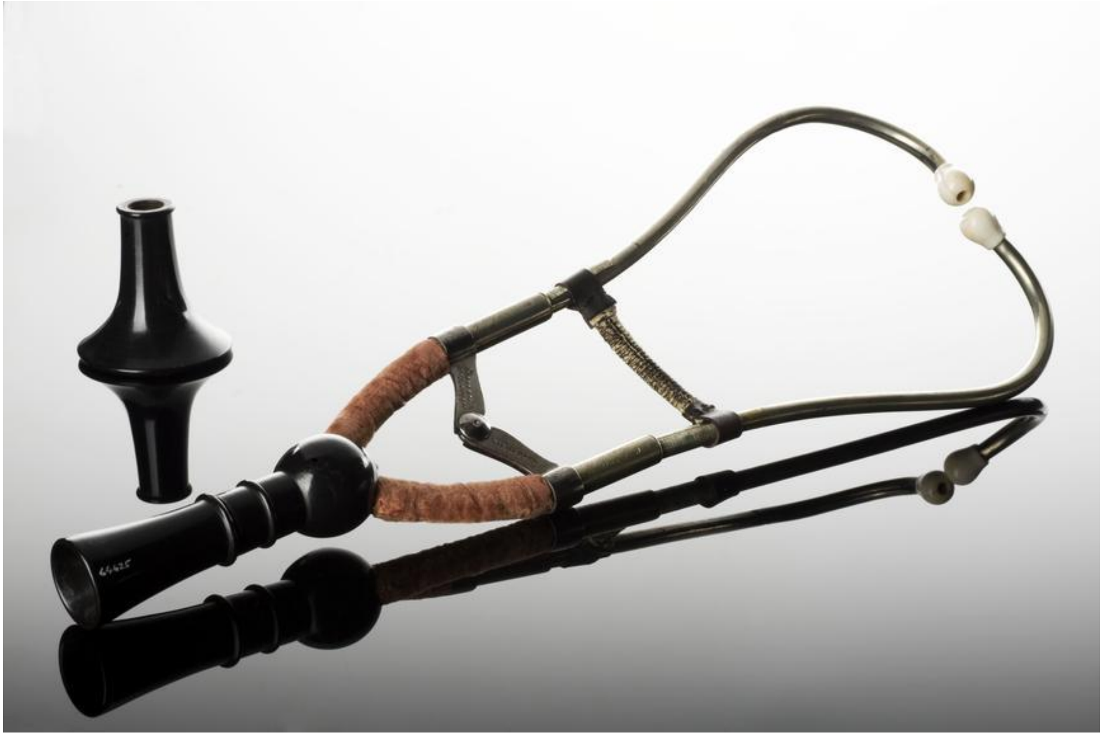

Figure 10. The first stethoscope made by Dr. George Cammann in 1855 [69].

However, it is also important to mention that, the stethoscopes, are always directly dependent on the medical staff's expertise and knowledge, which can lead to significant inter-observer variability, meaning that diagnosis could be discordant among different doctors [2,6–8,10,17,20,54]. Additionally, they might vary in the quality of their materials and even the technology could be diverse between models. More gravely, also the lack of a standardized nomenclature of the respiratory sounds could become a further deterrent for the reliability of sole auscultation diagnosis [17]. Since the accuracy of any conventional auscultation cannot be properly assessable by any known gold standard [70], electronic stethoscopes could potentially solve this issue since they can record the auscultation and store it for future classification or analysis.

For example, doctors could use this capacity of electronic stethoscopes to quickly make a preliminary diagnosis based on the recorded data so that the patients who require secondary prevention, after being discharged from the hospital, could still benefit from long-term monitoring. Additionally, if also an early abnormality detection feature could be integrated into the electronic stethoscopes through a reliable classification algorithm, this would enable the doctors to react in time and eventually reintegrate the patient into hospital care to avoid further complications. In detail, a digital stethoscope transforms the acoustic vibrations (sounds) into electronic impulses that can also be eventually amplified for better listening quality. These signals can then be further analyzed and converted to digital form before being transmitted to a laptop or personal computer. As a result, the digital stethoscope presents an evolutionary improvement to the patient digital healthcare system's landscape [3,70]. However, the majority of digital stethoscopes now available only offer a basic recording capability without automated analysis [3].

Air-conduction and contact-conduction electronic stethoscopes are two popular varieties of electronic stethoscopes. They are both characterized by different types of sensors. The electromagnetic coil or electret capacitor is used as the sensor in the air-conduction electronic stethoscope. Its task is to gather sound signals and it performs in a reliable way but with limited sensitivity. On the other hand, piezoelectric materials perform better against interferences, and they also have an overall better sensitivity. They are usually placed as sound sensors in contact-conduction electronic stethoscopes. However, piezoelectric materials tend to be quite fragile, therefore, any bending of the components during the manufacturing process might reduce substantially the future stethoscope's sensitivity. [57].

In synthesis, piezoelectric materials possess the ability to produce electrical signals when physical pressure is applied to them [71]. Pierre and Paul-Jacques Curie made the discovery of piezoelectricity in 1880. Their discovery took place when they noticed that crystals such as quartz, tourmaline, and Rochelle salt, created a voltage on their surface in response to being squeezed. The following year, they saw the opposite result. The same type of crystals grew longer when under the effect of an electric current, therefore establishing the reverse property of the materials. Today, many useful tools and equipment feature piezoelectric materials: microphones, telephones, and sonars, they all depend on the same discovery [72]. The following illustration summarizes the concept of piezoelectricity where he pressure that results in the interaction of the two opposing poles within the piezoelectric material generates a voltage.

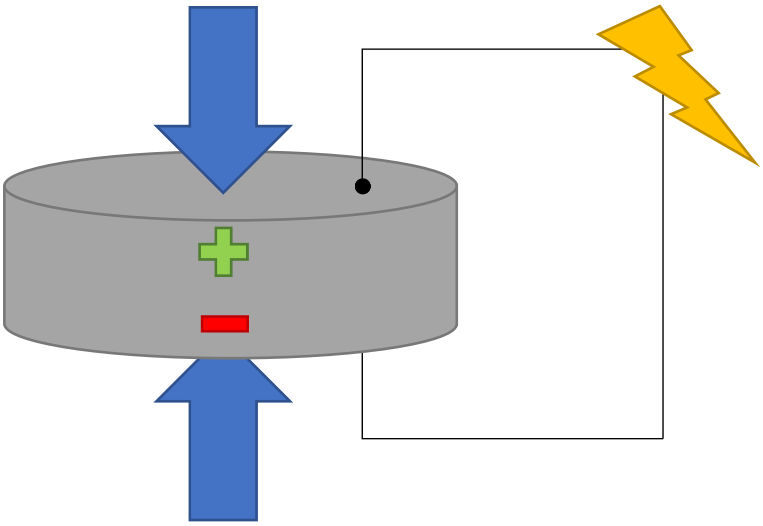

Figure 11. The piezoelectric material (grey) generates a hypothetical voltage from the pressure that causes the two opposite poles to interact with each other within the material.

Since abnormal cardiopulmonary sounds cannot be recorded with conventional purely mechanical stethoscopes, the results of the auscultation may change from doctor to doctor due to many factors such as hearing, experience, or just even opinion divergence [57], a new discipline called "computer-aided auscultation" has emerged after the introduction of the electronic stethoscopes. The acoustic-based automatic diagnosis of heart disease via electronic stethoscope has garnered a lot of attention in recent years and all thanks to the technological advancements in the signal processing techniques that involve machine learning [73].

Today, multiple typologies of stethoscopes are produced and used around the globe and their ability to record patient sounds in high quality has been made possible only by the improvements in stethoscope technology [74]. To successfully comprehend and analyze stethoscopes’ signals it is important to have at least a basic comprehension of the science behind it. For example, while digital stethoscopes may record sounds throughout the whole spectrum of detectable sound frequencies, traditional bell and diaphragm stethoscopes may attenuate higher-frequency noises like wheezing adventitious sounds [70]. Therefore, this brief conceptual presentation of piezoelectricity is offered as an incipit to guarantee the comprehension of how the signal is acquired.

1.10 AI Historical Incipit

Today, technology gives us the power to alter our reality in nearly any way. The active use of applied science has become necessary for not only social interactions and amusement but also for work-related activities and education, which can be achieved by and through the meaning of scientific application [75]. One of the most well-liked fields of information technology is artificial intelligence (AI). Computers were formerly created and utilized for the sole purpose of calculating algorithms. Regardless, computers, and more specifically machine learning, were also quickly employed to handle issues that were too complicated or incomprehensible for any algorithm to solve. [16,76,77]. Nevertheless, due to its impressive potential, artificial intelligence, after the advent of the digital era, has been recently identified as the fourth industrial revolution [16]. The following diagram illustrates the hierarchical entity of AI in correspondence with its sub-entities (Machine Learning & Deep Learning).

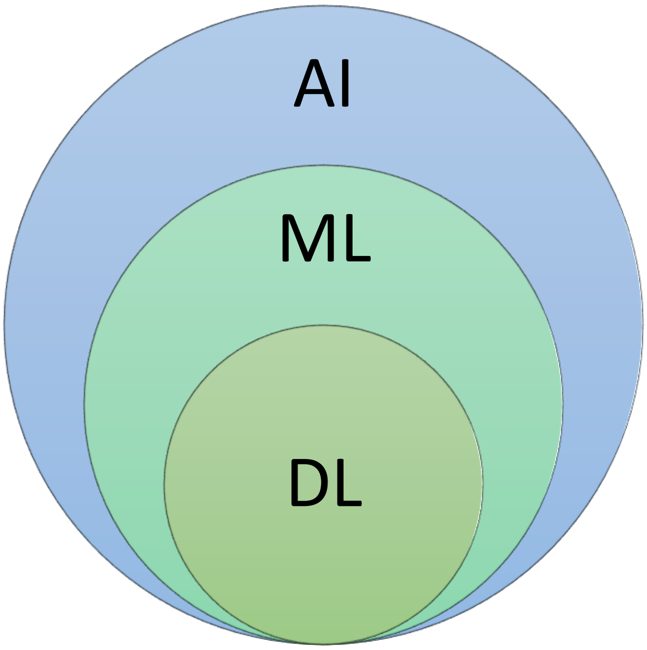

Figure 12. A hierarchical classification (Venn Diagram) of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

Machine learning (ML) is an interdisciplinary field that shares connections with mathematics' statistics and information technology. But why exactly? Since the final goal is to program computers so that they can learn, it makes sense that this field falls under the computer science umbrella. However, there are still significant distinctions that must be made though. For example, if a doctor proposes the theory that smoking is associated with a higher risk of developing heart disease, it is up to the statistician's (or nowadays data scientist's) responsibility to examine patient samples and determine whether the theory is valid through hypothesis testing. In contrast, ML seeks to define the causes of heart disease using the data collected from patient samples. There is still a possibility that automated methods will also be able to identify significant patterns or hypotheses that a human observer could have easily overlooked. Since the ability to transform experience into knowledge or to recognize significant patterns in complicated sensory input is a foundational aspect of human intelligence, ML can be seen as an attempt by humans to artificially reproduce the same learning mechanisms that they naturally possess. Nevertheless, it should be noted that machine learning, unlike traditional AI, does not aim to create an automated imitation of intelligent behavior. ML aims to use the powerful features of computers to augment human intelligence and increase human potential. Frequently this takes the form of carrying out tasks that are far beyond human capacity such as interpreting meaningful information from enormous databases [16,77].

