ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

DIAGNOPHONE: AN ELECTRONIC STETHOSCOPE FOR RESPIRATORY AUDIO ANALYSIS

M.Sc. THESIS

Ege YAĞ ÇAKIR

Department of Computer Engineering

Computer Engineering Programme

JUNE 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

DIAGNOPHONE: AN ELECTRONIC STETHOSCOPE FOR RESPIRATORY AUDIO ANALYSIS

M.Sc. THESIS

Ege YAĞ ÇAKIR
(504161510)

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Gökhan İNCE

JUNE 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DIAGNOPHONE: SOLUNUM SESİ ANALİZİ İÇİN BİR ELEKTRONİK STETESKOP TASARIMI

YÜKSEK LİSANS TEZİ

Ege YAĞ ÇAKIR
(504161510)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Gökhan İNCE

HAZİRAN 2019

Ege YAĞ ÇAKIR, a M.Sc. student of ITU Graduate School of ScienceEngineering and Technology 504161510 successfully defended the thesis entitled **"DIAGNOPHONE: AN ELECTRONIC STETHOSCOPE FOR RESPIRATORY AUDIO ANALYSIS"**, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Asst. Prof. Dr. Gökhan İNCE** Istanbul Technical University	....
Jury Members :	**Assoc. Prof. Dr. Sanem SARIEL UZER** Istanbul Technical University	....
	Assoc. Prof. Dr. M. Elif Karslıgil Yıldız Technical University	....

Date of Submission : **28 Haziran 2019**
Date of Defense : **12 June 2019**

FOREWORD

This thesis is made as a master project, as part of the requirements for the awarding of a degree in Master of Science in Engineering at the department of Computer Engineering at the Istanbul Technical University. I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Asst. Prof. Dr. Gökhan İNCE, my advisor for his countless hours of reflecting, reading, encouraging, and most of all patience throughout the entire process. I would like to also thank Dr. Erdem Anıl Çakır for his contributions in the user experience tests, help with the data collection and patiently answering all my questions.

June 2019

Ege YAĞ ÇAKIR
Computer Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
BMI	: Body Mass Index
CAD	: Computer Aided Diagnostics
CNN	: Convolutional Neural Networks
COPD	: Chronic Obstructive Pulmonary Disease
CT	: Computed Tomography
ECG	: Electrocardiogram
ER	: Emergency Room
FFT	: Fast Fourier Transform
GP	: General Practicioner
HCI	: Human Computer Interaction
ISO	: International Organization for Standardization
ITU	: Istanbul Technical University
K-NN	: K-Nearest Neighbors
LRTI	: Lower Respiratory Tract Infection
MFCC	: Mel Frequency Cepstral Coefficients
MRI	: Magnetic Eesonance Imaging
PSSUQ	: Post Study System Usability Questionnaire
SVN	: Support Vector Machines
VTLP	: Vocal Tract Length Perturbation
UI	: User Interface
URTI	: Upper Respiratory Tract Infection
UX	: User Experience

LIST OF TABLES

Page

LIST OF FIGURES

DIAGNOPHONE: AN ELECTRONIC STETHOSCOPE FOR RESPIRATORY AUDIO ANALYSIS

SUMMARY

Today, pulmonary diseases are one of the major causes of mortality in the world. According to the Turkish Institution of Statistics, there is a serious lack of expert doctors in Turkey trying to help people with such deadly diseases.

There are different tests available in order to diagnose an anomaly related to the lungs such as X-ray and tomography. Unfortunately, these types of equipment are not easy to find in every health clinic and they are both expensive and require time-consuming procedures for the patients. Because of the low availability of these types of equipment, patients cannot reach the results immediately because they have to wait a long time period for the queue to use them. In addition, for example, during MRI, the patient must remain idle for a long time which can be stressful for especially patients with claustrophobia. Furthermore, the patients who benefit from all these equipment are exposed to intense radiation during the use. In the tomography case, this rate is even greater.

As a result of the interviews conducted with many doctors, it is concluded that many diseases can be understood only by listening with a stethoscope, yet still, these tests are being applied by these kinds of equipment in order to validate the doctor's initial diagnosis.

Even though there are various diagnostic tests available, the stethoscope is still the first, cheapest and the most frequently used diagnosis device for the physicians. Therefore, in this thesis; a smart electronic stethoscope has been designed to help physicians with the identification of the anomaly and the diagnosis of the disease using Machine Learning. With this stethoscope, the lung sound can be heard and recorded at the same time. This feature can be used for teleconferencing when there is no specialist physician to be consulted, and also for storing the patient's audio data.

Nowadays, patients have access to the old test results (eg, blood tests, tomography results, etc.). Since data such as the previous lung sound is not stored in hospitals, the sound data cannot be used in the follow-up of the disease. However, in patients requiring follow-up such as asthma, storing the lung sound of the patient will facilitate the follow-up of the disease and will also help to determine the condition of the disease compared to the previous.

The audio recorded with Diagnophone can also be used in medical education. Today's medical education consists of listening to the lung sounds of the patients with the same stethoscope by the students, after the teachers, who are gathered at the beginning of the patient's bed. As a result of the user interviews, it was inferred that this was not an effective type of learning. However, the sounds recorded from the patient through Diagnophone can be played with the speaker of the mobile phone by the teacher while

pointing out the important parts to the students or these anomalies can be identified by the Diagnophone. That way better learning experience for the students will be achieved and this will make medical education more efficient and understandable for the students.

Before the development of this system, the users were interviewed before, during and after the implementation, by testing various prototypes of the system while various observations were made relating the experience of the users and the success of the experiments with the questionnaires was measured. As a result, the design was evolved and finalized.

In order to create a design that satisfies all the needs of the physicians in terms of user experience and human-computer interaction principles, 10 doctors and 5 medical students from several hospitals have been contacted and interviewed.

In this study, two datasets have been used. The first one is the dataset published by the International Conference of Biomedical and Health Informatics, and the second dataset is the one that we have presented which contains recordings from 44 patients (18 crackles, 5 wheezing, 11 crackles and wheezing, 10 healthy), taken from 6 different lung lobes, consisting of 370 audio recordings collected via the Diagnophone and labeled by the specialist doctors in Ümraniye Eğitim ve Araştırma Hastanesi.

Both datasets are first pre-processed before the classification step. In order to obtain more training data, data augmentation has been performed with two different techniques. Firstly, audio has been stretched (by speeding up and down randomly) along the time axis and secondly, after the creation of the Mel spectogram images, using Vocal Tract Length Perturbation (VTLP) with random linear warping, the Mel spectogram images have been transformed along the frequency axis randomly.

Following that, each audio recording has been divided into breathing cycles and it has been concluded that one breathing cycle takes approximately 5 seconds in 95% of the audio. Therefore, all of the audio has been divided into 5-second chunks and if the divided chunk is less then 5 seconds, the remaining of the audio has been filled with zero padding.

Especially with our data, when the patient firstly touched his body, the noise increase occurred due to exceeding the limit threshold values. This caused clipping and these parts have been identified and reduced from the audio.

After the pre-processed audios have been labeled using one-hot labels, the feature extraction step has been carried out. In this step, various temporal and spectral features of the audios have been calculated along with the Mel Frequency Cepstral Coefficients of the signals. Following this, the most efficient of these attributes have been selected using Principle Component Analysis and used in the classification phase.

Before the classification step, Mel spectrogram images of the signals have been calculated and these images have been classified alongside the features extracted and the success of both approaches have been compared.

In the classification step, first the data has been divided into 80% training, 10%test and 10% validation parts and for two datasets, models have been created using Convolutional Neural Networks, Support Vector Machines, K-Nearest Neighbour and Multi-class Adaboot algorithm with decision trees. The results of these algorithms have been calculated for both datasets. As a result, 81.1% accuracy was obtained with the CNN algorithm and the efficiency of the system was shown.

DIAGNOPHONE: SOLUNUM SESİ ANALİZİ İÇİN BİR ELEKTRONİK STETESKOP TASARIMI

ÖZET

Günümüzde solunum hastalıkları, dünyadaki ölümlerin ana nedenlerinden biridir. Türk İstatistik Kurumu verilerine göre, Türkiye'de bu alanda çalışan uzman doktor sayısında ciddi bir eksiklik bulunmaktadır.

Bu hastalıkların tanısı için röntgen, tomografi, MR gibi farklı testler mevcut olsa da, bu ekipmanların pahalı ekipmanlar olması sebebiyle her klinikte bulunması zordur. Az bulunması sebebiyle, bu ekipmanlardan yararlanmak için hastalar uzun süre sıra beklemek zorunda olması sebebiyle sonuçlara hemen ulaşamamaktadır.

Ayrıca örneğin MR sırasında hastanın hareketsiz şekilde uzun süre beklemesi gerekmektedir. Bu, özellikle klostrofobi sahibi hastalar için daha da stresli bir süreç haline gelmektedir. Bunun yanında, tüm bu ekipmanlardan yararlanan hastalar, yoğun miktarda radyasyona maruz kalmaktadır. Tomografi örneğinde, bu oran daha da fazladır.

Yapılan doktor görüşmeleri sonucunda, birçok hastalığın sadece stetoskop ile dinlenerek de anlaşılabileceği, fakat doktorların tanıdan emin olmak için tekrar bu tip ekipmanlarla bir test istediği, bu testleri bir validasyon niteliğinde kullandığı birçok doktor tarafından belirtilmiştir.

Anomali tespitinde bu tip ekipmanların mevcut olmasına rağmen, stetoskop doktorlar için hala ilk başvurulan, en ucuz ve en sık kullanılan tanı cihazıdır. Bu sebeple bu tezde, Makine Öğrenmesi yardımı ile hastalığın teşhisi konusunda hekimlere yardımcı olacak akıllı bir elektronik stetoskop tasarlanmıştır.

Tasarlanan bu stetoskop ile duyulan akciğer sesi aynı zamanda kayıt da edilebilecektir. Bu özellik, danışılacak uzman bir hekim bulunmadığı durumlarda telekonferans yapmak için kullanılabileceği gibi, hastanın ses verisinin saklanmasında da kullanılabilecektir. Günümüzde hastalar daha önce yaptırdıkları test sonuçlarına (örneğin kan testi, tomografi sonucu vb.) erişebilmektedirler. Fakat bir önceki akciğer sesi gibi bir bilgi saklanmadığı için, hastalığın takibinde ses verisi kullanılamamaktadır. Oysa astım gibi takip gerektiren hastalıklarda, hastanın akciğer sesinin saklanması, hastalığın takibini kolaylaştıracağı gibi, hastalığın önceye kıyasla durumunun belirlenmesi konusunda da yardımcı rol oynayacaktır.

Diagnophone ile kaydedilen bu ses aynı zamanda tıp eğitiminde de kullanılabilecektir. Çünkü günümüzde tıp eğitimi, hasta yatağı başında toplanan öğrencilerin, öğretmenlerinin ardından sıra ile aynı stetoskop ile hastanın akciğer seslerinin dinlenmesinden ibarettir. Yapılan kullanıcı görüşmeleri sonucu, bunun efektif bir öğrenim çeşidi olmadığı çıkarımı yapılmıştır. Fakat Diagnophone sayesinde hastadan kaydedilen sesler, cep telefonu hoparlöründen tekrar dinletilerek öğretmen tarafından anomalinin

olduğu yerin özellikle belirtilmesi, ya da Diagnophone tarafından sonucun öğrencilere gösterilmesi, tıp eğitimini öğrenciler için daha verimli ve anlaşılır hale getirebilecektir.

Tüm bunları gerçeklemeden önce, gerçekleme sırasında ve sonrasında kullanıcılar ile görüşülmüş, çeşitli prototiplerin kullanıcılar tarafından deneyimlenmesi sağlanmış, bu sırada çeşitli gözlemler yapılmış ve anketler ile yapılan deneylerin başarısı ölçülerek, tasarımların bu doğrultuda evrimleşmesi ve son halini alması sağlanmıştır.

Hekimlerin kullanıcı deneyimi ve bilgisayar insan etkileşimi açısından tüm ihtiyaçlarını karşılayacak bir tasarım oluşturabilmesi için, çeşitli hastanelerden 10 doktor ve 5 tıp fakültesi öğrencisi ile görüşmeler ve kullanıcı testleri yapılmıştır.

Bu çalışmada iki farklı veri kümesi kullanılmıştır. Bunlardan ilki, Uluslararası Biyomedikal ve Sağlık Bilişimi Konferansı tarafından yayınlanan, 920 adet ses dosyasından oluşan, 126 hastadan toplanmış, 6898 solunum sesi içeren veri seti, ikinci ise 44 hastadan (18 çatırtı, 5 hırıltı, 11 çatırtı ve hışıltılı, 10 sağlıklı) 6 farklı akciğer lobundan toplanan kayıtları içeren veri kümesidir. İkincisi ise, Ümraniye Eğitim ve Araştırma Hastanesi'nde Diagnophone aracılığıyla toplanan ve uzman doktorlar tarafından etiketlenen 370 ses kaydından oluşan veri kümesidir.

Her iki veri kümesi de sınıflandırma adımından önce ön işleme aşamasından geçirilmiştir. Daha fazla eğitim verisi elde etmek için, iki farklı teknikle veri çoklama işlemi gerçekleştirilmiştir. İlk olarak, sinyallere zaman ekseni boyunca rastgele bir oranda uzatma veya daraltma uygulanarak ses dosyaları kopyalanmıştır. İkinci olarak ise üretilen Mel spektogram görüntüleri, frekans ekseninde rastgele doğrusal çözgü (random linear warping) Vokal Kanal Uzunluğu Pertürbasyonu (VTLP) kullanılarak dönüştürülmüştür.

Bunu takiben, her bir ses kaydı bir solunum döngüsünden oluşacak şekilde bölütlenmiştir. Elde edilen yeni seslerin yaklaşık %95'inde bir nefes alış ve veriş döngüsünün 5 saniyede tamamlandığı sonucuna varılmıştır. Bu sebeple, tüm sinyaller, 5'er saniyelik parçalara bölünmüştür ve elde edilen bölütler 5 saniyeden azsa, bölütüm kalanı sıfır dolguyla (zero padding) doldurulmuştur.

Özellikle ikinci veri kümesinden elde edilen sinyallerde, diyafram hasta vücuduna ilk kez dokunduğunda, limit eşik değerlerini aşması nedeniyle gürültü artışı ve seste kırpılma (clipping) meydana geldiği gözlenmiştir. Bu bölümlerin belirlenerek, belirli limitler içinde kalacak hale getirilmesi sağlanmıştır.

Ön işleme tabi tutulmuş sesler one-hot etiketleme kullanılarak etiketlendikten sonra, özellik çıkarma adımı gerçekleştirilmiştir. Bu adımda, seslerin çeşitli zamansal ve spektral özellikleri, sinyallerin Mel Frekanslı Cepstral Katsayıları ile birlikte hesaplanmıştır. Bunu takiben, bu özelliklerden en etkin olanları Prensip Bileşen Analizi kullanılarak seçilmiş ve sınıflandırma aşamasında kullanılmıştır.

