

SmartScope: An AI-Powered Digital Auscultation Device To Detect Cardiopulmonary Diseases

Abstract

Cardiopulmonary diseases are leading causes of death worldwide, accounting for nearly 15 million deaths annually. Accurate diagnosis and routine monitoring of these diseases by auscultation are crucial for early intervention and treatment. However, auscultation using a conventional stethoscope is low in amplitude and subjective, leading to possible missed or delayed treatment. This research aimed to develop a stethoscope called SmartScope powered by machine-learning to aid physicians in rapid analysis, confirmation, and augmentation of cardiopulmonary auscultation. Additionally, SmartScope helps patients take personalized auscultation readings at home effectively as it performs an intelligent selection of auscultation points interactively and quickly using the reinforcement learning agent: Deep Q-Network. SmartScope consists of a Raspberry Pi-enabled device, machine-learning models, and an iOS app. Users initiate the auscultation process through the app. The app communicates with the device using MQTT messaging to record the auscultation, which is augmented by an active band-pass filter and an amplifier. Additionally, the auscultation readings are refined by a Gaussian-shaped frequency filter and segmented by a Long Short-Term Memory Network. The readings are then classified using two Convolutional Recurrent Neural Networks. The results are displayed within the app and LCD. After the machine-learning models were trained, 90% accuracy for cardiopulmonary diseases was achieved, and the number of auscultation points was reduced threefold. SmartScope is an affordable, comprehensive, and user-friendly device that patients and physicians can widely use to monitor and accurately diagnose diseases like COPD, COVID-19, Asthma, and Heart Murmur instantaneously, as time is a critical factor in saving lives.

1. Introduction

Respiratory and cardiovascular diseases are some of the leading causes of death globally, as nearly 15 million people die from them each year [5]. Early diagnosis and routine monitoring of cardiopulmonary diseases are crucial for intervention and treatment. Both cardiac and lung auscultation have been identified as a non-invasive and safe technique for detection. Despite the advantages of low cost and easy operation, traditional auscultation has certain limitations. Auscultation itself is subjective to false interpretations, variability between different physicians, instrument errors, and substantial cost [8]. Subjectivity is the most crucial issue as the "average number of correct detections [during] auscultation is a mere 36.5% for pulmonologists" [1]. Furthermore, the study concluded that the diagnosis of wheezes and rhonchi had a variability of nearly 30% among different physicians, which could be determinantal for patients as the diagnosis could delay access to specific treatments for serious diseases like asthma or chronic bronchitis. Additionally, heart and lung sound by traditional auscultation are low in amplitude and often distorted with ambient

noise leading to missed or delayed treatments. Although there have been technological advances in electronic auscultation devices, "digital auscultation is not yet entirely computational" [6]. Current solutions focus mainly on noise amplification and cancellation. For example, the EKO Core or the 3M Littmann 3200BK only focuses on noise amplification and cancellation features. Most electronic stethoscopes only focus on a specific feature, but no solution encompasses all the features in one user-friendly device. Also, the procedures like echo, MRI, and CT scans are usually only available in large institutions, leaving underserved areas without such advanced technologies to aid in the diagnosis and treatment of cardiopulmonary diseases. Thus, a physician can utilize a device like SmartScope for a timely diagnosis and treatment of cardiopulmonary diseases.

SmartScope achieves higher accuracy by using a novel preprocessing algorithm to segment the recorded auscultation using a Long-Short Term Memory Network into different regions, such as S1, Systole, S2, and Diastole for cardiovascular auscultation, and inspirations and expirations for respiratory auscultation. Identifying the specific regions removes any extraneous information so that the Convolutional Recurrent Neural Networks can achieve higher accuracy with less computation. Additionally, the segmentation for lung sounds was created, which currently does not exist in any other digital stethoscope. After the segmentation, the audio file is filtered by a Gaussian-shaped frequency filter. Not only software algorithms were utilized to augment the sound files but also hardware components, such as an amplifier to increase the sound and an active band-pass filter to remove any sound outside of the heart and lungs frequencies. All these novel preprocessing techniques assist the Convolutional Recurrent Neural Networks to classify cardiopulmonary diseases more accurately by eliminating subjectivity. The iOS app provides user-friendly access to the SmartScope device to take the reading and view the results instantaneously, making it easier for not only the physicians to diagnose the disease accurately but also the patients to monitor their heart and lung health and transfer the reading to doctors for further evaluation. Lastly, this solution is an affordable choice costing about \$50 compared to current solutions, which cost nearly \$500 plus an additional fee for diagnosis each month.