Although the initial steps of the creation of AI might be hard to date accurately, the 1940s, and more specifically 1942, is when the American writer, Isaac Asimov, started to theorize about AI by releasing his short novel “Runaround”. [78]. More specifically, the scientific history of AI can be traced back to the early 1950s, when the term “artificial intelligence” was first coined by a small group of scientists, led by Alan Turing and John McCarthy. These scientists began exploring, for the first time, the possibility of creating machines that could think like humans. The first AI research projects were solely focused on using computers to solve complex mathematical problems and it is only thanks to these early efforts we were able to lay the foundation of AI for the progress of our era [79–81].

In the 1960s, AI research solved most of their complex mathematical problems and began to focus on developing computer systems that could understand natural language and, therefore, interact with humans. These studies eventually led to the development of real “expert systems”, which used rule-based algorithms to make decisions and solve problems. One of the most successful examples was the ELIZA computer software. Originally developed between 1964 and 1966 by Joseph Weizenbaum at the Massachusetts Institute of Technology, it was one of the first algorithms that managed to “elude” the Turing Test. ELIZA was in fact a natural language processing tool, so powerful for the time, that it was able to successfully mimic a conversation with another human.

In the 1970s, AI research tilted towards creating even more sophisticated machine learning algorithms. At the same time, the U.S. Congress started to criticize the “unjustified” heavy expenses of AI research, and the British government ended its support for AI research. Both the U.S. and the U.K. were the pioneers in the field of AI research at the time. Despite this, the newly discovered algorithms still allowed computers to learn more from data and improve their performance over time. This poorly funded research gave birth to the foundation for modern AI, such as deep learning (DL), which powers many of today’s AI applications [78].

In the 1980s, AI research continued to develop more complex neural networks, such as Artificial Neural Networks (ANN), and other machine learning algorithms that are still used nowadays. This period also saw the birth of robotics which allowed, for the first time, machines to interact with their environment [82]. One of the main advantages of Artificial Neural Networks was that they offer the chance to solve a higher degree of non-linear problems that, the already employed classical statistical models, were not as capable of solving [83]. An example of the basic element of the Artificial Neural Network is offered below where the input gets processed by the structure and transformed into the output.

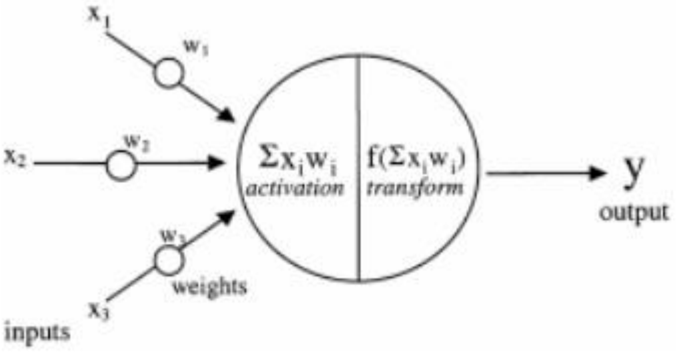

Figure 13. A biologically inspired artificial neural network neuron where the structure transforms the input into the output after processing it [49].

However, since the 1990s, AI research focused on creating even more intelligent agents, such as virtual personal assistants and autonomous robots. This is the time when Artificial Neural Networks (ANN) research decided to dedicate more to Deep Artificial Neural Networks [79]. Since that, deep learning has recently shown a lot of success in solving many machine learning challenges. Especially, since the early 2010s, AI has faced an even more important resurgence due to the efficient creation and utilization of image classifiers [16]. The use of deep neural networks (DNNs) for modeling and learning the unknown structures in machine learning tasks, which are typically characterized by some unknown function mappings or unknown probability distributions, is largely responsible for the success of deep learning [84]. This, for example, enabled the development of applications such as Amazon's Alexa and Apple's Siri.

The most significant question that we still face nowadays is: what exactly is intelligence, and if it can be human, could it also be completely artificial? A short glance in any dictionary will yield several definitions, some of which include understanding, knowledge, and the capacity for logic and experience-based learning. Each definition might potentially reflect what could be observed in both human and machine behavior. However, both human prejudice and the overall shallowness of modern artificial intelligence still prevent us from reaching a conclusive answer. What are the elements that might have hampered our comprehension of AI? Is it because we use AIs to study AIs or, maybe, have we not made enough progress yet? The modern artificial neural networks, whose roots were created in an effort to mimic the information processing seen in real neurons, are perhaps one of the key components of contemporary AI that could give us a hint of the direction in which we are proceeding. [82]

Intelligence could be classified as either cognitive, emotional, or social. Depending on the forms of intelligence it demonstrates, AIs can be categorized as either analytical, human-inspired, and humanized AI, or Artificial Narrow, General, and even Super Intelligence. The requirements for the latter are directly dependent on its evolutionary stage [78]. AI is generally depictable by the ability of a machine to emulate the various aspects of human intelligence. One of the main purposes of AI engineering is to try to build machines that can use human intelligence to solve issues human issues and, therefore, improve our environment [81].

Today, AI is being used in a variety of industries. From healthcare to finance, its applications range from automated customer service to self-driving cars. As the AI field continues to develop, its uses and applications are expected to become increasingly diverse and powerful. However, as the power range of AI technology keeps increasing, ethical concerns are also consequentially addressed [76,79,81], which includes also the infamous black-box effect [83]. The black-box effect is describable as the incapacity of the researcher to explain the reasons why the algorithm came to a certain conclusion. Thus, when an algorithm is specifically applied to the medical field, one should always ensure that the decision taken by algorithm follows the requirements of the ethical principles for healthcare-related decisions. Any reason or bias behind the algorithms should always be accurately describable to ensure, not just the efficacy of the latter, but also its reliability [85].

1.10.1 Supervised vs Unsupervised

Learning activities can generally be divided in accordance with the type of interaction that occurs between the learner and the environment [77]. Within the AI field, there are three "main" types of machine learning modalities; supervised, unsupervised, and reinforced [16,20]. The first difference worth mentioning is perhaps that of supervised learning methodologies against unsupervised ones.

Supervised learning scenarios could be described by scenarios in which the algorithm is provided with a training example that contains significant information about the data and a test one that does not. The information could be for example the spam or not spam labels present within an email account. It is the task of the algorithm to apply what it has learned in the training phase to the testing phase; therefore, we could see the training phase as a pure learning process for the algorithm to gain the experience required to successfully process the test samples. When the testing phase commences the algorithm must "guess" the missing information of the data based on the acquired experience. More specifically, supervised learning adopts a labeling system to communicate to the machine where the data belongs, therefore, this method is often compared to a teacher who actively supervises the learner. On the other hand, unsupervised learning does not provide any pre-determined separation of the data. The purpose of unsupervised methodologies is to provide a summary or a compressed version of the data that has been processed. More specifically, these tasks are describable as either clustering or dimensionality reduction [77].

Specifically, supervised machine learning is one of the most prominent and successful tools for addressing classification and regression tasks. For example, supervised learning should be used whenever there is a need to predict a certain outcome from a given input, and there are already examples of how the outputs should be. Building a training set from scratch is frequently considered a hard labor task, but if it is correctly done and the data is sufficiently good, supervised machine learning could guarantee great results. In synthesis, supervised learning enables us to address tasks that would have been otherwise impossible due to the magnitude of the data or just for its complexity. [86].

In supervised learning, all the inputs are automatically assigned to the necessary classes. The classes can be of any type. For example, if you want to categorize food you might want to create one class for vegetables and one for fruit. After you create the classes, you train the algorithm with enough elements so that it will learn the differences between those two families of elements that you provided. The algorithm will learn from the labels during the training phase which element corresponds to a determined set of features. The features are those components that can be extracted from the elements. In the case of fruit vs vegetables, the features could be shape, color, or even more complex factors like nutritional values, or water percentage composition. Basically, anything that could be extracted from either a visual or numerical source, could be used by the machine to learn potential associations. Nowadays, many companies rely on machine learning for their services such as tracking, data mining, object recognition, medical imaging, and a multitude of other tasks that grow exponentially every year and are all technically aimed at the final benefit of society. Overall, supervised learning guarantees more accurate classification results than unsupervised methodologies. Supervised learning also has a reduced cost in the overall complexity of the development and later deployment of the model. As always, this proceeds in accordance with the correct labeling and quality of the input data. Regardless, the supervised methodologies are also inherently dependent on the amount of training data provided to the model which, in enough, might prevent eventual over-fittings or under-fittings, depending on the amount and typology of provided data [11].

In contrast to conventional machine learning methods, deep learning ones adopt a series of non-linear functions to describe their features. These functions also serve the purpose of lowering mistakes and, therefore, they aim to increase overall accuracy [4]. Unsupervised dataset transformations use algorithms to represent data generally complex unlabeled data in a much simpler way so that it could be more easily processed and comprehended by humans or other less sophisticated machine learning algorithms. Dimensionality reduction is one of the most common applications of unsupervised learning. Dimensionality reduction starts from a high-dimensional representation of the data, generally also made up of numerous features, and tries to discover new ways to associate and describe it by reducing the actual number of features. This procedure is generally adopted to reduce the multidimensionality of the data to a number of two dimensions only, much easier to visualize and comprehend for the human brain. Another powerful ability for unsupervised learning is to find the elements that compose the data like, for example, extracting the topics out of documents folders. The aim of the algorithm might be to successfully identify the unidentified subjects that are discussed within each document, and further discover which topics each document covers. These kinds of tasks are already deployed for monitoring conversations on social media to prevent the distribution of illegal data. [86]. Other examples of unsupervised learning tasks could also be anomaly detection, clustering, density estimation, and association rule development [11]. The difference between a simple ANN and a more complex DNN, where more than 3 hidden layers are present, is offered below.

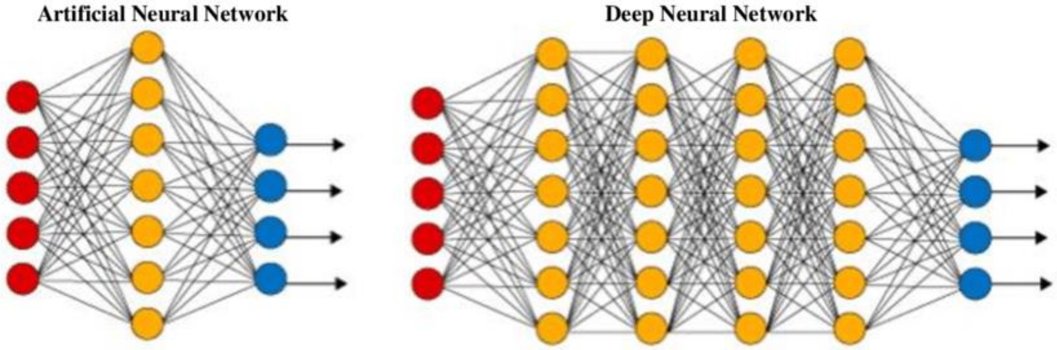

Figure 14. The difference between a "shallow" Artificial Neural Network and a Deep Neural one is in its number of hidden layers (3 or more) [87].