Sınıflandırma adımında ilk önce veriler %80 eğitim, %10 test ve %10 validasyon için kullanılmak üzere bölümlere ayrılmıştır. Daha sonra, sınıflama aşamasında iki çeşit yaklaşım izlenmiştir. Birinci olarak elde edilen bu öznitelikler, Destek Vektör Makineleri, K-En Yakın Komşu ve çok sınıflı Adaboot Karar Ağacı algoritması yardımı ile sınıflandırılarak başarıları kıyaslanmıştır. İkinci olarak ise, sinyallerin Mel spektrogram görüntüleri oluşturulmuş ve bu görüntüler Konvolüsyonel Sinir Ağları'na girdi olarak verilerek sınıflanması sağlanmıştır. Sonuç olarak, CNN algoritması %81.1 ile iki veri kümesi için de en yüksek doğruluk elde edilen algoritma olmuştur.

Literatürde akciğer oskültasyonunun vücudun hem ön hem de arkasından yapılmaktadır. Bu sebeple hastaların hem sırt hem de göğüs bölgesinden toplanan sesler işlenmiştir. Fakat göğüs bölgesinden toplanan ses dosyalarında kalp seslerinin akciğer seslerini baskıladığı ve bunun toplam başarıyı baskıladığı gözlemlenmiştir. Daha sonra doktorlar ile yapılan gözlemler ve görüşmeler sonucunda, teorik bilgide göğüsten oskültasyon önerilse de, pratikte doktorların akciğer oskültasyonunu sadece hastaların sırt bölgesinden yaptığı, göğüs bölgesini es geçtiği gözlemlenmiştir. Bu iki sebeple, projenin kapsamında sadece sırt bölgesinden alınan sesler benimsenmiştir.

1. INTRODUCTION

Today, respiratory diseases are one of the major causes of mortality in the world with 43% crude death rates according to World Health Organization (WHO) [3]. The report of the Turkish Statistical Institute states that in 2016, respiratory diseases constitute 11.7% of deaths with 49,295 patients, which have risen up to 12.0% with 49,855 deaths in 2017, in Turkey [4].

There are numerous diagnostic tests available that can be used to determine a variety of respiratory diseases, patient history, current symptoms of the patient. X-ray, Computed Tomography (CT scan), Magnetic Resonance Imaging (MRI) are extremely helpful during the diagnosis. However, these tests are expensive procedures hence hard to be found in rural areas. Since there is a limited number of these machines in hospitals, patients have to wait long time periods to benefit these opportunities. Furthermore, the CT scan causes radiation for the patients. Concerns about CT scans include the risks from exposure to ionizing radiation and possible reactions to the intravenous contrast agent, or a dye, which may be used to improve visualization. The exposure to ionizing radiation may cause a small increase in a person's lifetime risk of developing cancer [5]. Moreover, MRI is uncomfortable for the patients since the patients must remain still for extended periods of time during the scan.

Even though there are different tests available, the initial diagnosis, the physician's first step to identify the anomaly of respiratory diseases is mostly auscultation with the help of a stethoscope.

Auscultation of the lung is an important part of the respiratory examination. Auscultation assesses airflow through the trachea-bronchial tree. It is important to distinguish normal respiratory sounds from abnormal ones such as crackles, wheezes and pleural rub in order to make the correct diagnosis. The stethoscope is necessary for the auscultation to be performed in order to understand the underlying pathophysiology of various lung sounds generation for a better understanding of the disease processes.

Since its invention in 1816 by French physician Rene Laennec, the stethoscope has been indispensable, iconic equipment for the medical professionals [6]. The stethoscope is an acoustic medical device for auscultation or listening to the internal sounds of an animal or human body. It typically has a small disc-shaped resonator (diaphragm) that is placed against the chest of the patient, and two tubes connected to the earpieces. It is often used to listen to lung and heart sounds. It is also used to listen to intestines and blood flow in arteries and veins. In combination with a sphygmomanometer, it is commonly used for measurements of blood pressure [7].

The stethoscope now serves as the most commonly used diagnostic method in hospitals in primary health care, health centers, family health centers, and rural areas that do not have expensive equipment such as tomography or X-ray. However, traditional stethoscopes have some drawbacks because hearing the patient in low amplitude and frequency signals using a stethoscope, could lead to a false diagnosis especially in noisy environments such as Emergency Room (ER). Also using a stethoscope requires expertise and skills, which are gained over time. The solution to those bottlenecks is highly important in early detection of abnormality of lung sound by medical professionals [8].

According to the report of the Turkish Ministry of Health, in 2014, there are 6.2 internal medicine doctors per 100 thousand people [9]. It can be concluded from this data that there is a serious lack of experts in the area against such deadly diseases in our country.

In order to design a system to help the doctors with these drawbacks, understand their problems about the field and create a better user interaction and better user experience, interview sessions and user tests with 15 doctors (2 pediatric emergency doctors, 2 adult emergency doctors, 1 ambulance doctor, 1 internal medicine specialist, 1 specialist pediatrician, 1 home healthcare doctor from Ümraniye Eğitim ve Araştırma Hospital, 2 family healthcare doctors from Zeytinburnu Aile Sağlığı Merkezi and 5 intern doctors from İstanbul Cerrahpaşa Medical School) have been carried out.

According to the conclusions made from interviews with the doctors, general practitioners who work in rural areas have difficulty finding a specialist doctor to consult.

To eliminate all these shortcomings, in this thesis, a stethoscope, the most frequently used tool by the medical professionals, has been designed to diagnose pulmonary anomalies.

The proposed system consists of an electronic stethoscope and a mobile app that allows doctors to record the data from the patient and identify the abnormality with the help of Signal Processing and Machine Learning. This will assist physicians in diagnosing respiratory diseases and identifying the pathophysiology in the lungs, especially in rural areas where there is no expensive equipment such as x-ray or tomography, and lack of specialist physicians to consult. The system will contribute to reducing the time and money allocated to the diagnosis of the disease at the same time helping the problem of lack of specialist physicians.

In addition to these, the system will also be helpful with the practical training of medical students. The current education in medicine is carried out in the form of listening to the breathing sounds of the patients who are being treated in the university hospital only, by the intern students. For this reason, students are only able to distinguish the cases where they have the chance to meet in their educational lives.

However, since it is possible to record the voice heard in the proposed system, this will allow the recording of the rare cases to be recorded. Therefore, these records can be used in medical education, as well as in consultation with another specialist doctor through teleconference.

In addition, today, doctors have access to the previous tests of the patients, such as tomography results, blood test results, etc, but they do not have the opportunity of accessing the previous sounds of lungs and previous auscultation recordings of the patients. Yet this information is important in diseases requiring follow-up such as asthma.

The proposed system will eliminate those shortcomings with its ability to help with the diagnosis and sound recording.

In Chapter 2, studies in the literature about electronic stethoscopes and Artificial Intelligence (AI) approach to respiratory audio has been discussed in detail.

In Chapter 3, the design procedure of the system has been presented. The study and conclusions of user experience and user studies conducted are also presented in this section.

In Chapter 4, the respiratory audio analysis algorithms and different approaches followed and the experiments have been explained in detail with their results.

In Chapter 5, the thesis is concluded by summarizing the whole thesis and presenting future plans.

2. LITERATURE REVIEW

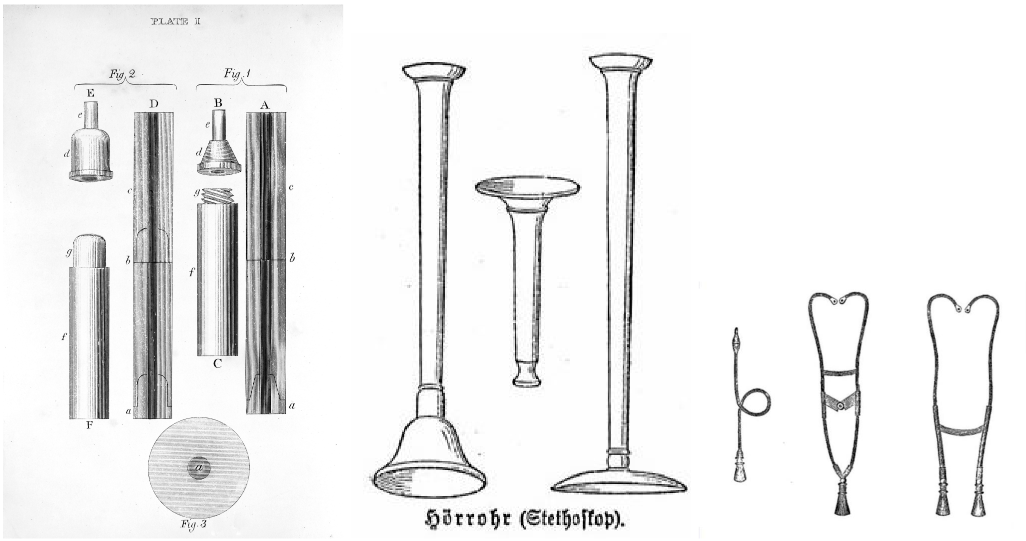
In this chapter, the evolution of typical stethoscopes since it's invention to present and different stethoscope kinds have been investigated. Studies in Computer-Aided Auscultation and works established in the literature relating the respiratory audio analysis has been explored and presented.

2.1 Stethoscopes

Since it has been invented in 1816 by French physician Rene Laennec, the stethoscope serves as the most commonly employed technique for diagnosis [7, 10]. The stethoscope is an acoustic medical device for auscultation which is performed by listening to the internal sounds of an animal or human body. It typically has a small disc-shaped resonator that is placed against the chest of the patient, and two tubes connected to earpieces. It is often used to listen to lung and heart sounds. Besides, it can also be used to listen to intestines and blood flow in arteries and veins. In combination with a sphygmomanometer, it is commonly used while measuring blood pressure. Even though many incremental improvements have been made to Laennec's original invention, including flexible tubes and binaural constructions, the essential function, and usage of the non-electronic stethoscope remained relatively unchanged for almost 150 years [11].

When first invented, the stethoscope was consisted of a wooden tube and was monaural as shown in Fig.2.1(a) . Laennec invented the stethoscope because he was uncomfortable placing his ear on women's chests to hear heart sounds. Laennec called his device the "stethoscope" (stetho- + -scope, "chest scope") [7] which then evolved as in Fig. 2.1(b) and Fig. 2.1(c).

The invention of the stethoscope made a major improvement of understanding the disease from only being a bundle of symptoms to the current sense of a disease as a problem with an anatomical system even if there are no observable symptoms.



(a) Laennec's Stethoscope (b) Hörrohr Stethoscope (c) Early flexible Stethoscopes

Figure 2.1 : Early stethoscopes invented.

In the 1960s, Dr. David Littmann designed a two-sided chest-piece, with a bell on one side and a diaphragm on the other, greatly improved stethoscope acoustics. This tunable diaphragm had a very hard (G-10) glass-epoxy resin diaphragm member with an over-molded silicone flexible acoustic surround which permitted increased excursion of the diaphragm member in a "z"-axis with respect to the plane of the sound collecting area. The left shift to a lower resonant frequency increases the volume of some low-frequency sounds due to the longer waves propagated by the increased excursion of the hard diaphragm member suspended in the concentric acoustic surround [11, 12].

Restricting excursion of the diaphragm by pressing the stethoscope diaphragm surface firmly against the anatomical area overlying the physiological sounds of interest, the acoustic surround could also be used to dampen excursion of the diaphragm in response to z-axis pressure against a concentric fret. This raises the frequency bias by shortening the wavelength to auscultate a higher range of physiological sounds [12].

Stethoscopes are often considered as a symbol of healthcare professionals, as various healthcare providers are often seen or depicted with stethoscopes hanging around their necks. Murphy claimed that the stethoscope, when compared to other medical equipment, had the highest positive impact on the perceived trustworthiness of the practitioner seen with it [12].

There are different kinds of stethoscope types invented which are stated below.

2.1.1 Acoustic stethoscopes

Acoustic stethoscopes are the most commonly used stethoscope types nowadays. They operate on the transmission of sound from the chest piece, via air-filled hollow tubes, to the listener's ears. The chest piece usually consists of two sides that can be placed against the patient for sensing sound; a diaphragm (plastic disc) or bell (hollow cup). When placed on the patient's body, the body sounds vibrate the diaphragm, creating acoustic pressure waves that travel up the tubing to the listener's ears. The bell transmits low-frequency sounds, while the diaphragm transmits higher frequency sounds. This two-sided stethoscope was invented by Rappaport and Sprague in the early part of the 20th century [13].

Figure 2.2 : A typical stethoscope [1].

2.1.2 Electronic stethoscopes

A digital stethoscope is able to convert an acoustic sound to electronic signals, which can be amplified for optimal listening. These electronic signals can be further processed and digitized to transmit to a personal computer or a laptop. It filters, buffers and amplifies the auscultated sounds after converting the acoustic sound to a digital signal. Therefore a clinical diagnostic decision can be concluded more easily by the physician [14].

Figure 2.3 : An electronic stethoscope [2].

The electronic stethoscope in general, consists of a sensor, a digital filter, and a user interface. The sensor is configured to detect acoustic signals from the human body and generate medical measurement signals based on the detected acoustic signals. The digital filter is capable of filtering the medical measurement signals with one or more headset filters, each headset filter operable to provide a transfer function to compensate for a characteristic frequency response of a particular headset. The user interface is configured to accept input from a user, wherein the user interface is further configured to allow a user to select a headset filter from the one or more headset filters and the digital filter is configured to filter the medical measurement signals with the selected headset filter [10].

There are various methods applied during the design of an electronic stethoscope. The most popular electronic stethoscopes are reviewed below.

Even though the majority of the electronic stethoscopes use a microphone for sound reproduction, The Welch Allyn Meditron Electronic Stethoscope utilizes a highly directional, recessed pressure sensor which eliminates much of the ambient noise in the environment while reproducing body sounds. Because they claim that no matter the fidelity, microphones pick up ambient noise in the environment transmitted through the stethoscope diaphragm and chest-piece [15].

3M Littmann launched two different electronic stethoscope models that use a piezo-electric crystal placed within foam behind a thick rubber-like diaphragm. The models both transmit sounds via Bluetooth technology, has a small LCD window that would display the patient's heart rate in real time, eliminating 85% of ambient noise by using both sources of ambient noise to cancel each other out, leaving only the unfiltered sound of the heart [16].

The Thinklabs One uses an electromagnetic diaphragm with a conductive inner surface to form a capacitive sensor. This diaphragm responds to sound waves, with changes in an electric field replacing changes in air pressure. It provides multiple filter choices ranging from very low frequency (low pitched sounds such as S3 for heart sounds), to higher frequencies (lung sounds) and a few in between. In addition to that, the diaphragm pressure can also be changed by applying more or less pressure on the diaphragm. As the user pushes on the diaphragm, the gap in the sensor changes, which modifies the sensor's response to lower frequencies [17].

The Eko has 2 different products namely CORE and DUO. CORE is a complete electronic stethoscope with digital and analog capabilities that amplifies heart and lung sounds and reduces ambient noise. The analog-digital toggle allows physicians to use their stethoscope even when it's turned off [18]. DUO is a portable stethoscope combined with an Electrocardiogram (ECG) designed for clinical use. By combining heart sounds and ECG, DUO enables a more comprehensive view of cardiovascular function. DUO makes it possible to capture ECG and systolic time intervals for effective heart failure and Atrial Fibrillation (AFib) detection. With their supported app, it can also be connected to a smartphone [19].

2.1.3 Problems of current stethoscopes

The stethoscope serves as the most frequently employed technique in primary health care. However, traditional stethoscopes have some drawbacks because hearing the patient in low amplitude and frequency signals using a stethoscope, could lead to a false diagnosis especially in noisy environments such as emergency rooms. Also using a stethoscope requires lots of expertise and skills, which is gained over time. Therefore the solution to those bottlenecks is highly important in early detection of abnormality of lung sound by medical professionals.