2. Materials and Methods

2.1 Hardware

The hardware's main component is Raspberry PI 4 powered by a 3.7v 4500mah Lipo rechargeable battery. An active band-pass filter is used to remove any sound outside of heart and lungs frequencies. It has two circuits for each mode: heart and lung. In heart mode, the band-pass filter controls a frequency range of 20 Hz-500 Hz, while in lung mode, it is 60 Hz-4000 Hz. Additional electronic components include the DAC HAT Sound Card and an LCD touchscreen. The device also includes the bell of a conventional stethoscope with an extra microphone added to it. The hardware components except the stethoscope bell are housed in a custom-built 3D-printed case designed using FreeCAD.

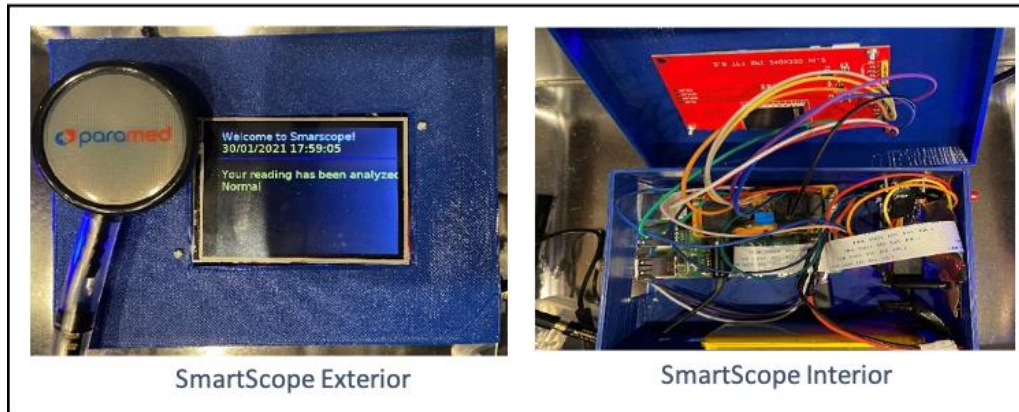


Figure 1: Exterior and Interior of SmartScope

2.2 Preprocessing Methods

The machine learning algorithm consists of multiple Python classes to create modularity and reduce the overall architecture's complexity. The first Python class called "Preprocessing" contains several methods that apply novel algorithms to filter and extract features from the stethoscope's audio. The first step is to convert the waveform into a spectrogram, a 2D tensor, which is "obtained using the short-time Fourier transform since it describes the evolution of the frequency components over time" [6]. The second step is to implement a noise suppression algorithm using Gaussian-shaped frequency filter for the spectrogram images using SciPy to limit their amplitude bandwidth [3]. To further remove any noise from the audio, the filter applies a standard deviation of 75 Hz and 150Hz for heart and lung sounds to ensure 99.7% of the frequencies are within the respective ranges.

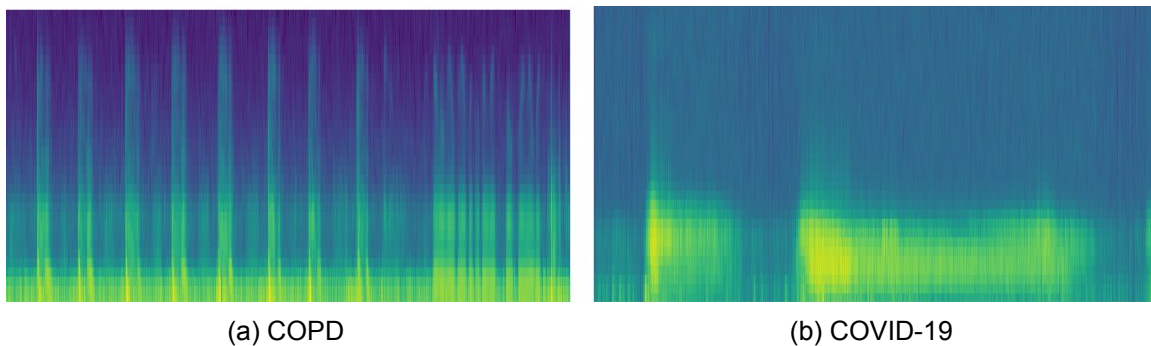


Figure 2: Spectrogram Images

The final step is to apply "feature segmentation, [which] is an essential step in the automatic analysis of heart sound recordings" [7]. As cardiovascular auscultation consists of these four fundamental sounds (S1, Systole, S2, and Diastole), identifying the regions between the S1 and S2 sounds allows for abnormal sound analysis such as clicks and murmurs. This segmentation was provided using a Long Short-Term Memory Network (LSTM). To apply this method, the audio from the stethoscope is randomly segmented into different

regions. The LSTM will evaluate each region based on the S1, Systole, S2, and Diastole sounds. Then S1 and S2 sounds are removed from the audio and the remaining segments (Systole and Diastole) are joined together. As respiratory auscultation does not have de-fined features, a novel algorithm was created to segment the auscultation audio into di erent inspirations and expirations based on their amplitude and duration.