In reinforcement learning, the data labels are obtained without any explicit instruction from the machine. Instead, the data is acquired from the interaction with the dynamic environment. The dynamic environment is composed of the feedback that the algorithm receives in accordance with its behavior and human expectations. The feedback could either reinforce positively the algorithm's logic, or it could discourage it negatively in case of unacceptable results. Reinforcement learning can also be thought of as a combination of both supervised and unsupervised methodologies [16]. Regardless of definitions, the objective of this study is to attempt to classify lung sounds binarily. A supervised approach has been preferred since the algorithm should be completely accountable for its decisions which must be made in accordance with a professional diagnosis. Therefore, given the complexity of respiratory signals, an unsupervised approach would have not proved medically reliable in its attempts to cluster the data without any sort of guidance.

1.11 Databases

Databases are digital storage structures that allow any sort of information to be stored and easily accessed for retrieval. Databases are specifically designed for accessibility, meaning that the stored information should be easily retrieved, modified, and re-stored again, from multiple locations and even at the same time. A database might be stored either as a file or a set of files. The files' contents can be divided up into records, each of which might have one or more fields within. The fields are the fundamental unit of data storage, and each field typically contains data related to one feature or attribute. Additionally, records can be arranged into tables that contain details on the connections between their various fields, meaning that different entries could still be entwined. Nowadays, almost all databases generally offer cross-referencing capabilities, meaning that users can quickly search, rearrange, organize, and choose the fields in numerous records. This enables the users to access or produce reports on certain data aggregates by using keywords and any variety of sorting instructions [88].

Inherently minor in its size, a dataset is a collection of separate sets of information that is treated as a single unit by a computer [89]. It is overall agreeable that databases might be composed of more than one dataset. Currently, there are four important open-access respiratory sound databases (repositories) : the King Abdullah University Hospital (KAUH) database, the International Conference on Biomedical and Health Informatics (ICBHI) 2017 database, the HF Lung V1 database , and the SJTU Pediatric Dataset [2].

1.11.1 SJTU Pediatric Dataset

The Shanghai Jiao Tong University and its affiliated hospitals jointly produced the first open-access pediatric respiratory sound database, known as SPRSound. The database has 2,683 recordings and 9,089 respiratory sound occurrences totaling 8.2 hours from 292 people. The database contains label annotations at the event and record levels. The count of Normal, Rhonchi, Wheeze, Stridor, Coarse Crackle, Fine Crackle, and Wheeze & Crackle events are 6,887, 53, 865, 17, 66, 1,167, and 34. At the record level, there are 1,785, 233, 347, 131, and 187 Normal, Continuous Adventitious Sounds (CAS), Discontinuous Adventitious Sounds (DAS), CAS & DAS, and even “Poor Quality” recordings, respectively. The average time between respiratory sound occurrences and records is 1.3 seconds, and it is 11 seconds.

The files are saved as .wav and it adopts to the following naming conventions. Each record's name includes five items of data—the number, age, gender, record location, and record number of the participants— all are separated by underscores. Each record's label annotations, including those at the record and event levels, are recorded in JavaScript Object

Notation (JSON) format with the same filenames (the beginning, the end, and the corresponding type of the respiratory sound event).

The digital stethoscope, Yunting II type II Stethoscope, was used to record the respiratory sounds. It possesses a sampling frequency of 8 kHz and a quantization resolution of 16 bits. Experienced doctors with a background in clinical measures and specimen collecting carried out the procedure. The authors acknowledged that the respiratory sounds of pediatric patients are inevitably weaker than the ones recordable from adults. Regardless, they also considered that the best locations for hearing pulmonary sounds generally are available on the dorsum, therefore they abstained from recording from the pectus since that would have been a more adequate zone for capturing heart sounds. Additionally, the researchers ensured that the collection time for each location is a minimum of 9 seconds so that at least two respiratory cycles would be present within the record. The patients were also ensured to be in a resting position during the auscultation and exhorted to cooperative quietness to ensure the quality of the recording. The recordings were further labeled with a custom-made sound annotation software, namely SoundAnn.

Additionally, it is worth mentioning that despite baby pneumonia mortality having decreased, pneumonia is still the largest infectious cause of death in children under the age of five, worldwide. Pneumonia can be difficult to diagnose, especially in areas with limited resources when chest radiographs and point-of-care diagnostic tools may not be easily accessible [70]. The importance of accuracy within data acquisition is utterly reinforced by this knowledge, nonetheless, the production of a high-quality database could really contribute one day toward diminishing infants’ mortality.

2 METHODS

The methods of this study have been already presented in the basics section of the introductory part. For the preprocessing part, the loudness normalization is the only "amplitude altering" technique that has been kept for evaluation, whilst, the list of features vector is the following:

- Variance
- Range
- Sum of Simple Moving Average (Coarse)
- Sum of Simple Moving Average (Fine) [128 window size, moving by 16]
- Spectrum Mean
- Spectral Kurtosis
- Zero-Crossing Rate
- Root Mean Square

Successively, the feature vectors are "fed" into the three already mentioned classification algorithms:

- Decision Tree
- Support Vector Machine (SMOTE, Cross Validation = 10, Bayesian Optimization)
- Deep Artificial Neural Network (Forward Feed, Adam Optimizer =1e-3, Binary Focal Loss Gamma = 2, Random Weight Sampling)

Specifically, the backend has been structured in the following way:

I. Pre-Processing
II. Features Extraction
III. Classification

I. The Pre-Processing part extracts the chosen breathing events in accordance with the given duration and the choice to apply the Loudness Normalization up to the mean value of the whole database or not apply it at all. The audio events are extracted into a table of amplitude values which are characterized by the ID of the event on the row and the time in the columns.

II. The Features Extraction part computes the eight different features a predisposes the binary classification by re-labeling them into 0 and 1 respectively for healthy and adventitious classes

III. The classification part runs separately for each of the classification algorithms.

The performance of the algorithms are evaluated via the same metrics that were also introduced in the basics section:

- F1-Score
- Receiver Operating Characteristic Area Under the Curve
- Precision-Recall Area Under the Curve

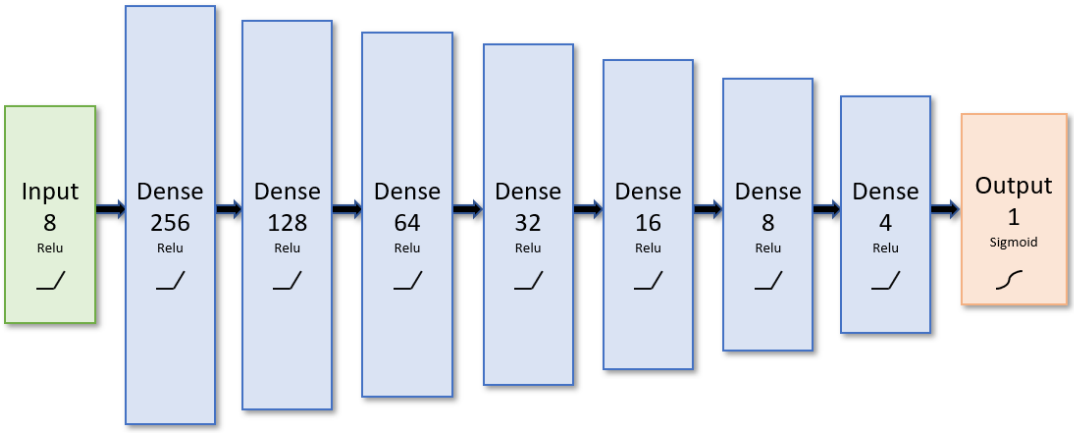

Figure 15. The structure of the Deep Neural Network.

Like shown in the image, the network runs in deep forward feed mode, with half number of perceptrons for each hidden layer, activated through only Rectified Linear Units (Relu) activation functions which ensure the positivity of the output for each layer. The last layer is necessarily sigmoid to ensure the binary classification. The optimizer is set to Adam Optimizer and with a learning rate of 1e-3, whilst the loss function is Binary Focal Loss with a Gamma of 2. Also, the neural network also performs a Random Weight Sampling by deducing the weight of the two classes and imputing artificial datapoints for the minority class in the attempt to balance the unevenness of the two. The imputation of the datapoints is based on the random reproduction of points that were already present within the database, therefore the weighted random sampler does follow a precise criterion that aims to produce information like the already present one.

Also, the SVM has been optimized into a pipeline that tests all different kernels and hyperparameters. The pipeline returns the kernel with the best results, and it has also been additionally programmed to address the class imbalance problem with the Synthetic Minority Oversampling Technique (SMOTE). Unlike random sampling, the SMOTE performs oversampling by using the Euclidean distance of the data points as a parameter. SMOTE, or more precisely the SMOTETomek variant, can even address under-sampling tasks by applying the same principle for the removal of the data points that are considered in excess [90]. Also, the cross-validation is set tenfold, and the Bayesian optimization is performed over the hyperparameters.

2.1 Materials

This section will cover all the time domain steps that have been adopted for the preprocessing of the recording. The following plot visualizes the clear division of the samples for the class of stridor, which despite being limited to only 14 samples still can offer a view of how the features should be ideally shaped to ensure class divisibility.

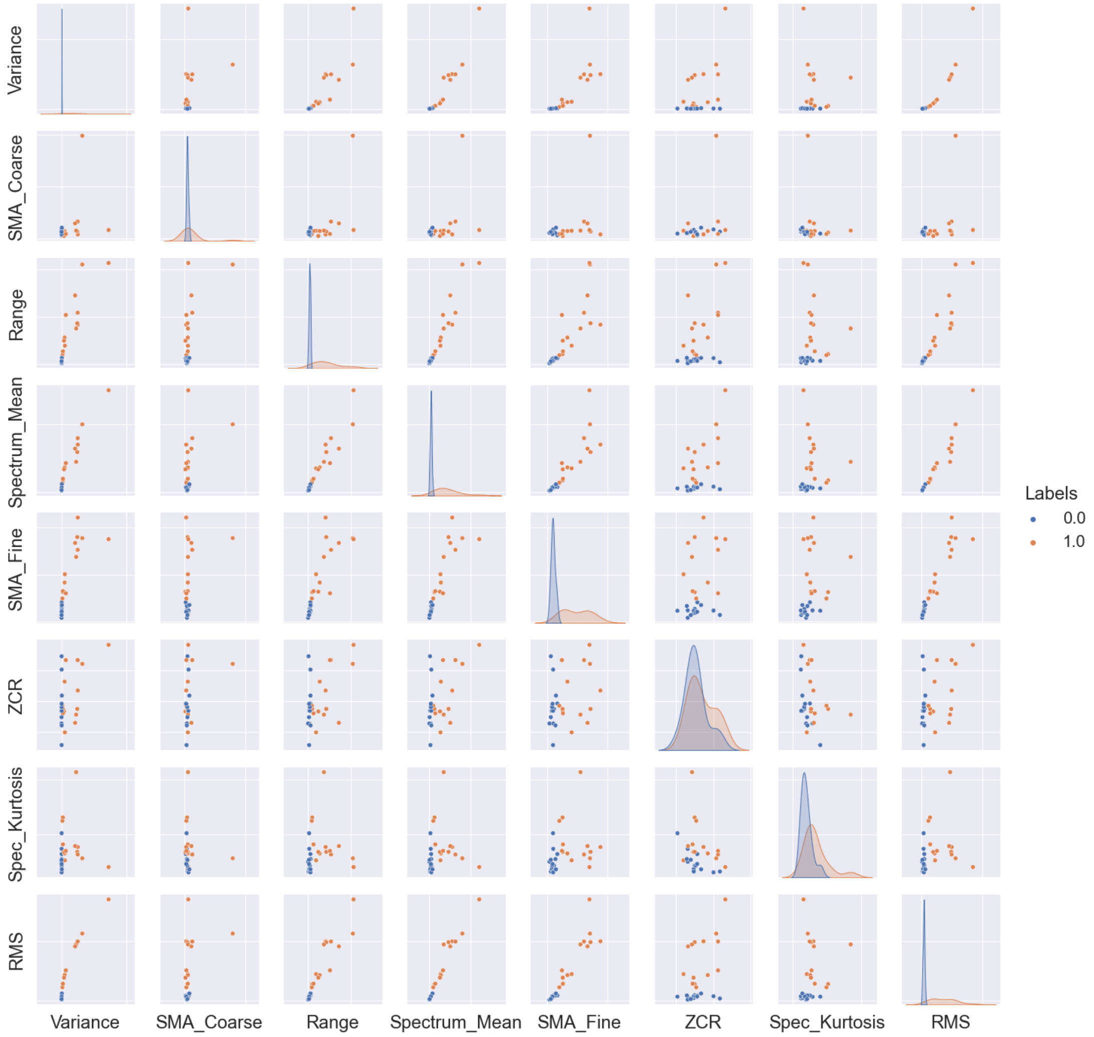

Figure 16. A scatterplot of the data points distribution for the Stridor events samples (orange) and the Healthy events samples (light blue).