Especially for the respiratory diseases, the diagnosis of some of the diseases may be hard to identify for non-experienced or peripheral doctors working in rural areas with no experts to consult.

Since there is no recording feature in the analog stethoscope, consulting to colleagues in long distances or re-analyzing the sounds offline, or during the preparation and publishing phases of academic articles is simply not possible.

Even though there are various electronic stethoscopes in the market, compared to the analog stethoscopes, they are approximately 10 times more expensive than the regular stethoscopes. Furthermore, they mainly focus on voice amplification and do not have the ability to identify the abnormality existing in the sounds heard from the patient's body.

2.1.4 Expected users and use context of the system

Expected users of the system are General Practitioners (GP), ER doctors, medical students and doctors who work in rural areas with no specialists to consult. They are expected to be using this technology during their lung examinations.

2.2 User Experience Studies on Healthcare

User experience (UX) is defined as a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service. User experience incorporates every one of the users' feelings, convictions, inclinations, recognition, physical and mental reactions, emotional aspects, practices, and achievements that happen previously, during and after use [20]. It is the experiences that influence a person after the interaction with a product in specific conditions [21,22]. The objective of user experience is to make a product that delivers a productive and efficient end-user experience by empowering users to accomplish their goals in the most ideal manner conceivable [23].

UX tests are one of the most important subjects in human-centered product development. The tests that focus on the user experience of a product can uncover plan or execution blames as well as spark some new ideas and future thoughts that could take the product to a new level. In this manner, conducting a user experience test gives insights about usability practice, since it gives direct input on how real users use the system.

Usability can incorporate the sort of perspective of the users' personal goals, which can include the kind of perceptual and emotional aspects typically associated with user experience. Usability criteria can be used to assess user experience. Any feedback from the users during or after the development process of a product passes on significance since it determines user acceptance, user experience and the quality of the product [24].

A user's experience and interaction with a product are affected by the user's personality and culture, social factors and the conditions in which the user uses the product. The user's prior experiences, expectations, emotions and the product's influential factors such as mobility and adoptive influence the experience that revealed from user-product interaction [22].

Design, particularly in healthcare, is about efficiency, usability, robustness and better user experience for patients as well as medical practitioners. User Experience methodologies and conducting user tests are a powerful approach to solve people's problems. We have applied these methods to healthcare with the help of technology and AI.

User experience in healthcare is an increasingly important topic due to the advances in technology. Factors that affect usability in healthcare include the profession being intense and stressful, requiring lots of late and tiresome working hours, being in a noisy, chaotic and crowded environment, the need to deciding and acting quick. Improving the user experience for clinicians is of utmost importance due to its impact on patient care as well as physician and clinician satisfaction considering the hard conditions they operate in.

Modern technology in the healthcare system brings patients and physicians better and more comfortable service. However, for this to work, the technology needs to be user-friendly. Healthcare workers need to feel as if they had always used these tools.

In order to capture user experience, there are several methods, such as, interviews, observation, surveys or questionnaires, diaries, storytelling, and prototyping [22]. In [25], authors stated that for the long-term use, surveys, diaries, and storytelling have been considered as effective tools to get written information about users' experiences. Moreover, observation and questionnaires are beneficial methods to get an insight about users' experiences from non-verbal expressions, which is significant since the users may not be aware of their experiences or be capable to express them verbally [22].

2.3 Studies on Respiratory Audio Processing

Computer-Aided Auscultation or Computerized Assisted Auscultation (CAA) is a digitized form of auscultation which includes the recording, visualization, storage, analysis and sharing of digital recordings of heart or lung sounds using an electronic stethoscope.

CAA is designed to assist health care professionals who perform auscultation as part of their diagnostic process. Automatic detection and classification of computer-aided pathological sounds are intended to help physicians diagnose or monitor diseases such as asthma, Chronic Obstructive Pulmonary Disease (COPD) and pneumonia. There are various approaches to this issue. In the following section, researches about respiratory audio analysis research are discussed.

Murphy et al. [12] presented a framework where they utilized a modernized multi-channel lung sound analyzer to decide if unbiasedly estimated lung sounds varied fundamentally in patients with pneumonia versus asymptomatic subjects. They also wanted to quantify the pneumonia sounds of patients for educational purposes. They studied 100 patients who were diagnosed by their physicians as having pneumonia and 100 patients without pneumonia with forty-eight percent female patients. Subjects were inspected with a multi-channel lung sound analyzer in which

14 microphones are placed into a soft foam pad. The microphone pad was secured with a uniquely designed, single-use, dispensable interface that forestalls the transmission of pathogens to the pad. The computer quickly performed a time-expanded waveform analysis of each channel, and the analysis was used to verify the automated analysis.

Morten et al. [26] propose an AI-based methodology for identifying crackles in lung sounds recorded utilizing an electronic stethoscope in a large health survey by analyzing 209 subjects which contain wheezing, crackles and normal lung sounds. They compared different feature extraction techniques with selecting variance, range, the sum of the simple moving average (coarse), the sum of the simple moving average (fine) and compared them with features obtained with the MFCC algorithm and defended that these 5 features outperformed MFCCs.

Ono et al. [27] researched whether spectral analysis with Fast Fourier Transform (FFT) of respiratory audio is useful in the diagnosis and evaluation of the severity of Interstitial Pneumonia (IP). They have studied 10 healthy and 21 IP patients. They have generated respiratory averaged linear intensities using FFT and the determined frequency at maximum sound intensity (Fmax), and quartile frequencies (f25, f50, and f75), compared these values between the groups, generated receiver operating characteristic curves to compare the detectability of IP between the indices and auscultation in all cases. As a result, they have discovered Both f50 and f75 were significantly higher in the IP group.

Chamberlain et al. [28] used time-frequency analysis and the Short Time Fourier Transform to identify sections of wheezing in recorded respiratory audio files by training and testing their model with data acquired from 38 patients at a pulmonary clinic. They identified a wheezing sound with 86% success.

Sinharay et al. [29] have introduced a system in which the direct acoustic coupling technique is employed to convert the smartphones into a digital stethoscope by using an attachment. They have used a smartphone, headset and a converter box (cavity) on a regular stethoscope and created an electronic stethoscope. The motivation of their work is to enable the heart patients to send their heart sound to their doctors from their homes without the need to go to a hospital. This is particularly useful in poor or developing countries where there is a scarcity of healthcare centers and patients have to travel long distances to visit the hospitals. Their work states that it would especially be helpful with patients who went through heart surgery and requires follow up visits or for the elderly population requiring routine checkups.

Kahya et al. [30] used different feature sets in conjunction with K-Nearest Neighbors (K-NN) and Artificial Neural Network (ANN) classifiers for the classification of respiratory audio signals and compared their performances with different feature sets derived from respiratory sound data acquired from one microphone placed on the posterior chest area. Each subject is represented by a single respiration cycle divided into sixty segments from which three different feature sets consisting of 6th order AR model coefficients, wavelet coefficients and crackle parameters in addition to AR model coefficients are extracted and the two classes (healthy and pathological) were identified.

Aguilera et al. [31] proposed a system of 3-D printed stethoscope connected to a smartphone. In their system, the chest piece consists of an electret microphone embedded into the drum of a 3D printed chest piece. An electronic dongle amplifies the signal acquired from the microphone and reduces external noises. The acquired audio is then shown in a graphical user interface in an Android display. They mostly focused on how to digitally get the sound waves, properly amplify them and show the sound waves in the mobile phone and did not include a diagnosing feature to their system.

What distinguishes our study from the ones in the literature is that, competitive technologies that built their own electronic stethoscope mainly focus on noise reduction and audio amplification in the electronic stethoscopes which did not include diagnostic features in their systems. The ones that included automatic classification of the sounds, either focused on heart sounds, only wheezing or only crackles detection, or used the stethoscope only for the data acquisition phase.

We, on the other hand, wanted to create an end to end product with our study. We both used out stethoscope during the data acquisition phase, and after creating a machine learning model with the collected data, we provided a product for the physicians to use as a new tool during their auscultation exams and get real-time results.

Furthermore, none of the related work focused on the concepts of HCI and UX. We, however, conducted detailed user experience tests and surveys with physicians before during and after using the product designed and evolved our design and planning according to the outcomes, which are explained in detail in Chapter 3.

3. DESIGN PROCEDURE OF DIAGNOPHONE

In order to create a better user experience for the users, the actual users of the product -the doctors- have been contacted, observed in their natural work setup and interviewed at Ümraniye Eğitim ve Araştırma Hospital, Zeytinburnu Aile Sağlığı Merkezi and İstanbul Cerrahpaşa Medical School. At the end of this process, the needs of the users and the struggles along the process have been identified and the design has been evolved accordingly, using Human Computer Interaction principles. In the following sections, first, the user experience tests and interviews with the doctors and medical students have been explained. Then in section 3.1, the results of the user tests and interviews, and design changes related to these results have been listed. In section 3.3, the stethoscope built in order to collect data and create a tool for the medical professionals has been explained. Then the mobile app built has been demonstrated.

In the following section, the process of developing the design, cognitive walk-through of the design with the users, processing the results from these sessions and changes that have been made based on the results has been explained.

3.1 Design Observations

The aim of this thesis is to identify a problem that healthcare professionals were experiencing and come up with a solution to make their lives easier with the help of technology. In order to do that, 15 doctors (2 pediatric emergency doctors, 2 adult emergency doctors, 1 ambulance doctor, 1 internal medicine specialist, 1 specialist pediatrician, 1 home healthcare doctor from Ümraniye Eğitim ve Araştırma Hospital, 2 family healthcare doctors from Zeytinburnu Aile Sağlığı Merkezi and 5 intern doctors from İstanbul Cerrahpaşa Medical School) have been contacted and observed in their natural environments during patient examination and auscultation which can be seen in Fig. 3.1(a) and Fig. 3.1(b).

During the observations and interviews, it has been concluded that the first and most frequently used tool by the medical professionals is the stethoscope. After noticing that, physicians' daily tasks which involve using the stethoscope have been the focal point of our observations. We have been given the chance to examine and use their stethoscopes and taught how to use it and shown which areas on the human body that the stethoscope should be placed by the physicians which are shown in Fig. 3.2.

During the interview, questions about how, when, why, how often they use the stethoscope, how do they understand there's an anomaly in the patient, do they find the stethoscope easy or ergonomic to use, is it sufficient only to have the stethoscope to diagnose a disease, what other tools do they use in order to help with the diagnosis alongside with the stethoscope has been asked. They are also asked to fill a Pre-Test Demographic Information Form in Appendix A.1 and the User Survey Questions in Appendix A.2 by using Product Reaction Cards in Appendix A.3.

After the design phase is complete, the same doctors have been contacted again for the usability tests with the low fidelity prototype and the second interviews have been carried out which can be seen in the Figure 3.4 . They were shown the sketches of the prototypes and were asked to imagine using that device instead of the stethoscope they currently possess. With "the speaking out loud" method, they described their feelings, features they liked, they could not adjust, found unnecessary, or technically impossible to make about the prototype and all of the feedback we discussed those topics which helped to improve the prototype. The users are also requested to fill the Post-Test User Satisfaction Form in Appendix A.4 after the prototype testing.

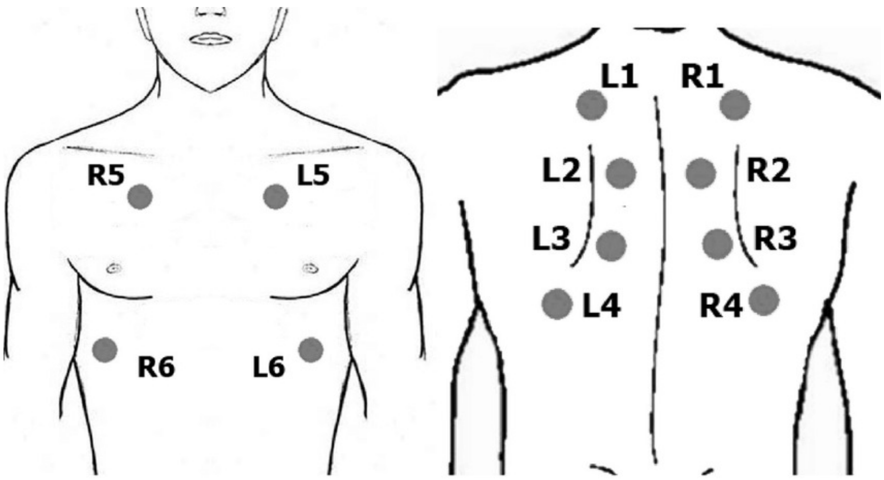
The system has been designed according to the Shneiderman's eight golden rules of design [32] using the data acquired from the user interviews. Since the system uses components that the users are familiar, it helps to reduce short-term memory load and retention over time. It has an interface similar to what the doctors are used to, just with extra features.

3.2 The Design Criteria

After the interviews and observations with the doctors and medical students about their use of the stethoscope and testing the low fidelity prototype of Diagnophone, the following requirements have been concluded.

(a) Adult examination (b) Pediatric examination

Figure 3.1 : Doctors observed in ER during auscultation.



(a) (b)

Figure 3.2 : Areas that a stethoscope should be placed for lung auscultation.

- Even though the stethoscope is the most frequently used tool, the physicians do not find it ergonomic. Especially the earpiece of the regular stethoscope has to be replaced with **ergonomic, soundproof headphones**.

- The system should be able to **record sounds** for keeping track of patient history or in order to be used in teleconferencing for education and consulting. [1]

- They have complained about the lack of necessary equipment such as X-ray, MRI or CT-scan in the rural areas and the lack of specialist doctors. Newly appointed doctors may find it hard to identify differences in respiratory audios in some cases. So the system should be able to **assist doctors with the abnormal sound classification**.

- Since especially the ER environment is so noisy, the system should be able to **isolate the noise and increase volume** in order to help doctors hear well.

- First, our aim was to detect anomalies from every part of the body that was heard by a stethoscope which is heart, lungs, intestines, and thyroid. But after consulting these ideas with expert doctors, these requirements have been removed and decided only focusing on heart and lung sounds. Because it has been told to us that the thyroid is a rare case and stethoscope is not sufficient by its own. In addition, the anomaly detection in the intestines is straightforward. The abdominal area gets partitioned into 4 imaginary parts. After listening to each part, the number of activity sounds is counted. If it is more than 4, which means the patient is most likely to have diarrhea, if it is less, the patient may have constipation, etc. Since this result can be achieved really easily just by counting, it has been decided to focus on only heart and lung sounds. However, after the interviews and user tests, it has been inferred that most of the doctors have a prejudice about not being able to trust a device other than an ECG device with the heart anomalies. And they have expressed that the respiratory sounds are harder to identify and they would appreciate the help with the lung sounds more. So, the scope of the project was identified accordingly and the only focus of the Diagnophone was determined as **lung sound anomaly detection**.

[1] Teleconferencing or Telemedicine is the interactive audiovisual communication between health care providers and their patients or among health care providers regardless of geographic distance [33]

- It has been observed that every doctor we have interviewed was using Littmann brand stethoscopes. When they were asked would they consider using a different brand, most of them prefer to change the setup. In their explanation, they have gotten used to the specific sound of a Littmann stethoscope, they were afraid that hearing with another stethoscope, may sound different and they would miss the anomaly sound. This led to the decision to design our product using the same **Littmann brand stethoscope**.
- It has been observed that **young doctors and interns** were more excited about the Diagnophone compared to the older physicians who rather had faith on their ears with years of experience rather than a device.
- We have concluded that the Diagnophone will be useful in education, especially for the **medical students in the training period**. The intern doctors told that they gathered around the bed of the patient and underwent auscultation in turns under the supervision of their instructors, and mostly did not understand anything they hear. Instead, using Diagnophone, the audio from the patient can be recorded and played identifying the abnormal behavior by the instructor at the same time.
- After the interviews, it has been deducted that some diseases require follow-up and control. In order to observe the status of patients who require follow-up in hospitals, doctors usually examine the previous and subsequent chest X-rays, but lung sound is also important in providing information about the condition of the disease in some cases, and there is no such **record kept of the patients to follow-up**.

The validation of the requirements was based on describing the physicians the system, explaining the requirements, what the system is capable of doing and what they should expect from the product at the same time giving them **the sketches, low-fidelity prototypes** and asking them to imagine using the device according to their needs while observing if the system and the requirements meet their needs and do what it is expected to do.

3.3 The Hardware Design

In Fig. 3.3, the overview of the system architecture is shown. The diaphragm of the stethoscope has been cut from the rubber tube and an electret microphone is attached into the rubber tube. With the help of a divider attachment for microphone and headphone, together the microphone and headphone are connected to a smartphone via 3.5 mm jack. As a result, the audio heard from the diaphragm of the stethoscope was able to be recorded by the smartphone in order to be classified. Littmann Classic II stethoscope is used with a Trust Lava microphone. This can be seen in Figure 3.4 where the user is using the Diagnophone in real time in Pediatric Intensive Care Unit of Ümraniye Eğitim Araştırma Hastanesi.

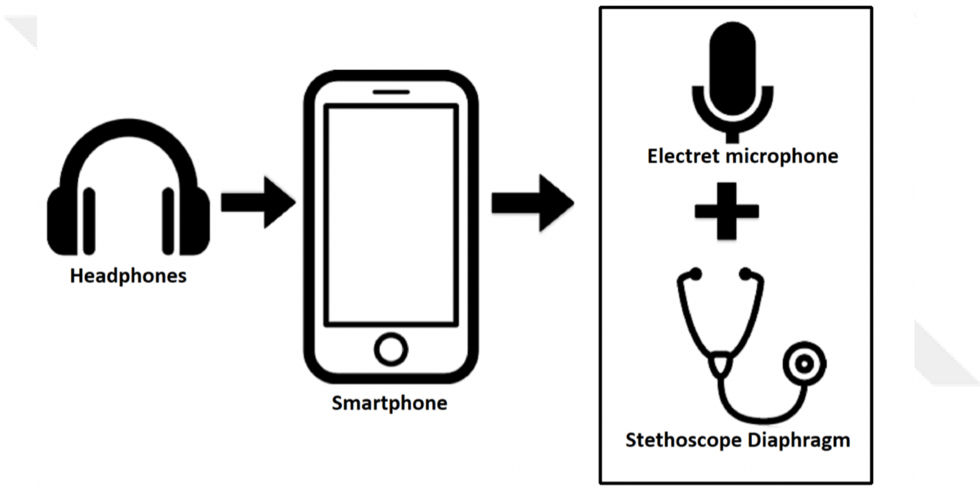


Figure 3.3 : The architecture of the system.

Fig. 3.5 shows the microphone and a diaphragm of the Littmann stethoscope before being assembled. The finished prototype can be seen in Fig. 3.6 with connected to a smartphone.In order to remove outside noise as much as possible, the microphone is wrapped with polyurethane heat shrink tubes. The physicians are asked to use this stethoscope on their patients by connecting their smartphones with a headphone of choice.

3.4 Mobile Application Design

A mobile application has been designed and developed in order to record the lung sounds of the patients and show the results in the user interface to the doctors. In the first interface in Fig.3.7(a), the physician is asked to fill some data about the patient.

Figure 3.4 : User with the diagnophone.

Then the physician is navigated to the audio recording. In this UI in Fig.3.7(b), the physician selects which lobe of the lung that the audio is going to be acquired and when pressed, the recording starts. If the physician thinks the recording time is sufficient, "stop recording" button is pressed and the recording is saved with the lung lobe name and the current date of the recording is prefixed to the name of the audio file. The physician also may choose to pause and continue the recording which is also another feature that is available in the user interface 2. In the third UI in Fig.3.7(c) the records of the patients can be accessed and re-listened. These records are named according to the lung lobe that the audio is being collected, patient name and the current date the audio is being recording. These audio data can then be shared with a colleague for telemedicine purposes and getting suggestions related to the diagnosis, or can be used while doing an academic research by the doctors to collect data of rare cases, or can be used during the medical education to teach the students different cases of the respiratory audio. The data has been recorded from 6 different locations of patients' lungs, in one channel, using 16 bits data with 4100 Hz sampling rate. The doctors have used our mobile application to record the respiratory audio for our dataset. This sound data is then collected from the doctors, labeled by the doctors with 4 classes namely healthy, crackles, wheezing, crackles and weezing (both) in order to be used to train the machine learning models. In the following images, the user interface of the mobile application is shown.

Figure 3.5 : Pieces of the electronic stethoscope before the assemble.

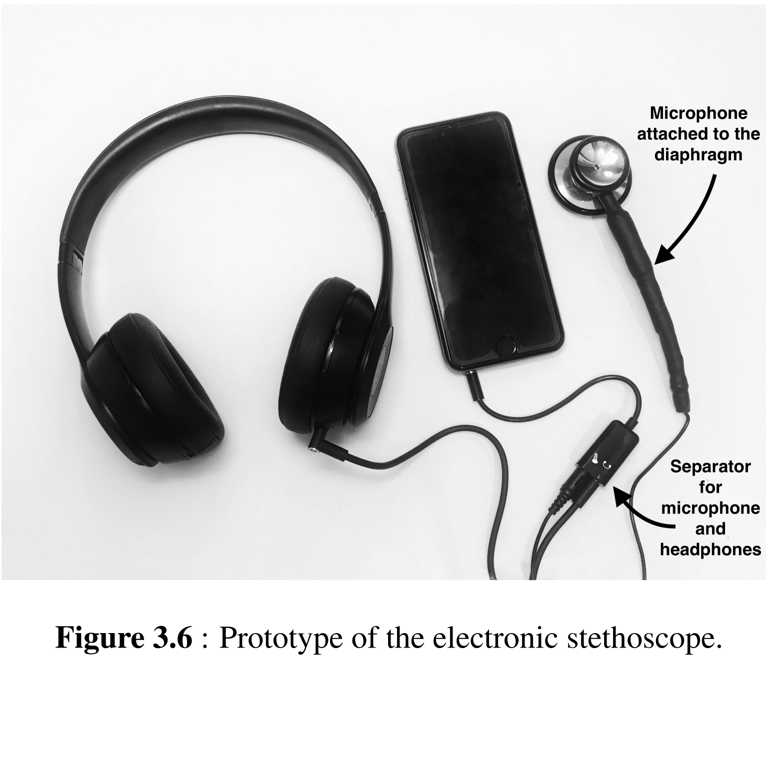
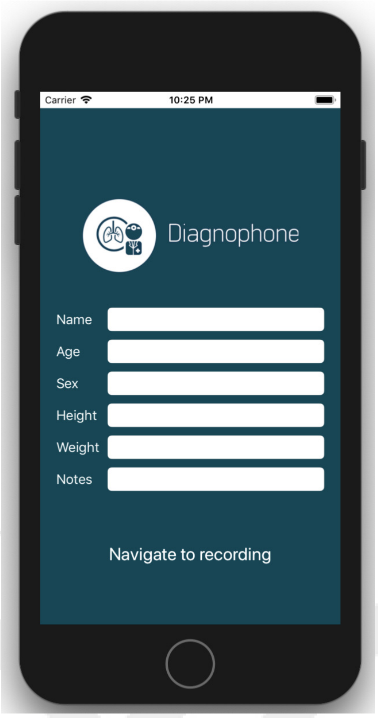
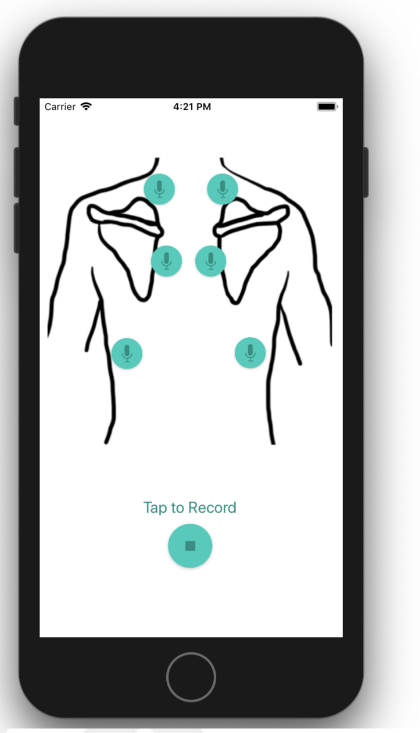


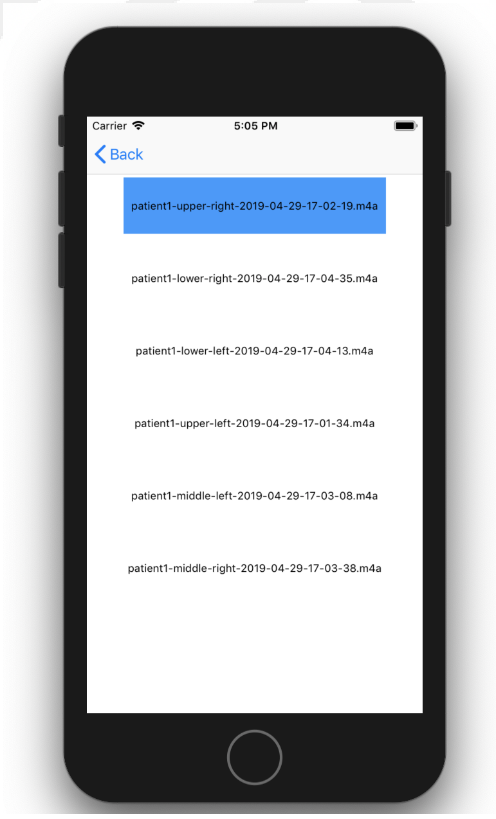
Figure 3.6 : Prototype of the electronic stethoscope.



(a) Information entry about the patient



(b) Recording audio from different auscultation points



(c) Playing previously recorded patient audios

Figure 3.7 : User Interfaces of the Diagnophone application.

4. RESPIRATORY AUDIO ANALYSIS

In this chapter, the audio processing system has been explained in consecutive steps. First, the data has been acquired, and this is followed by the pre-processing, feature extraction and feature selection steps which are then finalized with the classification step.

4.1 Overview Of The Audio Processing System

The overview of the system is shown in Fig.4.1. First, the audio data has been recorded from the electronic stethoscope from 6 different places of the patient's back, which is shown in Fig.3.2 using the electronic stethoscope and the mobile application built, afterwards, the temporal and spectral features have been extracted from the recorded audios signals. Following that step, the demographic information data of the patients have been added as new features and merged with the existing features. Following this step, Principle Component Analysis (PCA) has been applied in order to find the most relevant and efficient features. Finally, classification has been performed using Support Vector Machines (SVM) and as a result, the anomaly has been obtained.

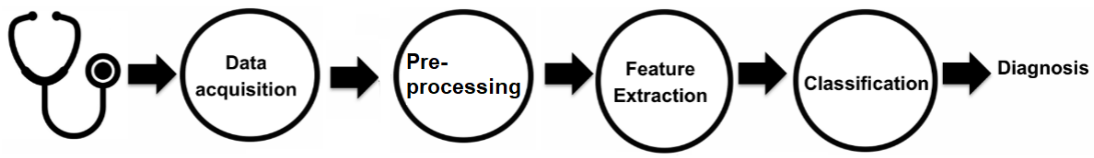


Figure 4.1 : Overview of the audio processing system.

4.2 Dataset Generation

In this thesis, two kinds of datasets have been used. The first dataset is a dataset published bu the Internatonal Conference of Biomedical Health Informatics - ICBHI 2017 [34] which is explained in detail in chapter 5.1.2 has been used. Afterwards, a system to record audio data of the auscultations and in the end create a dataset has been developed. With the help of doctors from Ümraiye Eğitim Araştırma Hastanesi, the audio from 44 patients has been recorded. Along with the audio data, some other demographic information about the patients has also been recorded and the dataset has been created.

4.3 Preprocessing

Both the ICBHI dataset and our data set collected in this thesis has information about which lobe of the lung the data has been obtained from, the age, sex, BMI, height, and weight of the patients along with the audio record itself.

First, the missing data in the dataset related the demographic information (the age, weight of the patient etc.) has been filled with the means of the feature of the same class.

Some of the data with so much background noise that represses breathing sound are eliminated from the training set. In the first dataset, audio with different sampling rates were present so all the audio files were resampled to the same sampling rate of 22000 which also helped to lower the memory need. Some of the 24 bit samples were resampled to 16 bits as well.

When the number of periods in the acquisition is not an integer, the endpoints are discontinuous. These artificial discontinuities show up in the FFT as high-frequency components not present in the original signal. These frequencies can be much higher than the Nyquist frequency and are aliased between 0 and half of the sampling rate. The spectrum acquired by using a FFT, therefore, is not the actual spectrum of the original signal, but a smeared version. It appears as if energy at one frequency leaks into other frequencies. This phenomenon is known as spectral leakage, which causes the fine spectral lines to spread into wider signals.

The effects of performing an FFT over a noninteger number of cycles can be minimized by using a technique called windowing. Windowing reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitizer. Windowing consists of multiplying the time record by a finite-length window with an amplitude that varies smoothly and gradually toward zero at the edges. This makes the endpoints of the waveform meet and, therefore, results in a continuous waveform without sharp transitions. This technique is also referred to as applying a window.

Inside a small enough window, it can be expect that the properties of the signal chunk do not vary too fast. However, when the features extracted from two consecutive frames are examined, the change of property between the frames may induce a discontinuity, or a jump (the difference of parameter values of neighbouring frames can be higher). In order to eliminate such shortcomings, a technique called window overlapping is employed in which some portion of the window is repeated in the next window.

Augmenting datasets by transforming inputs in a way that does not change the label is a crucial ingredient of the state of the art methods for object recognition using neural networks. In order to obtain more training data, data augmentation has been performed with two different techniques. Firstly, audio has been stretched (by speeding up and down randomly) along the time axis. Secondly, after the creation of the Mel spectogram images, using Vocal Tract Length Perturbation (VTLP) with random linear warping, the Mel spectogram images have been transformed along the frequency axis randomly.

Dynamic Time Wrapping (DTW) is one of the most used measures of similarity between two time series originally designed to treat automatic speech recognition. It is an algorithm that calculates an optimal warping path between two time series by calculating both warping path values between the two series and the distance between them using distance matrices.