The first preprocessing steps for this research have been formerly inspired by the study of Li, Wang et al. 2022. The audio recordings are segmented based on their labeling, where the segmentation is created at beginning of the event in the time domain. The set duration of the event is selected in accordance with the sparsity of the data, meaning that only events of a certain duration are selected for comparison. This is an important choice since it would be pointless to try to compare data entries that have completely different lengths. However, even if the events time frame is restricted, there is no real fixed length for those extracted episodes, therefore zero-padding is applied for filling the time differences between the events. This procedure also provides further help toward solving the class imbalance issue since most of the healthy events are longer than 1000ms meanwhile the adventitious events are tendentially shorter than that. As demonstrated in the image below:

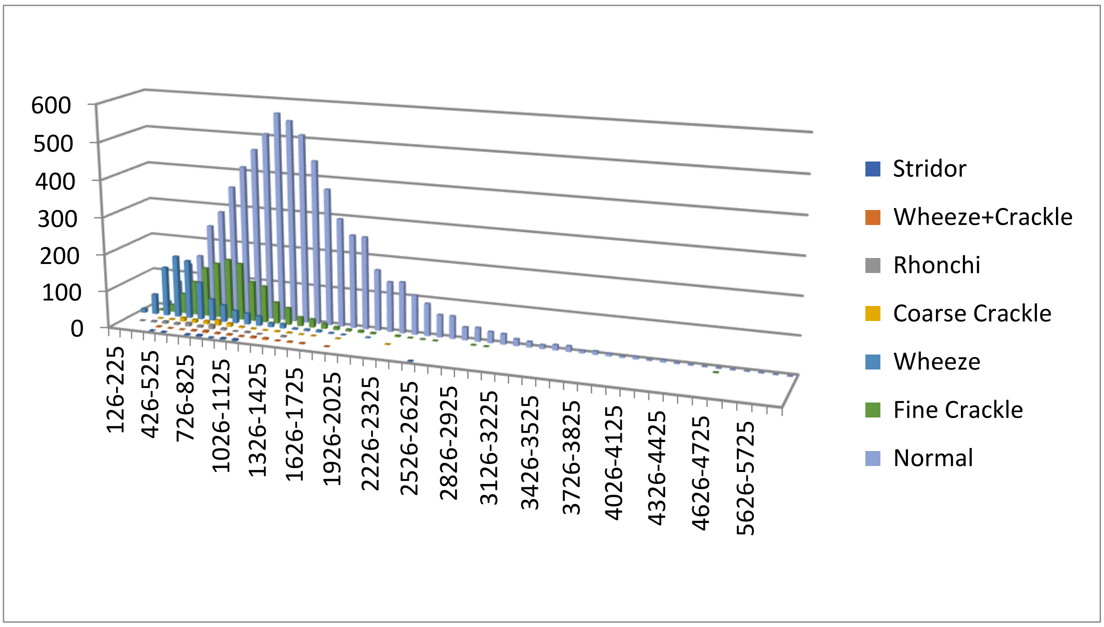

Figure 17. The distribution of the data based on the duration of the events and class division.

This decision enables the study to confront the events despite their different length in the time domain. As in every data science analysis, the data must be scaled to an extent that the line of similarity could be enough to enable statistical investigation. Another strategy to address the class imbalance problem is to limit furtherly the time sample by applying a threshold not only on the max amount of time for the selected segments, but also for the minimum amount of time. This creates a partition of the database that can further guarantee a less uneven ratio between the classes for latter comparison by establishing a similarity between the analyzed samples and minimizing the disturbance of the zero padding. In addition, since the overall number of healthy samples is still quantitatively bigger than the adventitious ones, a random but reproducible “drop” of the extra samples is adopted to finally balance the two classes, like in the picture below.

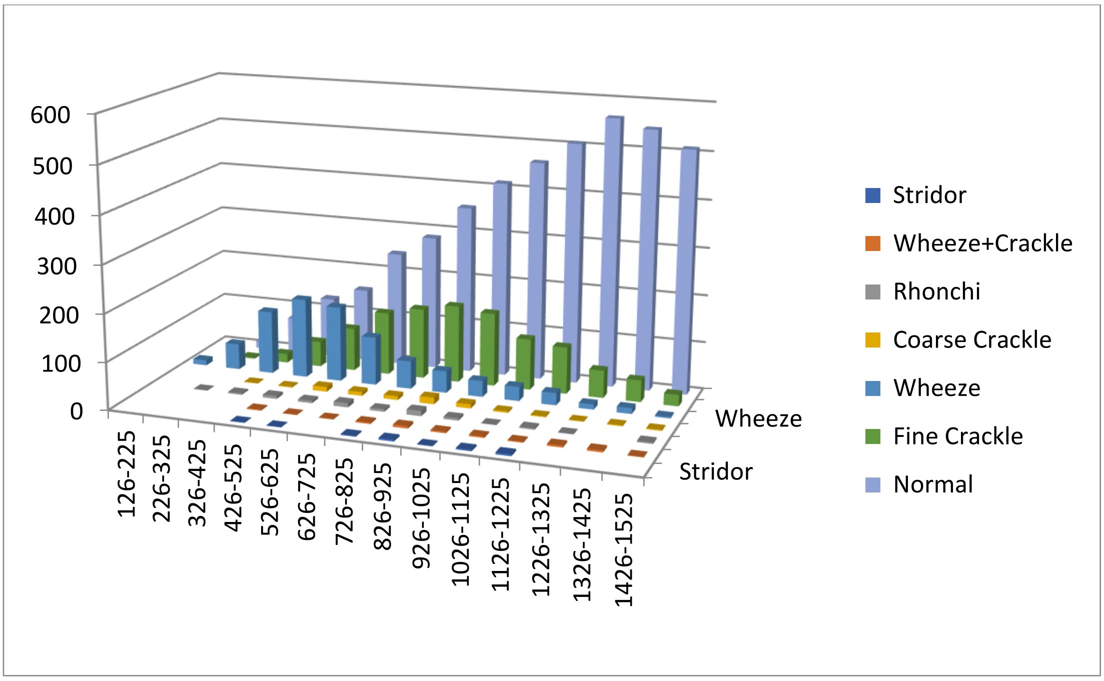

Figure 18. The distribution of the data based on the duration of the events and class division with focus on the part of the database with most concentration of adventitious sounds.

The different adventitious classes have been divided into different time windows depending on the amount of sample present within the chosen boundaries and with consideration for trying to avoid removing too many samples. For each adventitious class selection, there is an equal amount of samples, of random extraction, of the healthy class for comparison. All three classifiers were tested on each adventitious sound classification and each adventitious sound was tested binarily against the healthy counterparts. Additionally, a further comparison of the results in offered against Loudness Normalized recordings and not normalized ones. As shown in the image below, there is a noticeable amount of discrepancy between the main two classes, Fine Crackles & Wheezes, against the other four adventitious classes. As mentioned before, to face the class imbalance problem, each class has been faced with the same exact number of healthy samples that were selected also from the same time window. The following picture shows how many samples were used for each respective classification.

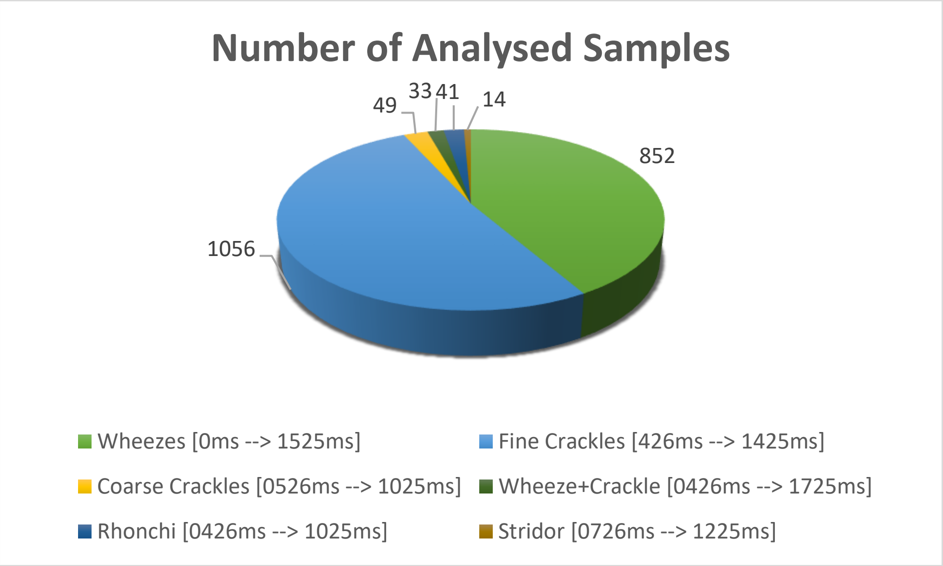

Figure 19. A Pie Chart of the number of samples that have been included in the classification task.

As also shown below, this decision is reinforced also by the enormous number of healthy samples that can be used for the binary classification but should not be necessarily used all at the same time.