VLPT is by using warping techniques in which a warp factor is generated randomly each time, during training, rather than fitting a single warp factor to each training and test instance [35]. At test time, a prediction is made by averaging the predictions over multiple warp factors. Given an audio signal $x(t)$, time warping by a factor α gives the signal $x(\alpha t)$. It can be seen from the Fourier transform of $x(\alpha t)$, $\alpha^{-1}\hat{x}(\alpha^{-1}\omega)$ that the warping factor produces shifts in the frequency components of the $\hat{x}(\omega)$by an amount proportional to frequency ω [36]. In VLPT, the speed of the signal is reduced, for values $\lambda < 1$ and there is a shift in the signal energy towards lower frequencies. This results in FFT bins with close to zero energy at higher frequen- cies. This likely means that some of the higher Mel bins end up with very small energies. In this study, we have used VLPT and melspecgogramsin which takes an array of the original mel spaced frequencies and returns a warped version of them (shifts the spectrum in the mel spectrogram). The window size has been selected 25ms with no overlapping and the warping factor $\lambda = 0.9$ for 7500 cut-off frequency. This causes a shift created in the signal energy towards lower frequencies which results in FFT bins with close to zero energy at higher frequencies. That means that some of the higher Mel bins end up with very small energies because we are interested in the change occur in the lower frequencies.

Following that, each audio recording has been divided into breathing cycles and it has been concluded that one breathing cycle takes approximately 5 seconds in 95% of the audio. Therefore, all of the audio has been divided into 5-second chunks and if the divided chunk is less then 5 seconds, the remaining of the audio has been filled with zero padding.

Especially with our data, when the patient firstly touched his body, the noise increase occurred due to exceeding the limit threshold values. This caused clipping and these parts have been identified and reduced from the audio.

After the preprocessed audios have been labeled using one-hot labels, the feature extraction step has been carried out. In this step, various temporal and spectral features of the audios have been calculated along with the Mel Frequency Cepstral Coefficients of the signals. Following this, the most efficient of these attributes have been selected using Principle Component Analysis and used in the classification phase.

In the data recorded, when the diaphragm of the stethoscope first touches the skin of the patient, audio clipping has been observed which can be seen in the red parts of the Fig.4.2.

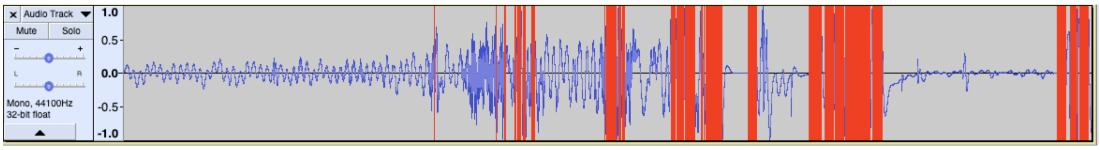


Figure 4.2 : Audio clipping occurring when the diaphragm touches to the patient's skin.

Clipping limits a signal when it exceeds allowed threshold and creates a distortion. It can occur when a signal is digitized. There are two kinds of clipping which are hard and soft clipping. In hard clipping, the signal is strictly limited at the threshold, which produces a flat cutoff whereas, in the soft clipping, the clipped signal continues to follow the original at a reduced gain. Hard clipping results in many high-frequency harmonics; soft clipping results in fewer higher order harmonics and inter-modulation distortion components [37].

In our study, audio parts that clipping has observed have been removed and only the data after the clipping part occurred has been used in order to be classified.

4.4 Feature Extraction

After preprocessing, Mel-frequency Cepstral Coefficients (MFCCs) has been extracted. The name Cepstral comes from the spectral with the "spec" reversed. Cepstrum is the information of rate of change in spectral bands. In the conventional analysis of time signals, any periodic component (for eg, echoes) shows up as sharp peaks in the corresponding frequency spectrum (ie, Fourier spectrum. This is obtained by applying a Fourier transform on the time signal).

On taking the log of the magnitude of this Fourier spectrum, and then again taking the spectrum of this log by a cosine transformation. we observe a peak wherever there is a periodic element in the original time signal. Since we apply a transform on the frequency spectrum itself, the resulting spectrum is neither in the frequency domain nor in the time domain and hence Bogert et al [38] called it the "quefrency" domain. And this spectrum of the log of the spectrum of the time signal was named cepstrum which was first introduced to characterize the seismic echoes resulting due to earthquakes.

Pitch is one of the characteristics of a speech signal and is measured as the frequency of the signal. Mel scale is a scale that relates the perceived frequency of a tone to the actual measured frequency. It scales the frequency in order to match more closely what the human ear can hear (humans are better at identifying small changes in speech at lower frequencies). This scale has been derived from sets of experiments on human subjects. Let me give you an intuitive explanation of what the mel scale captures. The mel scale tries to capture differences in the sound. So the MFCC is defined as the coefficients that make up the Mel-frequency cepstrum.

The Mel-Frequency Cepstrum (MFC) shows the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel Scale of frequency [39]. Here $Mel(f)$ is frequency scale of f's logarithmic scale.

$$Mel(f) = 2595\log_{10}(1+\frac{f}{100}) \tag{4.1}$$

In equation 4.2, MFCCs denoted as C_n are calculated from the Fast Fourier Transform (FFT) power coefficients, which are filtered by the filter bank with a triangular strip passage. The filter bank consists of 12 triangular filters. Here, Filter Bank refers to the mel filters (coverting to mel scale) and Cepstral Coefficients are the MFCCs.

$$C_n = \sqrt{\frac{2}{k}}\sum_{k=1}^{K}(\log_{10}S_k)\cos\left[n(k-0.5)\frac{\pi}{k}\right], n = 1,2,...N \tag{4.2}$$

Here the $S_k(k = 1,2,...K)$ shows the output of the filter banks where (N) is the total number of samples in a 20 s audio unit [40].

The lung sounds in the time domain of 4 classes, namely wheezing, crackles, both crackles and wheezing and normal lung sounds, are shown in Fig.4.3. The X-axis shows the time in seconds, the Y-axis shows the amplitude of the sound wave in Volts. Here, there are differences between the amplitude-time distributions of 5-second audio recordings can be seen.

A spectrogram is a visual representation of the frequency spectrum of the sound as a function of time or another variable. A Mel Spectrogram is a spectrogram that is transformed to have frequencies on the Mel scale, which is a logarithmic scale. In Fig.4.3, the Mel Spectrograms of wheezing, crackles, both crackles & wheezing and healthy lung sounds are shown. The X-axis shows the time in seconds, and the Y-axis shows the frequency in Hz.

To extract the features of each segment, MFCCs are calculated for $n = 12$, each row being formatted as one observation and each column will be an attribute. The coefficients are then scaled to be zero and have unit variance.

After this step, the following temporal and spectral features of the signals have been obtained and added to the existing features.

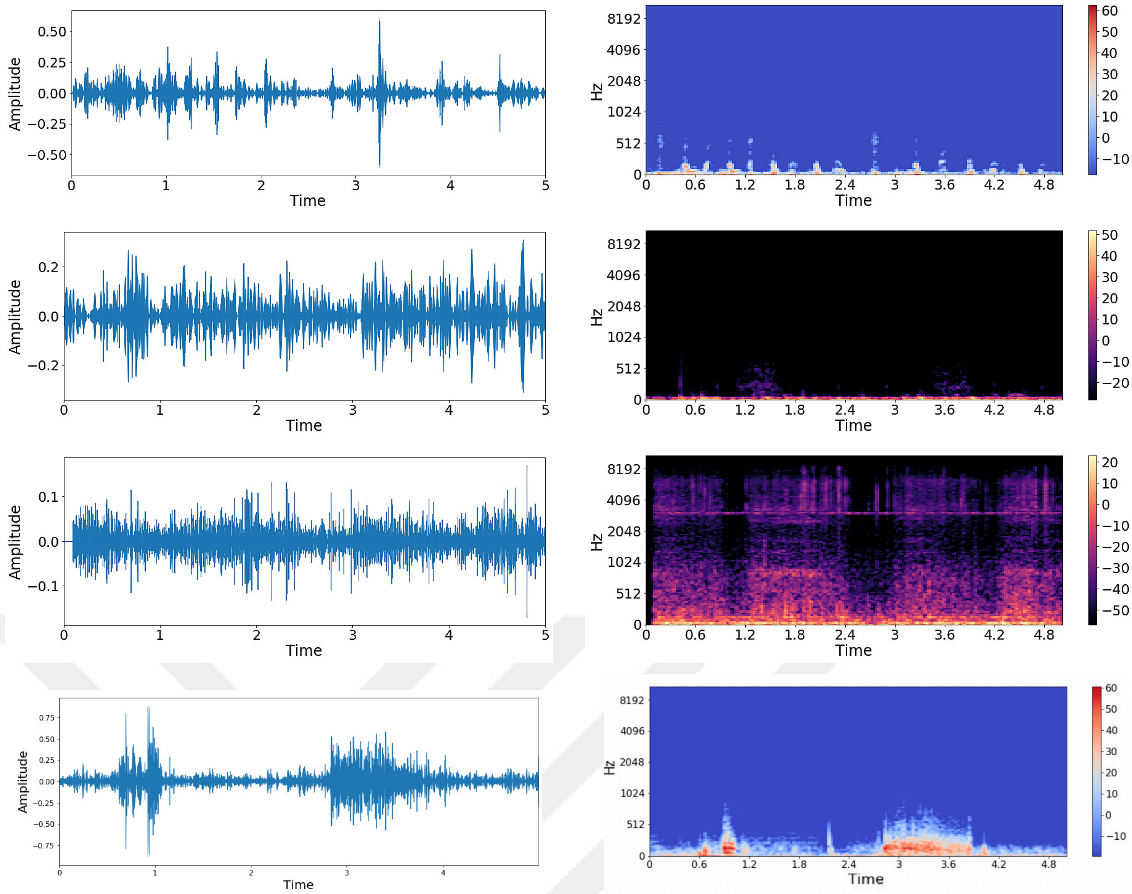


Figure 4.3 : Time-amplitude wave graph and Mel Spectrogram in Frequency domain graphs of crackles, wheezing, healthy patients in order.

Zero Crossing Rate (ZCR): The signal value change rate between positive and negative over time [41].

$$ZCR = \sum_{k=2}^{K} |sgn[x(k)] - sgn[x(k-1)]| \tag{4.3}$$

$$sgn(n) = \begin{cases} 1, n > 0 \\ 0, n = 0 \\ -1, n > 0 \end{cases}$$

Energy (E): Energy: The sum of squares of values of the signal. This is normalized by divided to the frame length.

$$E = \frac{1}{K}\left(\sum_{k=1}^{K} |X(k)^2|\right) \tag{4.4}$$

Spectral Flux (F): Spectral Flux: It shows the spectral change between two consecutive frames. It can be calculated as the squared difference between the normalized magnitudes of the spectra of two consecutive short-term windows [41].

$$F = E(n) - E(n-1) \tag{4.5}$$

Spectral Centroid (SC): This shows the center of the mass of spectrum which shows the character of the spectrum. .

$$SC = \left(\sum_{k=1}^{K} k \cdot |X(k)|^2 \right) / \left(\sum_{k=1}^{K} |X(k)|^2 \right) \tag{4.6}$$

here x denotes the observed data where $p(x)$ denotes the possibility of observing x [41].

Signal Bandwidth (BW): The bandwidth of a signal is defined as the difference between the upper and lower frequencies of the generated signal.

$$BW = \sqrt{\left(\sum_{k=1}^{K} (k - SC)^2 |X(k)|^2 \right)} / \left(\sum_{k=1}^{K} |X(k))|^2 \right) \tag{4.7}$$

Harmonicity (H): Harmonics is a measure of sound energy relative to the rest of the spectrum. It refers to the ratio of the frequency of a signal or wave to the frequency of the reference signal or wave.

$$H(f) = \sum_{k=1}^{K} h_k X(k \cdot f) \tag{4.8}$$

Pitch (P): The pitch is the relative height or low of a tone perceived by the ear; depends on the number of vibrations produced per second.

$$P = f : f >= 0 \Lambda \forall g >= 0, \quad H(f) >= H(g) \tag{4.9}$$

Energy Entropy (S): Measurement of sudden changes in signal. In Equation ref eqn: energy2, entropy is calculated as follows for the sample values of x with x_i. Here, $p(x_i)$ denotes the possibility of observing x_i.

$$S(x) = -\sum_{i} p(x_i).\log(p(x_i)) \tag{4.10}$$

4.5 Feature Selection

After all the attributes obtained with the MFCC, demographic information and temporal and spectral features were combined, the Principal Component Analysis (PCA) method was used to select the important ones.

PCA uses an orthogonal transformation to convert a set of observations of correlated variables into sets of non-linearly related variables called basic components [42].

In the equation below, Y denotes the matrix of scores whereas the W is the coefficients matrix [43].

$$Y = W'X \tag{4.11}$$

which can also be written as

$$y_i j = w_{1i}x_{1j} + w_{2i}x_{2j} + ... + w_{pi}x_{pj} \tag{4.12}$$

The matrix of weights W is calculated from the variance-covariance matrix S which is calculated using formula in Eq.(4.13).

$$S_{ij} = \frac{\sum_{k=1} n(x_{ik} - \bar{x_i})(x_{jk}\bar{x_j})}{n-1} \tag{4.13}$$

In PCA, the singular value decomposition of S is being calculted as in Eq.4.14.

$$U'SU = L \tag{4.14}$$

where L is a diagonal matrix of the eigen-values of S, and U is the matrix of eigen-vectors of S. W is calculated from L and U in equation 4.15 in which W is the scaled version of eigen-vector matrix U, in order to make the variance of each factor y_i equal to one.

$$W = UL^{-\frac{1}{2}} \tag{4.15}$$

The correlation between an i-th factor and the j-th original variable is calculated using Eq.4.16.

$$r_{ij} = \frac{u_{ji}\sqrt{l_i}}{s_{jj}} \tag{4.16}$$

Here u_{ij} is an element of U, l_i is a diagonal element of L, and s_{jj} is a diagonal element of S. The correlations are named the factor loadings that show the relative importance.

When the correlation matrix R is used instead of the covariance matrix S, the equation for Y then becomes as in Eq.(4.17).

$$Y = W'D^{-\frac{1}{2}}X \tag{4.17}$$

where D is a diagonal that matrix consists of the diagonal elements of S. In this case, the correlation formula may be simplified since the s_{jj} are equal to one [43].

4.6 Classification

In this study, 4 classes of respiratory audio signals namely crackles, wheezing, healthy and crackles and wheezing (both) have been classified in two different datasets using different algorithms. Before the model has been trained, the data has been divided into 80% training, 10% test and 10% validation parts. After that, various audio classification techniques have been applied and compared. In the first approach, the features calculated and explained in section 4.4 have been used as an input and the models have been trained using The Support-Vector Machines (SVM), K-Nearest Neighbor (KNN) and MultiClass AdaBoost Decision Tree algorithms which are explained further in detail below. In the second approach, the mel spectograms of the signals have been calculated which can be seen in Fig.4.4 and used as an input for a Convolutional Neural Network (ConvNet/CNN).

Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. CNN treats data as spatial. Instead of neutrons being connected to every neutron of the previous layer, they are instead only connected to the neutron close to it, all have the same weight. The simplification in the connections means the network holds the spatial aspect of the dataset.