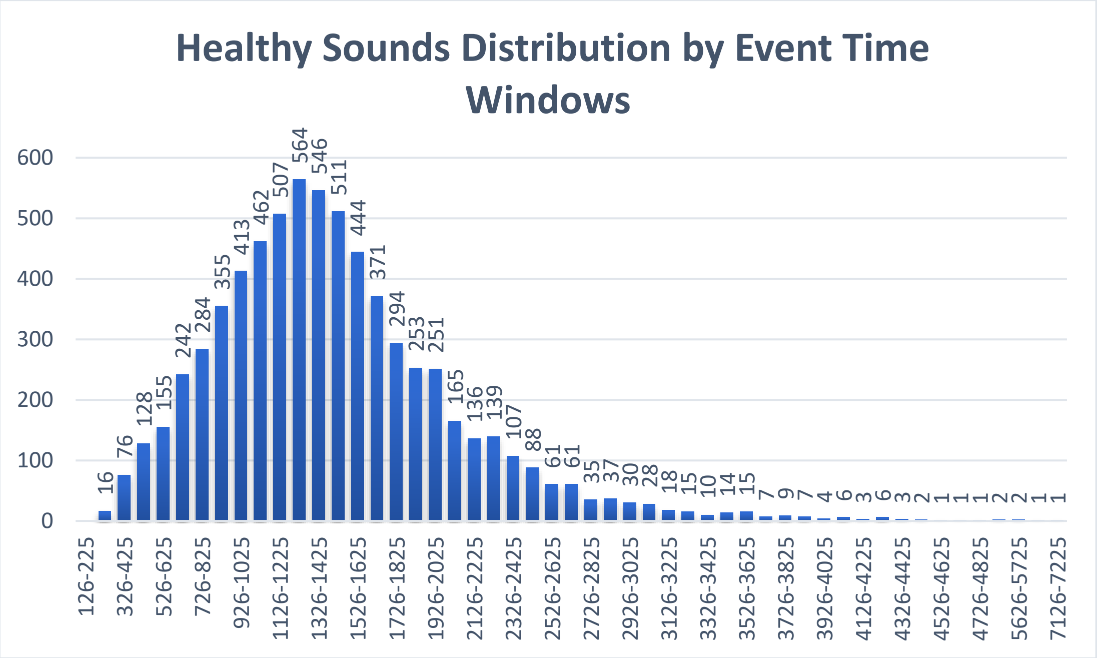

Figure 20. The distribution of Healthy events by time window.

The following plot visualizes that the most reliable number of selected samples for the classification are indeed belonging to the first two classes, namely Wheezes and Fine Crackles. Whilst the other classes exhibit a limited availability of samples, it was still possible to compare them successfully thanks to the even approach toward class imbalance which included the same exact number of samples from the healthy class that also belonged to the same time frame.

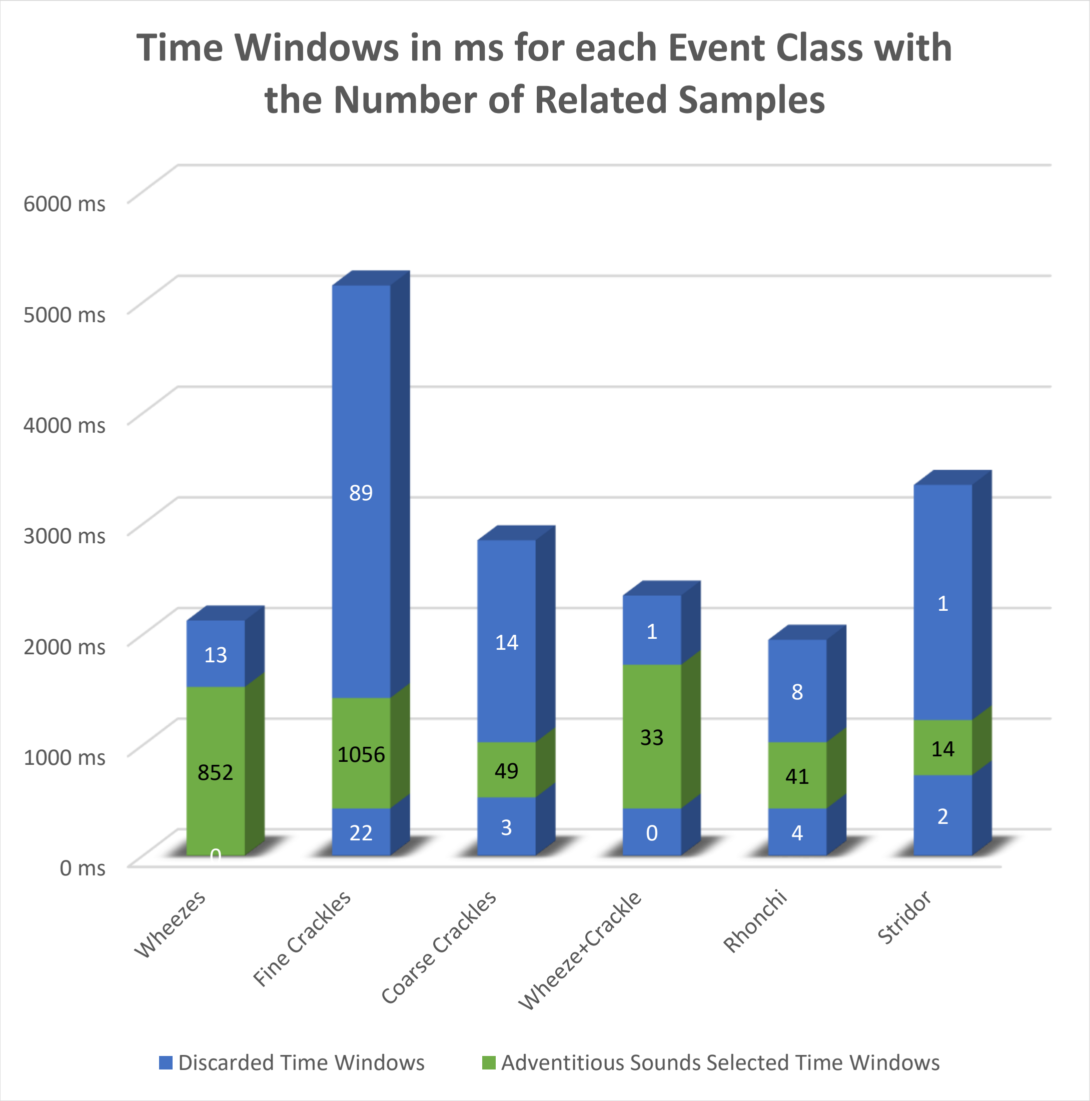

Figure 21. A visualization of the selected time windows for the analysis of each of the adventitious events (green bar), the excluded time frame where both adventitious and healthy events were present (light blue bar), and the excluded time frame where only healthy events were present (dark blue bar). The numbers on the green bars indicate the number of events that were selected for the classification from both classes, the numbers on the light blue bars indicate the number of events from the adventitious class that were excluded from the classification, and the numbers on the dark blue bars indicate the number of healthy events that were not considered for the classification. The last "healthy" bar serves as a reference for understanding the magnitude of the data. Time is expressed in milliseconds (ms).

3 RESULTS

The results are presented in the form of tabular representation, where all the respective adventitious sounds are scored with the related classification algorithm, and also the comparison with the loudness normalization preprocessing performance is offered. Visual bar graphs containing the scores of each class are also offered to ensure an easier understanding of the overall performance. Also, each table presents the results for the related metrics, namely (F1-Score, AUC, and PRC) the same ones that were digressed within the basics introductory part of this study. The last result is an averaged performance metric of all three scores. Additionally, the scores of the Deep Neural Network have been computed as an average of three different trials so that the lack of reproducibility could be minimized whilst the other classifiers results have been kept reproducible by the random state.

Classification Scores vs Healthy Class	DNN F1	SVM F1	DT F1
Wheezes	**87.49%**	**88.57%**	**80.00%**
Wheezes [LN]	86.85%	88.02%	79.29%
Fine Crackles	**77.66%**	**78.34%**	68.47%
Fine Crackles [LN]	74.90%	76.08%	**70.21%**
Coarse Crackles	60.95%	**72.00%**	**73.68%**
Coarse Crackles [LN]	**74.56%**	47.62%	33.33%
Wheeze+Crackle	**73.58%**	**84.21%**	73.68%
Wheeze+Crackle [LN]	62.38%	53.33%	**85.71%**
Rhonchi	93.54%	**96.00%**	88.00%
Rhonchi [LN]	**93.82%**	92.31%	**92.31%**
Stridor	**93.33%**	**100.00%**	66.67%
Stridor [LN]	71.11%	80.00%	**100.00%**
Average F1-Score	81.09%	**86.52%**	75.08%
Average F1-Score [LN]	**77.27%**	72.89%	76.81%

Table 1. The F1-Score results. [LN] is the abbreviation for Loudness Normalization, DNN is Deep Neural Network, SVM is Support Vector Machine, and DT is the Decision Tree. The numbers in "bold" are marked as the higher classifier scores for the adventitious class, whilst the "bold + highlight" are the best scores for that class.

Adventitious Classes	DNN AUC	SVM AUC	DT AUC
Wheezes	**91.30%**	**88.30%**	**79.84%**
Wheezes [LN]	90.32%	87.32%	79.62%
Fine Crackles	**82.15%**	**77.76%**	66.82%
Fine Crackles [LN]	78.96%	75.13%	**69.49%**
Coarse Crackles	59.33%	**65.00%**	**75.00%**
Coarse Crackles [LN]	**75.33%**	45.00%	40.00%
Wheeze + Crackle	**81.48%**	**74.44%**	58.89%
Wheeze + Crackle [LN]	58.15%	56.67%	**70.00%**
Rhonchi	95.51%	**96.15%**	79.81%
Rhonchi [LN]	**98.08%**	83.65%	**83.65%**
Stridor	**100.00%**	**100.00%**	75.00%
Stridor [LN]	**100.00%**	87.50%	**100.00%**
Average AUC Score	**84.96%**	83.61%	72.56%
Average AUC Score [LN]	**83.47%**	72.54%	73.79%

Table 2. The AUC Score results. [LN] is the abbreviation for Loudness Normalization, DNN is Deep Neural Network, SVM is Support Vector Machine, and DT is the Decision Tree. The numbers in “bold” are marked as the higher classifier scores for the adventitious class, whilst the “bold + highlight” are the best scores for that class.

Adventitious Classes	DNN PRC	SVM PRC	DT PRC
Wheezes	88.99%	87.57%	**77.97%**
Wheezes [LN]	**90.47%**	**89.27%**	75.71%
Fine Crackles	**78.45%**	**78.34%**	**70.05%**
Fine Crackles [LN]	76.76%	76.96%	**70.05%**
Coarse Crackles	58.37%	**90.00%**	**70.00%**
Coarse Crackles [LN]	**68.11%**	50.00%	30.00%
Wheeze + Crackle	**89.80%**	**88.89%**	77.78%
Wheeze + Crackle [LN]	63.04%	33.57%	**100.00%**
Rhonchi	98.81%	**92.31%**	84.62%
Rhonchi [LN]	**99.43%**	**92.31%**	**92.31%**
Stridor	**100.00%**	**100.00%**	50.00%
Stridor [LN]	**100.00%**	**100.00%**	**100.00%**
Average PRC Score	85.74%	**89.52%**	71.73%
Average PRC Score [LN]	**82.97%**	73.68%	78.01%

Table 3. The AUC Score results. [LN] is the abbreviation for Loudness Normalization, DNN is Deep Neural Network, SVM is Support Vector Machine, and DT is the Decision Tree. The numbers in “bold” are marked as the higher classifier scores for the adventitious class, whilst the “bold + highlight” are the best scores for that class.