The role of the ConvNet is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction. This is important during the design of an architecture which is not only good at learning features but also is scalable to massive datasets. The word convolutional refers to the filtering process that happens in this type of network. A CNN simplifies a complex images it can be better processed and understood. Like a normal neural network, a convolutional neural network is made up of multiple layers. There are a couple of layers that make it unique, the convolutional layer and the pooling layer. However like normal neural networks, it will also have a ReLU or a rectified linear unit layer and a fully connected layer. The ReLU layer acts as an activation function, insuring non-linearity as the data moves through each layer in the network. Without it, the data being fed into each layer would lose the dimensionality that we want to maintain.

The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the Kernel/Filter, K. The filter moves to the right with a certain Stride Value till it parses the complete width. Moving on, it hops down to the beginning (left) of the image with the same Stride Value and repeats the process until the entire image is traversed. The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset. It works by placing a filter over an array of image pixels.This then creates what's called a convolved feature map.

There are two types of results to the operation: one in which the convolved feature is reduced in dimensionality as compared to the input, and the other in which the dimensionality is either increased or remains the same. This is done by applying Valid Padding in case of the former, or Same Padding in the case of the latter. In Same Padding, when a 5x5x1 image is augmented the image into a 6x6x1 image and then the 3x3x1 kernel applied over it, in the result, the dimensions of the convolved matrix would be again 5x5x1. On the other hand in Valid Padding, if the same operation is performed without padding, the end result would be a matrix which has dimensions of the Kernel (3x3x1) itself .

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model. There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel.Max Pooling also performs as a Noise Suppressant. It discards the noisy activations altogether and also performs de-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, it can be concluded that Max Pooling performs a lot better than Average Pooling. The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details even further, but at the cost of more computational power.

The fully connected layer meanwhile, allows to perform classification on the dataset. Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space. After the fully connected layer, the input image has been converted into a suitable form for the Multi-Level Perceptron. After that, the image has been flattened into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the Softmax Classification technique. The CNN architecture used in this study can be seen in Fig.**??** and Fig.**??**

The Support-Vector Machines (SVM) is a supervised learning model that is used for pattern recognition, classification, regression, or other tasks like outlier detection. An SVM model shows the mapped points of examples in space, in a way that the examples of the different categories are divided by a clear gap which is as large as possible. By doing so, the new examples are mapped into the same space and classified by determining which category they belong by looking at which side of the gap they fall [44].

In SVM, a data point is viewed as a p-dimensional vector and the aim is to separate these points with a $(p-1)$-dimensional hyperplane in case of a linear classifier. There may be many hyperplanes that might classify the data. The best one is considered to be the one that creates the largest separation (margin) between the two classes. So the aim is to choose the hyperplane in which the distance from it to the nearest data points on each side is maximum which is called the maximum-margin hyperplane. The linear classifier it defines is named a "maximum-margin classifier"; or the perceptron of optimal stability. The larger the margin gets, the lower the generalization error of the classifier becomes. This hyperplane can be written as in Eq.(4.18), where w is normal to the hyperplane and b is the perpendicular distance between the hyperplane and the origin [45].

$$w \Delta x + b = 0 \tag{4.18}$$

If the data cannot be separatable in linear space, the SVM model can be used to construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, making the separation easier in that space. The dot products of pairs of input data vectors are defined in terms of a kernel function $k(x,y)$ selected according to the problem. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors x_i that occur in the database. With this choice of a hyperplane, the points in the feature space that is mapped into the hyperplane are defined by the relation

When applying our SVM to linearly separable data we have started by creating a matrix H from the dot product of our input variables:

$$H_{ij} = y_i y_j k(x_i, x_j) = x_i \Delta x_j = x_i^T x_j \tag{4.19}$$

$k(x_i, x_j)$ is an example of Kernel Functions. $k(x_i, x_j) = x_i^T x_j$ is known as Linear Kernel. The set of kernel functions are all based on calculating inner products of two vectors.

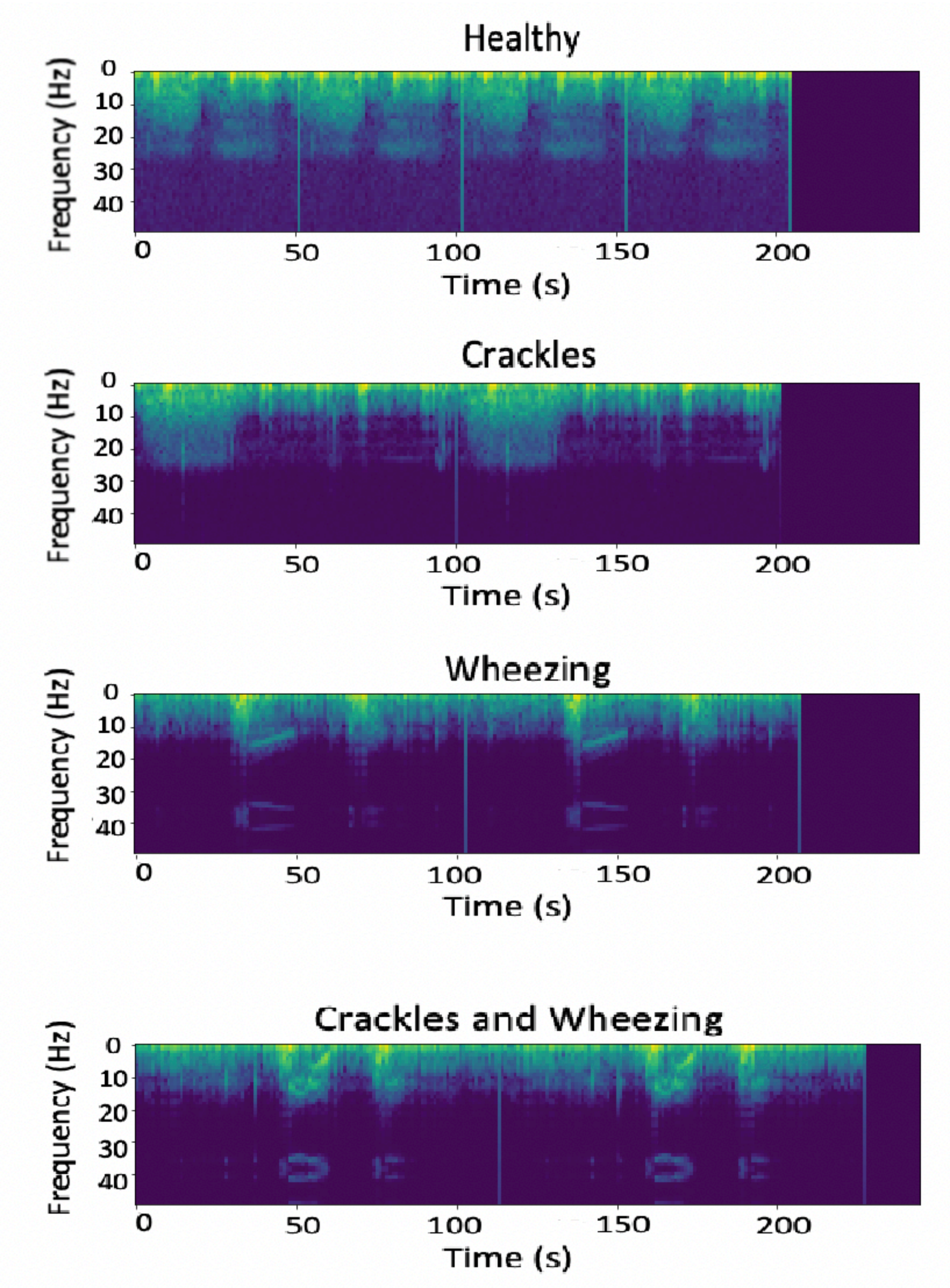


Figure 4.4 : The Mel spectograms calculated with VTLP of 4 different classes.

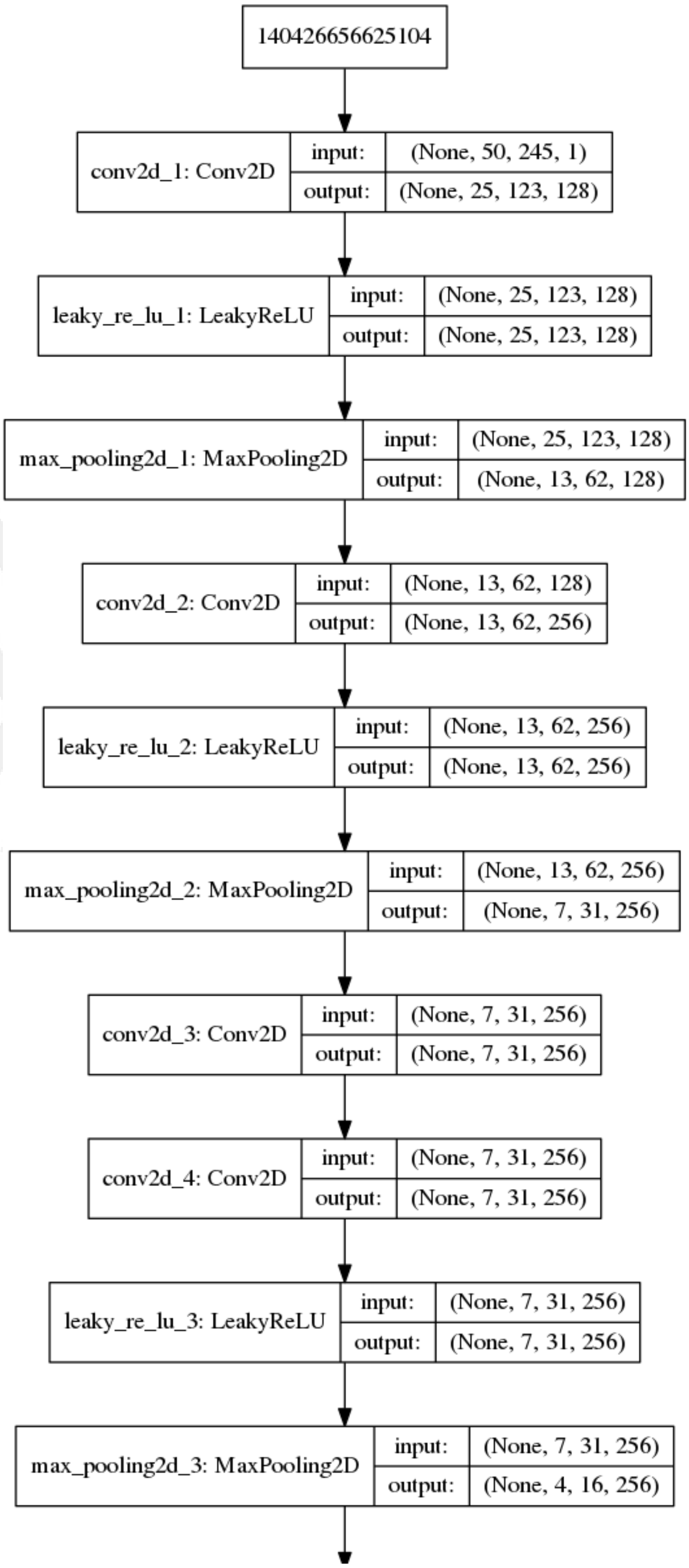


Figure 4.5 : The CNN architecture- part 1.

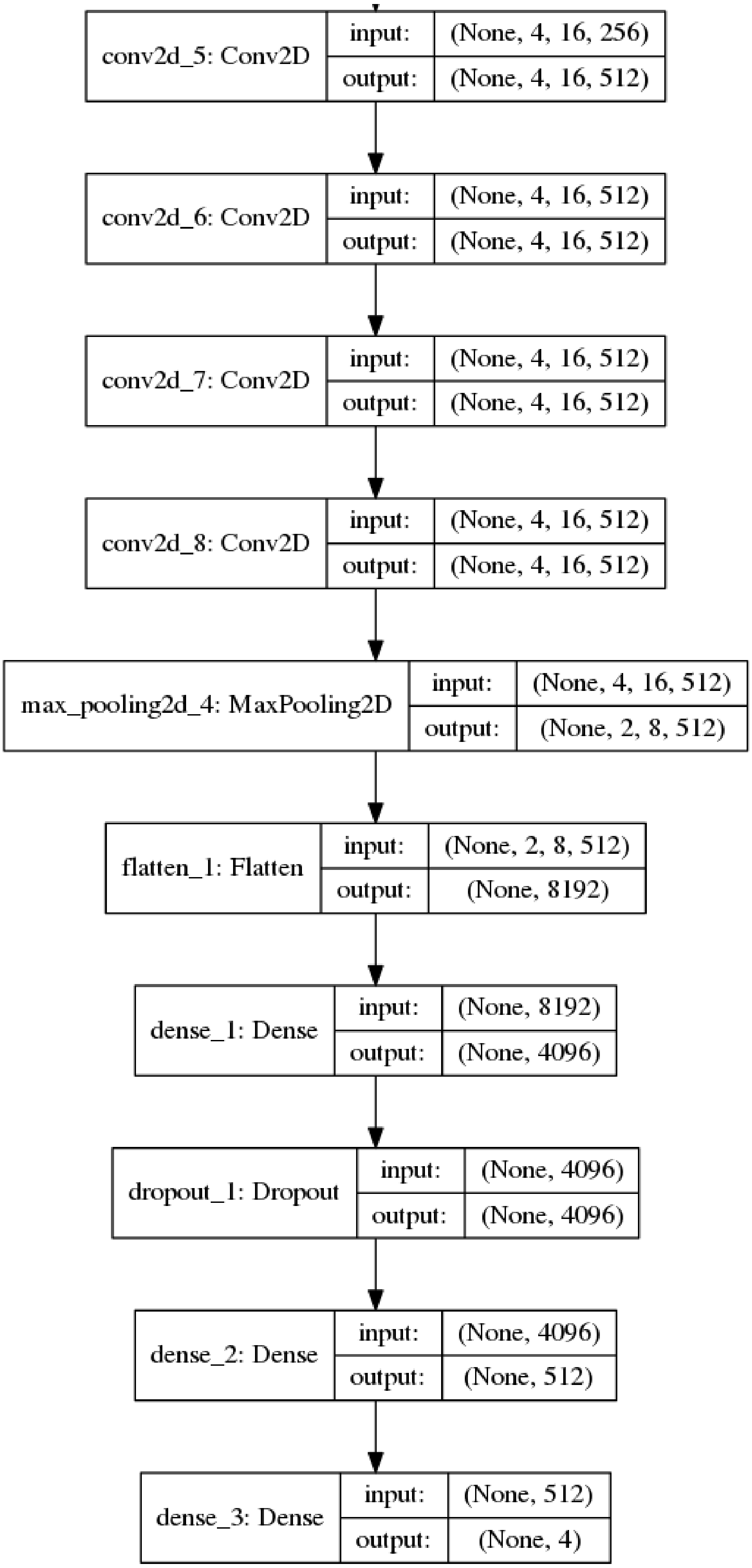


Figure 4.6 : The CNN architecture- part 2.

5. EXPERIMENTS AND RESULTS

In this section, conducted experiments and their results have been discussed in detail. First the user tests performed with the personas has been examined. Then different datasets and data acquisition has been described.

5.1 Experimental Setup

In order to help the doctors with the help of the technology, using the user experience and design thinking methodologies, the problems of the users in their daily lives, the needs of the users, and the requirements of a system that would solve these problems was designed.

To achieve this, in total of 15 doctors were interviewed and observed in their natural environments in the hospital. These were consisting of 2 pediatric emergency doctors, 2 adult emergency doctors, 1 ambulance doctor, 1 internal medicine specialist, 1 specialist pediatrician, 1 home healthcare doctor from Ümraniye Eğitim ve Araştırma Hospital, 2 family healthcare doctors from Zeytinburnu Aile Sağlığı Merkezi and 5 intern doctors from İstanbul Cerrahpaşa Medical School.

After the observations it has been inferred that the problem was that there were too many patients, not enough expert doctors and not enough diagnostic tools available related to pulmonary diseases.

In order to test the user experiences with the Diagnophone, a user experiment was designed in which the methods of questionnaires, think aloud, and observation applied.

5.1.1 Demographic information of participants

The user experience tests were conducted with 7 doctors with different years of experiences in different medical areas with different frequencies of stethoscope usage shown in Table 5.1.

Table 5.1 : Demographic information of participants.

	Age	Gender	Expertise	Frequency of stethoscope usage	Experience in years
User 1	27	Male	Adult ER Doctor	Very Frequent	1
User 2	28	Male	Pediatric ER Doctor	Very Frequent	2
User 3	34	Female	Internal Medicine Specialist	Very Frequent	5
User 4	54	Male	Family Health Doctor	Frequent	19
User 5	23	Male	Intern Doctor	Frequent	0
User 6	39	Male	Intensive care doctor (cardiovascular surgery)	Frequent	10
User 7	24	Female	Intern doctor	Frequent	1

5.1.2 Dataset and data acquisition

In this thesis, two kinds of data has been used. First, a dataset from ICBHI has been used to create a model, and second, our own collected dataset has been created with audio data we have collected from 32 patients. Both are explained in detail below.

The Respiratory Sound database used was created to support the scientific challenge at International Conference on Biomedical Health Informatics - ICBHI 2017 which is freely available for research [34].

The database contains audio samples that are collected by two different teams from two different countries. First one is School of Health Sciences, University of Aveiro (ESSUA) research team at the Respiratory Research and Rehabilitation Laboratory (Lab3R), ESSUA and at Hospital Infante D. Pedro, Aveiro, Portugal.

The second research team collected, respiratory sounds at the Papanikolaou General Hospital, Thessaloniki and General Hospital of Imathia, Greece which are from Aristotle University of Thessaloniki (AUTH) and the University of Coimbra (UC).

The database consists of 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes and 3642 healthy, in 920 annotated audio samples from 126 subjects which is total of 5.5 hours of recordings with 24bits wav files.

The breathing cycles were labeled by the expert doctors weather or not it contains crackles, wheezes, both of them, or no diseases. The audios were recorded using different equipments with different durations ranged from 10s to 90s and from different chest locations Trachea, Anterior left, Anterior right, Posterior left, Posterior right,

	patient_id	diagnosis
0	101	URTI
1	102	Healthy
2	103	Asthma
3	104	COPD
4	105	URTI
5	106	COPD
6	107	COPD
7	108	LRTI
8	109	COPD
9	110	COPD

Figure 5.1 : Diagnosis of patients.

	Beginning_of_respiratory_cycle	End_of_respiratory_cycle	Presence/absence_of_crackles	Presence/absence_of_wheezes
0	1.330	3.804	0	0
1	3.804	6.396	0	0
2	6.396	8.938	1	0
3	8.938	11.580	1	0
4	11.580	14.072	1	0
5	14.072	17.049	1	0
6	17.049	19.490	1	0

Figure 5.2 : Respiratory cycles of patients.

	patient_id	age	sex	adult_bmi	child_weight	child_height
0	101	3.00	F	NaN	19.0	99.0
1	102	0.75	F	NaN	9.8	73.0
2	103	70.00	F	33.00	NaN	NaN
3	104	70.00	F	28.47	NaN	NaN
4	105	7.00	F	NaN	32.0	135.0
5	106	73.00	F	21.00	NaN	NaN
6	107	75.00	F	33.70	NaN	NaN
7	108	3.00	M	NaN	NaN	NaN
8	109	84.00	F	33.53	NaN	NaN
9	110	75.00	M	25.21	NaN	NaN

Figure 5.3 : Demographic info of patients.

Table 5.2 : Distribution of data classes in the first dataset.

Class	Instance Number
Healthy	3642
Crackles only	1864
Wheezes only	886
Crackles and wheezees	506

Table 5.3 : Patient diagnosis results in the first dataset.

Disease	Count
COPD	64
Healthy	26
URTI	14
Bronchiectasis	7
Bronchiolitis	6
Pneumonia	6
LRTI	2
Asthma	1

Table 5.4 : Distribution of respiratory cycle lengths in the first dataset.

Cycle	Time
Longest cycle	16.16
Shortest cycle	0.20
Fraction of samples less than 5 seconds	0.96

Lateral left, Lateral right. In some of the data, the background noise is high which simulates real life.

In the second dataset, respiratory auscultation audio has been collected from 44 patients containing 18 crackles, 5 wheezing, 11 wheezing and crackles (both), 10 healthy patients from 6 different locations namely posterior upper left, posterior upper right, posterior middle left, posterior middle right, posterior lower left and posterior lower right. A total of 370 audio recordings were collected with each audio consists of approximately 20s.

5.2 Results

Table 5.5 : Available samples (breath cycles) in the first dataset after data augmentation.

	Training set	Test Set
Healthy	6094	756
Crackles	6036	379
Wheezing	7310	178
Both	6824	108

Table 5.6 : Available samples (breath cycles) in the second dataset after data augmentation.

	Training set	Test Set
Healthy	1542	192
Crackles	3280	204
Wheezing	2310	56
Both	3296	52

The results of the work has been evaluated in terms of accuracy, recall precision and f1-score which are described in the section 5.2.1. After that, the results of the classification algorithms are presented.

5.2.1 Evaluation criteria results

True Positives (TP) denotes the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes.

True Negatives (TN) denote the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. False positive and false negative values occur when the actual class contradicts with the predicted class.

False Positives (FP) denotes when the actual class is no and predicted class is yes.

False Negatives (FN) occurs when actual class is yes but predicted class in no.

These four parameters are used to calculate Accuracy, Precision, Recall and F1 score.

Accuracy is the most intuitive performance measure and it is a ratio of correctly predicted observation to the total observations. Accuracy is a great measure in case of a symmetric datasets where values of false positive and false negatives are almost same. In case of the unbalanced dataset, the other parameters are useful for the evaluation of the performance of the model.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{5.1}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate.

$$Precision = \frac{TP}{TP+FP} \tag{5.2}$$

Recall (Sensitivity) is the ratio of correctly predicted positive observations to the all observations in actual class.

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 is usually more useful than accuracy, especially in case of an uneven class distribution. In the following equation, P indicates precision where R indicates the recall value.

$$F1 - Score = \frac{2 * R * P}{R + P} \tag{5.4}$$

Macro averaging gives equal weight to each class, whereas micro averaging gives equal weight to each per document classification decision [46]. The macro-averaged results can be computed for a two classed classification result as indicated by the following equations in which P_1 indicates the precision value calculated for the class 1 and R_1 indicates the recall value calculated for class 1.

$$Macro\ Average\ Precision = \frac{P_1 + P_2}{2} \tag{5.5}$$

$$Macro\ Average\ Recall = \frac{R_1 + R_2}{2} \tag{5.6}$$

Micro average on the other hand, takes class weights into account and individual true positives, false positives, and false negatives of the system for different sets is summed up and applied in order to get the statistics. In the following equations, Micro Average Precision and Recall have been shown for 2 classes.

$$Micro\ Average\ Precision = \frac{TP_1 + TP_2}{TP_1 + TP_2 + FP_1 + FP_2} \tag{5.7}$$

$$Micro\ Average\ Recall = \frac{TP_1 + TP_2}{TP_1 + TP_2 + FN_1 + FN_2} \tag{5.8}$$

After PCA was applied, it has been observed that none of the properties of the patient were removed. When the classifications were repeated before and after PCA

Table 5.7 : Accuracy, macro average precision, macro recall average results for first dataset.

Algorithm	Accuracy	Macro Average Precision	Macro Average Recall
CNN	0.81	0.77	0.78
KNN	0.76	0.75	0.79
SVM	0.69	0.67	0.64
AdaBoost	0.69	0.65	0.72

Table 5.8 : Confusion Matrix for CNN in the first dataset.

		Predicted			
		Healthy	Crackles	Hırıltı	Crackles & Wheezing
Actual	Healthy	653	48	33	12
	Crackles	79	277	21	2
	Wheezing	12	27	129	10
	Crackles And Wheezing	3	13	9	84

application, there was an increase of 3% success rate for SVM and 6% for AdaBoost while no success increase was observed for KNN.

There are two different brands of stethoscope diaphragms used during the data collection. In 12 of the patients, Littmann Classic II model, in 32 of the patients, a MicroLife brand two sided classical stethoscope has been used. A majority of the doctors were preferring to use the Littmann stethoscope and when the audio collected from Littmann and a much cheaper brand compared, there were significant differences observed in terms of quality in the digital audio files and overall classification results. It has been observed that the audio audio collected from chest of the patients (arteior right and arterior left lung locations) were containing heart sounds and this was repressing the respiratory sounds which lowered the overall success of the system, so the scope of the project was determined as to process the respiratory audio only acquired from the back of the patients.

Table 5.9 : CNN results for 20 epoch in first dataset.

	Precision	Recall	F1-score	Support
Healthy	0.87	0.88	0.87	756
Crackles	0.76	0.73	0.74	379
Wheezing	0.68	0.72	0.70	178
Crackles and Wheezing	0.78	0.77	0.77	108

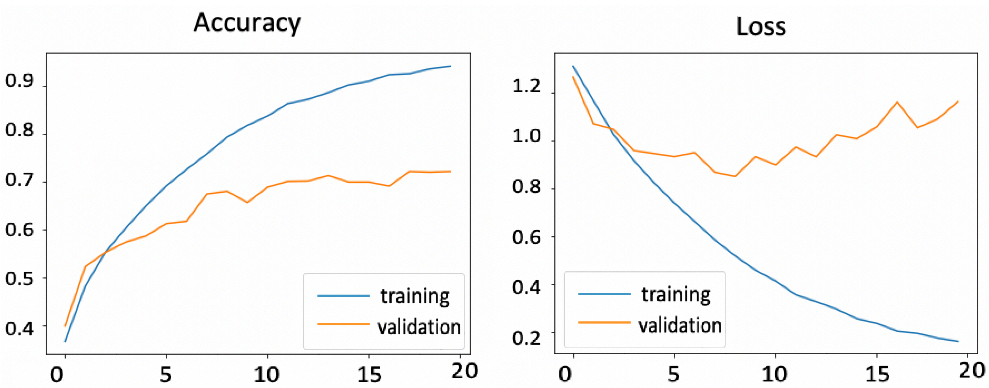


Figure 5.4 : Accuracy and Loss graphs for the first dataset.

Table 5.10 : Accuracy, macro average precision, macro recall average results for second datase.

Algorithm	Accuracy	Macro Average Precision	Macro Average Recall
CNN	0.71	0.71	0.71
SVM	0.67	0.70	0.72
KNN	0.65	0.69	0.70
AdaBoost	0.64	0.68	0.68

Table 5.11 : Confusion Matrix for CNN in the second dataset.

		Predicted			
		Healthy	Crackles	Wheezing	Crackles & Wheezing
Actual	Healthy	25	7	5	0
	Crackles	2	22	0	0
	Wheezing	2	0	14	2
	Crackles & Wheezing	0	0	11	9

Table 5.12 : CNN results for 20 epoch in second dataset.

	Precision	Recall	F1-score	Support
Healthy	0.86	0.68	0.76	37
Crackles	0.76	0.92	0.83	24
Wheezing	0.47	0.78	0.58	16
Crackles & Wheezing	0.82	0.45	0.58	20

Figure 5.5 : Accuracy and Loss graphs for the second dataset.

6. CONCLUSION

Nowadays, AI is revolutionizing the healthcare sector to reduce spendings and improve patient experiences. In this thesis, using the power of AI and UX, a smart electronic stethoscope named Diagnophone is designed to help physicians with the diagnosis of the crackles and wheezing respiratory sounds acquired from the patients' lungs.

The stethoscope is essential in a physician's daily life because this iconic equipment is the first and most frequently used tool for the initial diagnosis.

Even though there are various tests available to diagnose a respiratory disease such as MRI, CT-Scan, Tomography, X-rays, yet the initial diagnosis is always concluded with the help of a stethoscope.

Comparing to these tools, a stethoscope is much more cost-efficient and easier to use. In addition, because MRI, X-rays and other similar equipment are really expensive, unfortunately, they are not easy to find in every health clinic especially in rural areas.

Because of the low availability of these types of equipment, and the serious lack of specialist doctors, the patients cannot reach the results immediately because they have to wait a long time period in a queue.

That's why Diagnophone will be an asset for the physicians especially in rural areas with the lack of expert doctors to consult to or the lack of such expensive medical equipment.

In addition, for example, during MRI, the patient must remain idle for a long time which can be stressful for especially patients with claustrophobia. Furthermore, the patients who benefit from all these equipment are exposed to intense amount of radiation during the use of these types of equipment. In the case of the tomography, this rate is even greater.

As a result of the doctor interviews, it is concluded that many diseases can be understood only by listening with a stethoscope, yet still, these tests are being applied by these types of equipment in order to validate the doctor's initial diagnosis. The

Diagnophone will assist physicians with the diagnose while providing them a much simpler tool for validation.

Today, there are different electronic stethoscope models created in the market but they are approximately 10 times the price of the regular stethoscopes and they do not introduce and anomaly detection and classification feature.

What distinguishes our study from the ones in the market and in the literature is that Diagnophone is built with User Experience and Human-Computer Interaction principles in mind in a way best offers usability for a physician. The requirements of the Diagnophone are identified after user tests, surveys and interviews.

In this thesis, with the help of 15 doctors contacted, the user tests have been conducted in terms of human-computer interaction and usability, the needs were determined and the design was developed according to the results.

With this stethoscope, the lung sound can be heard and recorded at the same time. This feature can be used for teleconferencing when there is no specialist physician to be consulted, and also for storing the patient's audio data.

Nowadays, patients have access to the old test results (eg, blood tests, tomography results, etc.). However, since data such as the previous lung sound is not stored in hospitals, the sound data cannot be used in the follow-up of the disease. However, in patients requiring follow-up such as asthma, storing the lung sound of the patient will facilitate the follow-up of the disease and will also help to determine the condition of the disease compared to the previous.

The audio recorded with Diagnophone can also be used in medical education. Today's medical education consists of listening to the lung sounds of the patients with the same stethoscope by the students, after the teachers, who are gathered at the beginning of the patient's bed. As a result of the user interviews, it was inferred that this was not an effective type of learning. However, the sounds recorded from the patient through Diagnophone can be played with the speaker of the mobile phone by the teacher while pointing out the important parts to the students or these anomalies can be identified by the Diagnophone. That way better learning experience for the students will be achieved and this will make medical education more efficient and understandable for the students.

Before the development of this system, the users were interviewed before, during and after the implementation, by testing various prototypes of the system while various observations were made relating the experience of the users and the success of the experiments with the questionnaires was measured. As a result, the design was evolved and finalized.

In order to create a design that satisfies all the needs of the physicians in terms of user experience and human-computer interaction principles, 10 doctors and 5 medical students from several hospitals have been contacted and interviewed. After the requirements are identified, with the help of the electronic stethoscope and the mobile app built, data has been recorded during the auscultation and labeled by the doctors from 44 patients from 6 different lobes of the lungs in a total of 370 recordings approximately 10ms each.