Classification Scores vs Healthy Class	DNN	SVM	DT
F1-Score	**81.09%**	**86.52%**	75.08%
F1-Score [LN]	77.27%	72.89%	**76.81%**
AUC Score	**84.96%**	**83.61%**	72.56%
AUC Score [LN]	83.47%	72.54%	**73.79%**
PRC Score	**85.74%**	**89.52%**	71.73%
PRC Score [LN]	82.97%	73.68%	**78.01%**
Average Total Score	**83.93%**	**86.55%**	73.13%
Avarage Total Score [LN]	81.24%	73.04%	**76.20%**

Table 4. The total score results. [LN] is the abbreviation for Loudness Normalization, DNN is Deep Neural Network, SVM is Support Vector Machine, and DT is the Decision Tree. The numbers in "bold" are marked as the higher classifier scores for the adventitious class, whilst the "bold + highlight" are the best scores for that class.

4 DISCUSSION

This study went through several changes during its developmental course. Firstly, the former approach to the analysis of the adventitious sounds was inspired by the 2016 research of Grønnesby, which involved the segmentation of the audio signals with an overlapping of 50% between segments so that the adventitious events would not be missed. This method has shown some promising results with cherry-picked data but as soon as it was adopted on a larger amount of data it became immediately unreliable due to the incapacity of the algorithms to differentiate between the classes almost at all. Probably, the main reason for such an outcome lies in statistics and is fueled by human inaccuracy. In a perfect scenario, the adventitious events would be labeled at their exact location in the time domain rather than "just" on the general breathing cycle location. As Grønnesby states, the crackles are an evasive phenomenon that occurs in very restricted time windows (~92ms), therefore, it is highly improbable that someone who is not specifically looking for such time windows will not label them locally. As the evidence demonstrates, the labeling of this database is pursued in a more general manner. If the medical doctors detected any clue for the specific adventitious class to be present, the whole breathing cycle will be labeled as that class. Specifically for crackles, every local event becomes consequentially also potentially surrounded by healthy sounds that are still labeled as adventitious. Also, it is impossible to consider the labeling conducted by human doctors as 100% accurate for the many reasons mentioned in the introduction of this study. Even if we assume that the doctors managed to achieve an outstanding 80/90% of precision for the windows, that 10/20% of mislabeled information would still confuse the algorithm up to a point that it would be virtually impossible for it to differentiate the two classes when this situation is furtherly reinforced by a drastic class imbalance. This is also the reason why this believes that the classification of adventitious sounds should be differentiated by the type of sounds, since every adventitious class has distinct patterns in both time and frequency domains, therefore, mixing them all up together will not improve the classification accuracy.

Grønnesby was aware of this scenario and attempted to reduce this limitation by manually fixing the labels, therefore, creating his own dataset. However, to ensure a higher degree of reliability for the study, such a procedure would require the additional participation of external and unbiased medical personnel since otherwise the results could be unethically weighted in favor of the performance of the database. Additionally, despite the researcher of this study is a former Registered Nurse that can perform pulmonary auscultation, he lacks the capacity to establish a diagnosis, therefore, if he would have decided to pursue the manual labeling of the database its validity would have been still technically questionable. Fortunately, this study was started in the year 2021 as an academic multi-stage project and the available tools and online repositories are indeed higher in number than what Grønnesby had in 2016. This brought the study toward a different approach that did not limit manually the number of healthy samples. The new approach proceeded on a logical base of equal confrontation (50% / 50%), with random extraction for the excessive unnecessary samples, and all based on compatible time windows. Specifically, the extraction is aimed at the single labeled events so that the unlabeled and non-usable data would be excluded from the classification task. Also, the randomicity of the extraction frees the study from any bias that could belong to the notorious cherry-picking procedures. This procedure itself limits drastically the noise in the data, however, it is still not enough to ensure that the classification might be pursued correctly.

This new approach was inspired by the more recent study from 2022 made by Li et al., which applied a selection of single events based on fixed time-event windows and with zero padding of the shorter time frames. Unfortunately, the zero-padding means that also more artificial factors are introduced within the data. To limit the amount of noise generated by the artificial padding, this study opted to limit the time frame of each class evaluation based also on the evidence of the precedent study from Grønnesby which suggested that each class of adventitious event, by its morphological nature, belongs to a certain time frame. As already mentioned, the crackles belong to a very strict time window and that might be because they often are related to an alteration of the tissual surface of the lungs (generally overproduction of mucus), whilst the wheezes are a longer event because they are generally caused by a narrowing of the breathing structure (like asthma, or COPD). Consequently, the crackles can be accurately detected only in the tiny time frame where it is possible to hear the “popping” sound, whilst the wheezes are a constant phenomenon that can potentially alter the signal to longer extents.

This concept might also be reinforced not only but the lack of high performance for the Fine Crackles class, but also by the fact that the standard deviation for each adventitious class is tendentially higher for those classes than the healthy one. This notion is concordant with Grønnesby's findings and for the specific case of the Fine Crackles might give an additional hint toward the potential inadequate labeling procedure for that specific type of event. However, unless an unbiased third-party medical doctor were available to locally assess the labels, it is doubtful that any sort of further narrowing of the time window could potentially improve the results. In fact, this would limit even more the already mixed-up training data even more and compromise the analysis further for the classifier.

Features StDev	Healthy	Wheezes	Healthy	Fine Crackles
Variance	6.01627E-05	0.000350909	0.000102451	6.91002E-05
SMA_Coarse	4.15654E-06	6.8885E-06	4.5526E-06	6.43727E-06
Range	0.062044398	0.113273477	0.064098783	0.07710193
Spectrum_Mean	0.087335403	0.220413072	0.091710524	0.099288822
SMA_Fine	9.54836E-05	0.000202328	0.000144244	0.000105573
ZCR	8.206145433	10.0882691	7.895314806	7.352613911
Spec_Kurtosis	17.96588833	43.46806442	26.91753162	17.85382692
RMS	0.00235077	0.006795876	0.002900665	0.002665376
Features StDev	**Healthy**	**Coarse Crackles**	**Healthy**	**Crackles + Wheezes**
Variance	4.91105E-05	0.000167289	1.48638E-05	0.000140158
SMA_Coarse	5.72417E-06	6.89211E-06	3.83301E-06	5.67807E-06
Range	0.103819163	0.129673544	0.062729347	0.123533713
Spectrum_Mean	0.101057653	0.194450988	0.061617384	0.236441291
SMA_Fine	0.000104085	0.000333208	6.98787E-05	0.000284919
ZCR	7.852502808	9.051745208	9.307769867	10.42920566
Spec_Kurtosis	15.70124756	19.0991872	22.24190706	15.33626017
RMS	0.002953392	0.006253564	0.001736493	0.006264987
Features StDev	**Healthy**	**Rhonchi**	**Healthy**	**Stridor**
Variance	1.09834E-05	0.000180614	1.01478E-05	0.001911915
SMA_Coarse	3.97867E-06	1.85303E-05	3.86328E-06	4.97197E-05
Range	0.026503678	0.146834942	0.017545363	0.319123028
Spectrum_Mean	0.04955124	0.182378078	0.04858007	0.742471822
SMA_Fine	8.3151E-05	0.000413027	8.89814E-05	0.000579444
ZCR	7.609061081	6.403893492	7.246125788	8.826240553
Spec_Kurtosis	15.82924232	118.0437031	19.56446866	44.52597774
RMS	0.001396778	0.006388185	0.00139012	0.021302794

Table 5. The standard deviations of all eight features are divided by class while, in green, the highest values on each row are highlighted respectively.

As shown in the table, the Fine Crackles are the only class that has a predominantly lower amount of standard deviation than the healthy samples. This might be due to either a presence of undetected crackles within the healthy samples (also healthy people can have crackles), or just the too-large local window that integrates healthy samples into the adventitious class, or perhaps even a combination of the two. One could also think that such behavior might be typical for the specific class, however, since all the other adventitious classes exhibit the same pattern, it is most probable that the reasons behind this behavior are tied to the labeling instead. This is the reason why this study selected different time frames for each adventitious class. Additionally, this study formerly tried to solve the class imbalance problem by adopting random sampling generator techniques such as SMOTE, SMOTETomek, and Weighted Random Sampling directly on the classifier. However, those methods, despite being fast in their deployment and slightly improving the results, still did not guarantee a satisfactory outcome. Regardless, using random sampling involves the creation, or more specifically, the replication of already existing data points that would still be artificial, therefore, not entirely reliable for this type of medical data analysis. These methods have still been kept within the code used for the classification, however, due to the new nature of the preprocessing via time windows and equal number sampling, their outcome is overall nullified. The question that we should ask ourselves in this regard is: would you trust a medical machine learning algorithm that has been trained on the artificial reproduction of the same 14 data points (in the case of the stridor class), or would it be more reliable to have instead that x number of data points from real samples instead? Regardless of the answer, that little number of samples can still be used to understand more about the nature of the signal and its unique features.

As all the metrics show, there is a general tendency for the algorithm to perform better on the smaller samples, specifically the Stridor and Rhonchi. This might be due to the diversity of these sounds when paired with healthy samples, or that even the scarcity of available samples induced a higher degree of accuracy for the labeling of these events. Overall, the SVM (86.55%) seems to outperform both the DT (76.20%) and the DNN (83.93% without LN) in the average total score, however, when applied to Loudness Normalized events, the SVM (73.04%) does not seem to benefit at all from that type of preprocessing. On the other hand, the DT (73.13% without LN) seems overall to benefit highly from the Loudness Normalization whilst the DNN (81.24%) does not but is still not up to the same pejorative degree as the SVM. However, it is also important to mention that in terms of computational efficiency, the SVM is the algorithm that requests more power and time, whilst the DT, despite giving the lowest results, is indeed the fastest and easiest algorithm of the three, and the DNN sits in between the two. The Loudness Normalization does not seem to directly improve all kinds of classification, but this might be due to the disparity of the overall data, which might lead to the introduction of additional noises. Such factors might not be as easily compensated by the whole normalization process, especially when the data is processed by complex algorithms, whilst the overall effect seems still beneficial for the simpler classifiers. It is also important to denote that the complexity of the algorithm might even inhibit it from processing too big databases. For example, in the case of classifying the whole database, the SVM would have not been able to finish the task in a reasonable amount of time, whilst both the DT and DNN could have guaranteed the results anyway. The following plot shows how effective the combination of each feature proved to be for the classification of the wheezes sounds within the Support Vector Machine.

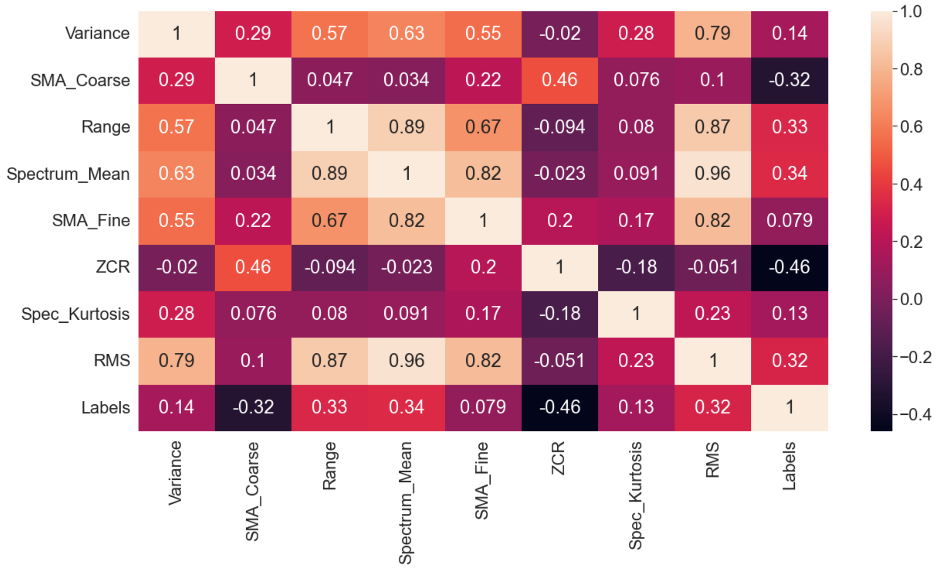

Figure 22. The correlation plot for the variables’ association for the Wheezes classification.