Firstly, various experiments were performed by using a dataset publicly available for research consisting of lung sounds. After the completion of the hardware and mobile application required for the creation of our own data, the audio data we have collected has been combined with this data and various experiments have been carried out. A portion of the data is segmented in the training of the data and a part is used in the test.

The obtained audio data is then put to the preprocessing phase in which clippings were removed. Especially when the patient firstly touched his body, the noise increase occurred due to exceeding the limit threshold values. This part of the sound is trimmed and the part containing the main lung sound is taken into operation. Mel spectograms and different features have been calculated from each audio and used during the training of the model such as temporal shape, temporal features, energy features, spectral shape features, harmonic features, perceptual features along with the MFCCs and demographic information of the patients.

Following this, the most important of these attributes have been identified using Principle Component Analysis so that the success of the model is improved by 3% by using only these features.

After this process, the success of CNN, AdaBoost, KNN, and SVM algorithms are compared for the dataset. As a result, the SVM algorithm performed best by obtaining 81% accuracy.

In the studies, it was observed that lung auscultation was done both at the front and the back of the body. For this reason, the sounds collected from both the back and chest regions of the patients were processed. However, in the sound files collected from the chest region, it was observed that heart sounds suppressed lung sounds and this suppressed overall success. As a result of observations and interviews with physicians, it was inferred that even though theoretically it is necessary to listen to the chest during the auscultation, in practice, the doctors only make lung auscultation from the back area of the patients and bypass the chest area. For these two reasons, only the sounds received from the back region were adopted within the scope of the project.

As future work, the heart sounds will be identified and eliminated from the audio. This data then will be able to be used during the training or test of the system. Also, the data collection will be concluded with for the noise canceling, an improvement for the system in the hardware can be applied.

REFERENCES

[1] **3M-Littman**, (2019(accessed March 13, 2019)). 3M^{TM} Littmann® Electronic Stethoscope, `https://www.littmann.com/3M/en_US/littmann-stethoscopes/products/~/3M-Littmann-Electronic-Stethoscope-Model-3100/?N=5932256+8711017+3293188506+3294857497&preselect=8727094+3293786499&rt=rud`.

[2] **3M-Littman**, (2019(accessed March 13, 2019)). 3M^{TM} Littmann® Classic II Stethoscope, `https://www.littmann.com/3M/en_US/littmann-stethoscopes/products/~/3M-Littmann-Classic-II-Pediatric-Stethoscope/?N=5932256+8711017+3293188503+3294857497&preselect=8727096+3293786499&rt=rud`.

[3] **Organization, W.W.H.**, (2016(accessed March 13, 2019)). Top 10 Causes Of Death, `https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death`.

[4] **Statistics, T.I.O.** (2018). Statistics Of Causes Of Death In Turkey, 2017, *27620*, 1–2.

[5] **of Health, U.D. and Services, H.**, (2018(accessed March 13, 2019)). Computed Tomography (CT), `https://www.fda.gov/Radiation-EmittingProducts/ucm115317.htm`.

[6] **Rene, A.** (1819). Laennec's new system of diagnosis, *The Quarterly Journal of Foreign Medicine and Surgery and of the Sciences Connected with Them*, *2*(2), 51–68.

[7] **Roguin, A.** (2006). Rene Theophile Hyacinthe Laënnec (1781–1826): The Man Behind the Stethoscope, *Clin Med Res*, *4*(2), 230–235.

[8] **Tiwari, H.K. and Harsola., A.** (2016). Development of embedded stethoscope for Heart Sound., *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 1547–1551.

[9] **Turkey Higher Education Board, Turkey Ministry of Health, T.M.o.F..T.M.o.D.** (2014). Health Education and Health Manpower Status Report of Turkey, *1*(1), 1–2.

[10] **ROGERS, D.J.** (2012). Electronic Stethoscopes With User Selectable Digital Filters, *World Intellectual Property Organization ,3M Center Office of Intellectual Property Counsel*.

[11] **3M-Littman**, (2011(accessed March 13, 2019)). Sound Strategy, `https://multimedia.3m.com/mws/media/749312O/history-of-3m-littmann-electronic-stethoscopes.pdf`.

[12] **Murphy, R.** (2007). Computerized Multichannel Lung Sound Analysis, Development of Acoustic Instruments for Diagnosis and Management of Medical Conditions, *IEEE Engineering In Medicine And Biology Magazine*, 16–19.

[13] **Maurice BRappaport E.E., H.B.M.** (1941). Physiologic and physical laws that govern auscultation, and their clinical application: The acoustic stethoscope and the electrical amplifying stethoscope and stethograph, *American Heart Journal*, *21*, 257–318.

[14] **Leng S, Tan RS, C.K.W.C.G.D.Z.L.** (2015). The electronic stethoscope, *Biomed Eng Online*, *14*, 0–66.

[15] **Meditron, T.W.A.**, (2018(accessed March 13, 2019)). The Stethoscope, `https://www.welchallyn.com/content/dam/welchallyn/documents/upload-docs/Training-and-Use/User-Manual/Master\%20Elite\%20Plus,\%20User\%20Manual.pdf`.

[16] **3M-Littman**, (2012(accessed March 13, 2019)). Redefining what a stethoscope can do for you, `https://multimedia.3m.com/mws/media/843622O/littmann-electronic-stethoscopes.pdf`.

[17] **ThinkLabs**, (2015(accessed March 13, 2019)). ThinkLabs One Digital Stethoscope, `https://www.thinklabs.com`.

[18] **Health, E.**, (2019(accessed March 13, 2019)). An Acoustic + Electronic Stethoscope. The Best of Both Worlds., `https://ekohealth.com/shop/core/`.

[19] **Health, E.**, (2019(accessed March 13, 2019)). Eko DUO ECG & Digital Stethoscope, `https://ekohealth.com/shop/duo/`.

[20] **ISO**, `https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en`, 2010(accessed April 20, 2019).

[21] **Olsson, T. and Salo, M.** (2011). Online user survey on current mobile augmented reality applications, *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pp.75–84.

[22] **Arhippainen, L. and Tähti, M.** (2003). Empirical evaluation of user experience in two adaptive mobile application prototypes, *Proceedings of the 2nd International Conference on Mobile*.

[23] **Gube, J.**, What Is User Experience Design? Overview, Tools And Resources., `https://www.smashingmagazine.com/2010/10/what-is-user-experience-design-overview-tools-and-resources/`, 2010(accessed April 20, 2019).

[24] **Krusche, S. and Bruegge, B.** (2014). User Feedback in Mobile Development, *Proceedings of the 2Nd International Workshop on Mobile Development Lifecycle*, MobileDeLi '14, ACM, New York, NY, USA, pp.25–26.

[25] **Palen, L. and Salzman, M.** (2002). Voice-mail Diary Studies for Naturalistic Data Capture Under Mobile Conditions, *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, CSCW '02, ACM, New York, NY, USA, pp.87–95, `http://doi.acm.org/10.1145/587078.587092`.

[26] **Morten Grønnesby, Juan Carlos Aviles Solis, E.H.H.M.L.A.B.** (2017). Feature Extraction for Machine Learning Based Crackle Detection in Lung Sounds from a Health Survey, *UiT The Arctic University of Norway*.

[27] **Hiroshi Ono, Yasuyuki Taniguchi, K.S.T.S.S.K. and Gemma, A.** (2009). Evaluation of the Usefulness of Spectral Analysis of Inspiratory Lung Sounds Recorded with Phonopneumography in Patients with Interstitial Pneumonia, *Journal of Nippon Medical School*.

[28] **Gorkem Serbes, C. Okan Sakar, Y.P.K.N.A.** (2013). Pulmonary crackle detection using time–frequency and time–scale analysis, *Digital Signal Processing*, (2), 1012–1021.

[29] **A. Sinharay, D. Ghosh, P.D.S.A.R.B. and Pal, A.** (2016). Smartphone Based Digital Stethoscope for Connected Health – A Direct Acoustic Coupling Technique, *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 193–198.

[30] **Kahya YP, Yeginer M, B.B.** (2006). Classifying respiratory sounds with different feature sets., *Conf Proc IEEE Eng Med Biol Soc*, (1).

[31] **C. Aguilera-Astudillo, M. Chavez-Campos, A.G.S. and Garcia-Cordero, J.L.** (2016). A low-cost 3-D printed stethoscope connected to a smartphone, *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando*, 4365–4368.

[32] **Shneiderman, B., P.C.C.M.J.S. and N., E.** (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Sixth Edition*, Pearson.

[33] **Sachpazidis, I.** (2008). Image and Medical Data Communication Protocols for Telemedicine and Teleradiology.

[34] **B. M. Rocha, D. Filos, L.M.I.V.E.P.E.K.P.N.A.O.C.J.A.M.R.P.P.I.C.P.C. and Maglaveras, N.** (2017). A Respiratory Sound Database for the Development of Automated Classification, *In Precision Medicine Powered by pHealth and Connected Health, Int. Conf. on Biomedical Health Informatics*, (1), 33–37.

[35] **Navdeep Jaitly, G.E.H.** (2013). Vocal Tract Length Perturbation (VTLP) improves speech recognition.

[36] **Tom Ko1, Vijayaditya Peddinti, D.P.S.K.** (2015). Audio Augmentation for Speech Recognition, *Sixteenth Annual Conference of the International Speech Communication, INTERSPEECH 2015*.

[37] **Zottola, T.** (1996). *Vacuum Tube and Guitar and Bass Amplifier Servicing. Bold Strummer.*, Pan American Union.

[38] **B.P. Bogert, M.H. and Tukey, J.** (1963). *The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking, in Time Series Analysis*, M. Rosenblatt.

[39] **Min Xu, Kiyoharu Aizawa, Y.N.S.S.** (2004). HMM-based Audio Keyword Generation In Advances in Multimedia Information Processing, *5th Pacific Rim Conference on Multimedia*.

[40] **Fang Zheng, G.Z. and Song., Z.** (2001). Comparison of Different Implementations of MFCC, *J. Computer Science & Technology*, *16*, 582–589.

[41] (2014). Introduction to Audio Analysis, **T. Giannakopoulos and A. Pikrakis**, editors, Introduction to Audio Analysis, Academic Press, Oxford, p. i.

[42] **Pearson, K.** (1901). On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine*, (2), 559–572.

[43] **Software, N.S.**, (2019(accessed March 20, 2019)). Principal Components Analysis, `https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Principal_Components_Analysis.pdf`.

[44] **Corinna Cortes, V.V.** (1995). Support-vector networks, *Machine Learning*, *20*, 273–297.

[45] **Fletcher, T.** (2008). Support Vector Machines Explained, *UCL*.

[46] **Tsoumakas, G., K.I..V.I.P.** (2010). *Mining multi-label data. In O. Maimon, & L. Rokach (Eds.) Data Mining and Knowledge Dis- covery Handbook*, Springer-Verlag, 2 edition.

APPENDICES

APPENDIXA.1

Pre-Test Demographic From

Demographic Information Form

Please fill the form below. Your answers to the following questions will help the study researchers to analyze the test results.

Participant ID	(prefilled)
Age	
Gender	Female ▢ Male ▢
Country	
Expertise	
Frequency of the usage of similar technology (in this case regular stethoscope)	
Experinece in related area	

APPENDIXA.2

User Survey Questionnaire

The purpose of this questionnaire is to help us gain an understanding of the people who will use our stethoscope, and to get any additional feedback or comments about our product. We will use this information to try to ensure that our product meets the needs of the people who will be using it. All the information you provide is confidential. Your name is not stored with this questionnaire, and the information you provide will not be used for any other purpose. On the second question, please use the product reaction cards provided.

1. Do you have an electronic stethoscope or ever used one before?
2. Would you recommend this stethoscope to a friend?
3. How would you describe this stethoscope in one words?
4. If this stethoscope were a car, what car would it be?
5. Which brand of stethoscope are you using?
6. Have you always used that brand?
7. Have you ever used an electronic stethoscope?
8. How does this stethoscope compare to Littmann electronic stethoscopes?
9. Why choose us over Littmann?
10. How does Diagnophone compare to other stethoscopes you have used?
11. If you were to review this stethoscope what score would you give it out of 10?
12. What do you find most frustrating about this stethoscope?
13. Overall, how easy to use do you find this stethoscope?
14. If you could change one thing about this stethoscope what would it be and why?
15. How easy is Diagnophone to use?
16. When you think about Diagnophone, do you think of it as something you need or don't need?
17. What features could you not live without?
18. What do you like best about this stethoscope?
19. What do you like least about this stethoscope?
20. How can we improve this stethoscope? Tell me your ideas and suggestions.
21. Anything else you care to share or get off your chest?

APPENDIXA.3

User Survey Questionnaire

Product Reaction Cards

Joey Benedek and Trish Miner
Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052
joeyb@microsoft.com and trishmi@microsoft.com

The following table contains all of the words used on the product reaction cards described in the paper *Measuring Desirability: New methods for measuring desirability in the usability lab setting*.

If you choose to use these cards for your own research, we are very interested in your experience, so we can continue to refine the method. Please contact us and let us know how it works for you.

The complete set of 118 Product Reaction Cards				
Accessible	Creative	Fast	Meaningful	Slow
Advanced	Customizable	Flexible	Motivating	Sophisticated
Annoying	Cutting edge	Fragile	Not Secure	Stable
Appealing	Dated	Fresh	Not Valuable	Sterile
Approachable	Desirable	Friendly	Novel	Stimulating
Attractive	Difficult	Frustrating	Old	Straight Forward
Boring	Disconnected	Fun	Optimistic	Stressful
Business-like	Disruptive	Gets in the way	Ordinary	Time-consuming
Busy	Distracting	Hard to Use	Organized	Time-Saving
Calm	Dull	Helpful	Overbearing	Too Technical
Clean	Easy to use	High quality	Overwhelming	Trustworthy
Clear	Effective	Impersonal	Patronizing	Unapproachable
Collaborative	Efficient	Impressive	Personal	Unattractive
Comfortable	Effortless	Incomprehensible	Poor quality	Uncontrollable
Compatible	Empowering	Inconsistent	Powerful	Unconventional
Compelling	Energetic	Ineffective	Predictable	Understandable
Complex	Engaging	Innovative	Professional	Undesirable
Comprehensive	Entertaining	Inspiring	Relevant	Unpredictable
Confident	Enthusiastic	Integrated	Reliable	Unrefined
Confusing	Essential	Intimidating	Responsive	Usable
Connected	Exceptional	Intuitive	Rigid	Useful
Consistent	Exciting	Inviting	Satisfying	Valuable
Controllable	Expected	Irrelevant	Secure	
Convenient	Familiar	Low Maintenance	Simplistic	

APPENDIXA.4

Post Test User Satisfaction Form

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	○	⊗	○	○	○	○	○	unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the product now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

CURRICULUM VITAE

Name Surname: Ege Yağ Çakır

Place and Date of Birth: Samsun/TURKEY, 17.01.1992

Adress: Kadıköy, Istanbul-TURKEY

E-Mail: egeyag@gmail.com

M.Sc.: Computer Engineering in Istanbul Technical University, 2019

B.Sc.: Computer Engineering in Yıldız Technical University, 2015

WORK EXPERIENCE

- 2015-2017: Software Engineer- Garanti Technology
- 2017-present: Software Engineer- Yapı Kredi Bank