Compared with Grønnesby’s research results, this study shows an overall slightly lower, but still acceptable, F1-Score (83.5% vs 78.34). On the other hand, the results for this study on the other adventitious sounds, particularly for the wheezes, show very promising results (88.57%) compared to Grønnesby’s (63.7%). Unfortunately, both the SJTU publication and the study from Li et al. used a different set of methods for the feature extraction and different metrics for the evaluation of the results, therefore, it would be impossible to compare them directly with this study. Another difference is that neither Grønnesby nor Li et al. codes were available, meanwhile, this study pursues the importance of reproducibility so the code will remain available on GitHub (check appendix). Regardless, this research would have never been possible without the inspiration and knowledge these studies provided.

However, despite the limited availability of the data for the minority classes, and the non-specific labeling of the events such as crackles, this study still led to more than acceptable results, which, could be furtherly improved by future studies if they the researchers will also consider these findings. Perhaps, the labeling of the next hypothetical pulmonary sounds database should be executed not just by medical doctors, but machine learning experts might also participate and supervise or instruct the doctors on what are the principles behind the automated classification of the lung sounds, therefore, identify and satisfy the needs of the machine to improve the potential further performance and put an important step in science toward the digitalization of healthcare.

4.1 Reproducibility

To encourage the pursuit of future research and enable complete reproducibility, all the code used for this study remains available at the following link: https://github.com/vradx/ACLSvr/tree/main

5 CONCLUSION

This study's findings suggest that Artificial Intelligence can establish a reliable pulmonary sound classification based on digital stethoscope recordings as long as certain requirements are met.

- Pulmonary recordings should be labelled as precisely as possible, whilst the machine learning algorithm should be provided with enough samples for each class, and class imbalance should always be minimized.
- One-Dimensional features can be used for this typology of classification as they prove an easier and quicker alternative to more complex Two-Dimensional features such as Spectrograms.
- Loudness normalization, as a pre-processing tool, can be beneficial for the detection of certain classes and particularly with Decision Trees. However, on average it lowers the overall performance of the main performing algorithms, therefore, it should be furtherly investigated.
- Support Vector Machines can achieve the highest degree of precision at the cost of computational effort, whilst Decision Trees can be used for an early understanding of the data, and Deep Neural Networks can be more computationally efficient than SVMs but do not guarantee the same degree of classification accuracy.
- To avoid data overlapping, class labelling should be conducted with the most achievable precision, especially for the short-timed events such as the crackles.
- Biomedical signals can contain a lot of information, therefore, only the most relevant parts of the pulmonary recordings should be included within the classification to avoid ”confusing” the algorithm.

Overall, Artificial Intelligence could be implemented soon as a support tool for the training of healthcare professionals. Once it becomes more stable in its results and the standards for its development and deployment will be determined, it might even potentially be used as an augmentative tool to support healthcare professionals in their daily tasks. Until then, more studies should be pursued to ascertain the reliability of AI’s performance for this type of task, nevertheless, future research should be kept completely reproducible to ensure the scientific community's progress by enabling future studies to access not only the results but also the exact techniques that have been developed for the task. Also, to ensure the future success of these types of tasks, artificial intelligence experts should ideally collaborate with the doctors for the acquisition and labeling of the data since their knowledge of the machines' decision criteria can prove fundamental for the future success of the classification tasks and, therefore, the creation of a standard for the whole acquisition process.

6 LIMITATIONS

This study has been conducted as a “learning” project by a single student during his master’s degree four semesters, therefore, the technical limitations of the latter are directly inherent to the development process of his scientific skills. As learning is both a passive and active process, certain decisions for the methodologies of this study could have been affected by the general ongoing acquisition of new knowledge concerning the topic. Additionally, since this study tried to adopt the methodologies of other studies whose code remains unpublished, the reproduction of those approaches can only be guaranteed at the best possible guessing and interpretation of the scientific words.

REFERENCES

References

[1] Rocha, B.M., Filos, D., Mendes, L.*, et al.*: 'A Respiratory Sound Database for the Development of Automated Classification', in Maglaveras, N., Chouvarda, I., Carvalho, P. de (Eds.): 'Precision Medicine Powered by pHealth and Connected Health' (Springer Singapore, Singapore, 2018), pp. 33–37

[2] Zhang, Q., Zhang, J., Yuan, J.*, et al.*: 'SPRSound: Open-Source SJTU Paediatric Respiratory Sound Database', *IEEE transactions on biomedical circuits and systems*, 2022, **16**, (5), pp. 867–881

[3] Ma, Y., Xu, X., Yu, Q.*, et al.*: 'LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using bi-ResNet Deep Learning Algorithm'. 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17/10/2019 - 19/10/2019, pp. 1–4

[4] Alshmrani, G.M.M., Ni, Q., Jiang, R., Pervaiz, H., Elshennawy, N.M.: 'A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images', *Alexandria Engineering Journal*, 2023, **64**, pp. 923–935

[5] Geddes, D.: 'The history of respiratory disease management', *Medicine (Abingdon, England : UK ed.)*, 2020, **48**, (4), pp. 239–243

[6] Demir, F., Sengur, A., Bajaj, V.: 'Convolutional neural networks based efficient approach for classification of lung diseases', *Health information science and systems*, 2020, **8**, (1), p. 4

[7] Abbas, A., Fahim, A.: 'An automated computerized auscultation and diagnostic system for pulmonary diseases', *Journal of medical systems*, 2010, **34**, (6), pp. 1149–1155

[8] Zhu, H., Lai, J., Liu, B.*, et al.*: 'Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: A cohort study', *Computer methods and programs in biomedicine*, 2022, **213**, p. 106500

[9] Aviles-Solis, J.C., Jácome, C., Davidsen, A.*, et al.*: 'Prevalence and clinical associations of wheezes and crackles in the general population: the Tromsø study', *BMC pulmonary medicine*, 2019, **19**, (1), p. 173

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

[11] Mahadevkar, S.V., Khemani, B., Patil, S.*, et al.*: 'A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions', *IEEE Access*, 2022, **10**, pp. 107293–107329

[12] Lalouani, W., Younis, M., Emokpae, R.N., Emokpae, L.E.: 'Enabling effective breathing sound analysis for automated diagnosis of lung diseases', *Smart health (Amsterdam, Netherlands)*, 2022, **26**, p. 100329

[13] Kumar, Y., Koul, A., Singla, R., Ijaz, M.F.: 'Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda', *Journal of ambient intelligence and humanized computing*, 2022, pp. 1–28

[14] Speizer FE, Horton S, Batt J, et al.: 'Disease Control Priorities in Developing Countries. 2nd edition.' (2006)

[15] Barry, S.J., Dane, A.D., Morice, A.H., Walmsley, A.D.: 'The automatic recognition and counting of cough', *Cough (London, England)*, 2006, **2**, p. 8

[16] Shen, Y.-T., Chen, L., Yue, W.-W., Xu, H.-X.: 'Artificial intelligence in ultrasound', *European journal of radiology*, 2021, **139**, p. 109717

[17] 'Respiratory sound analysis as a diagnosis tool for breathing disorders.', https://shura.shu.ac.uk/id/eprint/24966

[18] Suzuki, K., Shimizu, Y., Ohshimo, S.*, et al.*: 'Real-time assessment of swallowing sound using an electronic stethoscope and an artificial intelligence system', *Clinical and experimental dental research*, 2022, **8**, (1), pp. 225–230

[19] Sarkar, M., Madabhavi, I., Niranjan, N., Dogra, M.: 'Auscultation of the respiratory system', *Annals of thoracic medicine*, 2015, **10**, (3), pp. 158–168

[20] Grønnesby, M.: 'Automated Lung Sound Analysis'. Master's thesis, UiT The Arctic University of Norway, 2016

[21] Reichert, S., Gass, R., Brandt, C., Andrès, E.: 'Analysis of respiratory sounds: state of the art', *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2008, **2**, pp. 45–58

[22] Hsu, F.-S., Huang, S.-R., Huang, C.-W.*, et al.*: 'An Update on a Progressively Expanded Database for Automated Lung Sound Analysis' (arXiv, 2021)

[23] Li, J., Wang, X., Wang, X., Qiao, S., Zhou, Y.: 'Improving The ResNet-based Respiratory Sound Classification Systems With Focal Loss', in : '2022 IEEE Biomedical Circuits 2022', pp. 223–227

[24] Fukuda, T., Ichikawa, O., Nishimura, M.: ‘Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition’, *Speech Communication*, 2018, **98**, pp. 95–103

[25] Obaid, H.S., Dheyab, S.A., Sabry, S.S.: ‘The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning’, in : ‘2019 9th Annual Information Technology 2019’, pp. 279–283

[26] ‘Loudness Normalization in Accordance with EBU R 128 Standard - MATLAB & Simulink - MathWorks United Kingdom’, https://uk.mathworks.com/help/audio/ug/loudness-normalization-in-accordance-with-ebu-r-128-standard.html, accessed May, 2023

[27] ‘Audio normalization’, https://en.wikipedia.org/w/index.php?title=Audio_normalization&oldid=1123057462, accessed May, 2023

[28] ‘Auphonic Blog: Loudness Normalization and Compression of Podcasts and Speech Audio’, https://auphonic.com/blog/2011/07/25/loudness-normalization-and-compression-podcasts-and-speech-audio/, accessed May, 2023

[29] Salau, A.O., Jain, S.: ‘Feature Extraction: A Survey of the Types, Techniques, Applications’. 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 158–164

[30] ‘How to Use the Numpy Variance Function’, accessed April, 2023

[31] Arie, R., Brand, A., Engelberg, S.: ‘Compressive sensing and sub-Nyquist sampling’, *IEEE Instrum. Meas. Mag.*, 2020, **23**, (2), pp. 94–101

[32] ‘Discrete cosine transform’, https://en.wikipedia.org/w/index.php?title=Discrete_cosine_transform&oldid=1148536269, accessed April, 2023

[33] Arvid Trapp, P.W.: ‘2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC): Proceedings : 14-17 December 2021, Tokyo, Japan’ (IEEE, Piscataway, NJ, 2021)

[34] ‘Spectral kurtosis from signal or spectrogram - MATLAB pkurtosis’, https://www.mathworks.com/help/signal/ref/pkurtosis.html, accessed April, 2023

[35] Kathirvel, P., Sabarimalai Manikandan, M., Senthilkumar, S., Soman, K.P.: ‘Noise robust zerocrossing rate computation for audio signal classification’. Computing (TISC), Chennai, India, 12/8/2011 - 12/9/2011, pp. 65–69

[36] Albu, M., Heydt, G.T.: ‘On the use of rms values in power quality assessment’, *IEEE Trans. Power Delivery*, 2003, **18**, (4), pp. 1586–1587

[37] Keim, R.: ‘How Standard Deviation Relates to Root-Mean-Square Values’, *All About Circuits*, 29/07/2020

[38] ‘Root-Mean-Square -- from Wolfram MathWorld’, https://mathworld.wolfram.com/Root-Mean-Square.html, accessed April, 2023

[39] Damilola, S.: ‘A Review of Unsupervised Artificial Neural Networks with Applications’, *IJCA*, 2019, **181**, (40), pp. 22–26

[40] Rudin, C., Radin, J.: ‘1.2’, *Harvard Data Science Review*, 2019, **1**, (2)

[41] Loh, W.-Y.: ‘Fifty Years of Classification and Regression Trees’, *International Statistical Review*, 2014, **82**, (3), pp. 329–348

[42] Qomariyah, N.N., Heriyanni, E., Fajar, A.N., Kazakov, D.: ‘Comparative Analysis of Decision Tree Algorithm for Learning Ordinal Data Expressed as Pairwise Comparisons’. 2020 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 6/24/2020 - 6/26/2020, pp. 1–4

[43] Quinlan, J.R.: ‘Decision trees and decision-making’, *IEEE Trans. Syst., Man, Cybern.*, 1990, **20**, (2), pp. 339–346

[44] Al Hamad, M., Zeki, A.M.: ‘Accuracy vs. Cost in Decision Trees: A Survey’. 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain, 11/18/2018 - 11/20/2018, pp. 1–4

[45] Krzywinski, M., Altman, N.: ‘Classification and regression trees’, *Nat Methods*, 2017, **14**, (8), pp. 757–758

[46] ‘1.10. Decision Trees’, https://scikit-learn.org/stable/modules/tree.html#tree-mathematical-formulation, accessed April, 2023

[47] ‘Support Vector Machine Explained-Theory, Implementation, and Visualization | LinkedIn’, https://www.linkedin.com/pulse/support-vector-machine-explained-theory-visualization-zixuan-zhang/, accessed April, 2023

[48] Wang, Q.: ‘Support Vector Machine Algorithm in Machine Learning’. 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 6/24/2022 - 6/26/2022, pp. 750–756

[49] Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S.: ‘Activation Functions: Comparison of trends in Practice and Research for Deep Learning’ (08/11/2018)

[50] Author: Anthony Chang, MD, MBA, MPH, MS: ‘Is deep learning supervised or unsupervised?’, *AIMed*, 02/11/2021

[51] Mallick, S.: ‘Activation Functions in Deep Learning – A Complete Overview’, *Satya Mallick*, 30/10/2017

[52] Kanstrén, T.: ‘A Look at Precision, Recall, and F1-Score - Towards Data Science’, *Towards Data Science*, 11/09/2020

[53] ‘tf.keras.metrics.AUC &nbsp;|&nbsp; TensorFlow v2.12.0’, https://www.tensorflow.org/api_docs/python/tf/keras/metrics/AUC, accessed May, 2023

[54] Hsu, F., How, C.-H., Huang, S.-R., Chen, Y.-T., Chen, J.-S., Hsin, H.-T.: ‘Locating stridor caused by tumor compression by using a multichannel electronic stethoscope: a case report’, *Journal of clinical monitoring and computing*, 2021, **35**, (3), pp. 663–670

[55] Hamilton-Fairley, D., Coakley, J., Moss, F.: ‘Hospital at night: an organizational design that provides safer care at night’, *BMC medical education*, 2014, **14 Suppl 1**, (Suppl 1), S17

[56] Pramono, R.X.A., Bowyer, S., Rodriguez-Villegas, E.: ‘Automatic adventitious respiratory sound analysis: A systematic review’, *PloS one*, 2017, **12**, (5), e0177926

[57] Wu, Y.-C., Han, C.-C., Chang, C.-S.*, et al.*: ‘Development of an Electronic Stethoscope and a Classification Algorithm for Cardiopulmonary Sounds’, *Sensors (Basel, Switzerland)*, 2022, **22**, (11)

[58] Profant, O., Tintěra, J., Balogová, Z., Ibrahim, I., Jilek, M., Syka, J.: ‘Functional changes in the human auditory cortex in ageing’, *PloS one*, 2015, **10**, (3), e0116692

[59] Liu, X.Z., Yan, D.: ‘Ageing and hearing loss’, *The Journal of pathology*, 2007, **211**, (2), pp. 188–197

[60] Tortora, G.J., Derrickson, B.: ‘Tortora's principles of anatomy and physiology’ (Wiley, Hoboken, New Jersey, 2017)

[61] ‘Human ear - Transmission of sound within the inner ear’, https://www.britannica.com/science/ear/Transmission-of-sound-within-the-inner-ear, accessed April, 2023

[62] Drew, T., Võ, M.L.-H., Wolfe, J.M.: ‘The Invisible Gorilla Strikes Again’, *Psychological science*, 2013, **24**, (9), pp. 1848–1853

[63] Langner, R., Eickhoff, S.B.: ‘Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention’, *Psychological bulletin*, 2013, **139**, (4), pp. 870–900

[64] Laffere, A., Dick, F., Holt, L.L., Tierney, A.: ‘Attentional modulation of neural entrainment to sound streams in children with and without ADHD’, *NeuroImage*, 2021, **224**, p. 117396

[65] Lin, T.-H., Tsao, Y.: ‘Source separation in ecoacoustics: a roadmap towards versatile soundscape information retrieval’, *Remote Sens Ecol Conserv*, 2020, **6**, (3), pp. 236–247

[66] Madore, K.P., Wagner, A.D.: ‘Multicosts of Multitasking’, *Cerebrum: the Dana Forum on Brain Science*, 2019, **2019**

[67] Suozzo, A.C., Malta, S.M., Gil, G., Tintori, F., Lacerda, S.S., Nogueira-Martins, L.A.: ‘Attention and memory of medical residents after a night on call: a cross-sectional study’, *Clinics*, 2011, **66**, (3), pp. 505–508

[68] ‘Stethoscope. Encyclopedia Britannica.’, https://www.britannica.com/technology/stethoscope

[69] ‘Cammann type binaural stethoscope | Science Museum Group Collection’, https://collection.sciencemuseumgroup.org.uk/objects/co90885/cammann-type-binaural-stethoscope-stethoscope, accessed April, 2023

[70] Park, D.E., Watson, N.L., Focht, C.*, et al.*: ‘Digitally recorded and remotely classified lung auscultation compared with conventional stethoscope classifications among children aged 1-59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study’, *BMJ open respiratory research*, 2022, **9**, (1)

[71] Fattah, S.A., Rahman, N.M., Maksud, A.*, et al.*: ‘Stetho-phone: Low-cost digital stethoscope for remote personalized healthcare’. 2017 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, 19/10/2017 - 22/10/2017, pp. 1–7

[72] ‘Piezoelectricity . Encyclopedia Britannica.’, https://www.britannica.com/science/piezoelectricity

[73] Leng, S., Tan, R.S., Chai, K.T.C., Wang, C., Ghista, D., Zhong, L.: ‘The electronic stethoscope’, *Biomedical engineering online*, 2015, **14**, p. 66

[74] Fraiwan, M., Fraiwan, L., Khassawneh, B., Ibnian, A.: ‘A dataset of lung sounds recorded from the chest wall using an electronic stethoscope’, *Data in brief*, 2021, **35**, p. 106913

[75] Radicchi, V.: ‘Technology and Nursing Education in Europe: A Literature Review’. Bachelor's thesis, Satakunta University of Applied Sciences, 2020

[76] Ondrisova, M., Molnar, L., Juhas, G., Juhasova, A., Mazari, J., Mladoniczky, M.: ‘Story of Artificial Intelligence: research and publications behind’. 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA), Starý Smokovec, Slovakia, 21/11/2019 - 22/11/2019, pp. 581–586

[77] Shalev-Shwartz, S., Ben-David, S.: ‘Understanding machine learning: From theory to algorithms’ (Cambridge University Press, New York NY USA, 2014)

[78] Haenlein, M., Kaplan, A.: ‘A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence’, *California Management Review*, 2019, **61**, (4), pp. 5–14

[79] Jones, L.D., Golan, D., Hanna, S.A., Ramachandran, M.: ‘Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern?’, *Bone & joint research*, 2018, **7**, (3), pp. 223–225

[80] Zakharov, V.: ‘About the Evolution of the Concept of “Artificial Intelligence”’. 2021 International Conference Engineering Technologies and Computer Science (EnT), Moscow, Russian Federation, 18/08/2021 - 19/08/2021, pp. 20–23

[81] Sennott, S.C., Akagi, L., Lee, M., Rhodes, A.: ‘AAC and Artificial Intelligence (AI)’, *Topics in language disorders*, 2019, **39**, (4), pp. 389–403

[82] Warwick, K., Nasuto, S.: ‘Historical and current machine intelligence’, *IEEE Instrum. Meas. Mag.*, 2006, **9**, (6), pp. 20–26

[83] Gentiluomo, L., Roessner, D., Augustijn, D*., et al.*: ‘Application of interpretable artificial neural networks to early monoclonal antibodies development’, *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik e.V*, 2019, **141**, pp. 81–89

[84] Lu, Y., Lu, J.: ‘A Universal Approximation Theorem of Deep Neural Networks for Expressing Probability Distributions’ (arXiv, 2020)

[85] Rugolon, F.: ‘Using counterfactuals to explain survival predictions in lung transplant recipients’. Master's Thesis, Karolinska Institutet, 2022

[86] Müller Andreas C. & Guido S.: ‘Introduction to Machine Learning with Python’ (2016, 1st edn.)

[87] Mostafa, B., El-Attar, N., Abd-Elhafeez, S., Awad, W.: ‘Machine and Deep Learning Approaches in Genome: Review Article’, *Alfarama Journal of Basic & Applied Sciences*, 2020, **0**, (0), p. 0

[88] ‘Database. Encyclopedia Britannica.’, https://www.britannica.com/technology/database

[89] ‘Dataset. Cambridge Dictionary’, https://dictionary.cambridge.org/dictionary/english/dataset

[90] Pradipta, G.A., Wardoyo, R., Musdholifah, A., Sanjaya, I.N.H., Ismail, M.: ‘SMOTE for Handling Imbalanced Data Problem : A Review’. 2021 Sixth International Conference on Informatics and Computing (ICIC), Jakarta, Indonesia, 11/3/2021 - 11/4/2021, pp. 1–8

KÄRNTEN
University of
Applied Sciences

AFFIDAVIT (DECLARATION OF ORIGINALITY)

I hereby declare that

- I have independently written the presented Bachelor/Master thesis by myself;
- I have prepared this Bachelor/Master thesis without outside help and without using any sources or aids other than those cited by me; moreover, I have identified as such any passages taken verbatim or in terms of content from the sources used;
- in addition, I have fully indicated the use of generative AI models (e.g. ChatGPT) by specifying the product name and the reference source (e.g. URL);
- I have not used any other unauthorized aids and have consistently worked independently and when using generative AI models, I realize that I am responsible how the content will be used and to what extent
- have not yet submitted this Bachelor/Master thesis in the same or similar form to any (other) educational institution as an examination performance or (scientific) thesis.

I am aware that any violation ("use of unauthorized aids") violates academic integrity and may result in (academic-related) legal consequences.

Klagenfurt, 07/06/2023

Date Signature