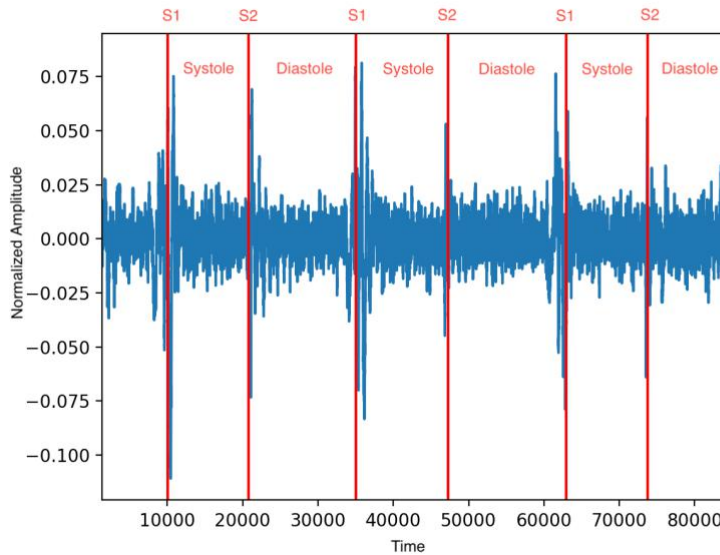


Figure 3: Segmentation of Cardiovascular Auscultation

2.3 Machine Learning Model

The primary component of SmartScope is the class, named "Model," a novel machine learning algorithm that implements separate convolutional recurrent networks for respiratory and cardiovascular auscultation using the TensorFlow and Keras Library. The fundamental layer was ConvLSTM2D, which was primarily used to compute convolutional operations on recurrent transformations for the most optimal time-series forecasting and analysis [2]. The first layer is part of the Keras Input class and receives the respective spectrogram image. The second and third layers resize the images and apply a normalization method to convert each pixel to a value between 0 and 1, respectively. The next four layers are constructed using two ConvLSTM2D and MaxPooling2D layers. The first LSTM convolutional layer applies 128 2x2 filters with the Leaky ReLU activation function, and the second one applies 64 3x3 filters with the ReLU activation function. The max-pooling layers utilize a pool size of 2x2. The next layer is the dropout layer, with a rate of 0.25 to prevent overfitting. The next four layers are three dense layers with 256, 512, and 256 units, followed by a dropout layer with a rate of 0.25. The final dense layer contains three units for each respiratory disease with SoftMax as its activation function. Similarly, cardiovascular auscultation follows the same design but utilizes a different set of parameters. Both models are compiled with Adam's optimizer and a sigmoid cross-entropy loss function.

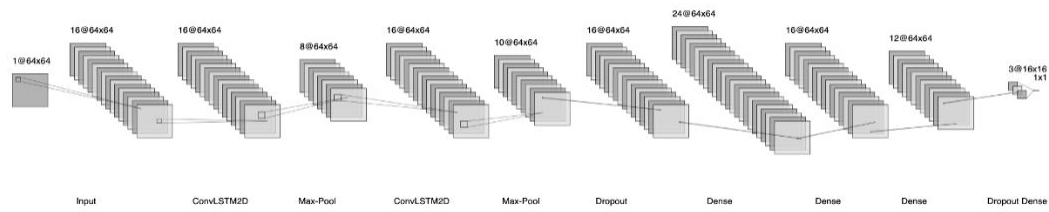


Figure 4: Convolutional Recurrent Neural Network Architecture

The `nal` class, called "TrainModel," enables the respiratory and cardiac auscultation machine learning models to be trained with datasets. The respiratory dataset is the ICBHI 2017 Respiratory Sound Database and Coswara dataset from GitHub. The cardiovascular dataset is from the "PhysioNet/Computing in Cardiology (CinC) Challenge 2016" [4]. The datasets are split according to this ratio 70:20:10 for training, validation, and testing. The SmartScope mobile application lets patients and doctors take auscultation readings of the heart and lungs. There are four main user interface screens in the app: the login page, the home page, the auscultation page, and the profile page. The auscultation page features a button "Record Auscultation" to record the auscultation reading. After the recording is finished, the auscultation results will be displayed in the app and the LCD touchscreen. The auscultation reading and patient data are stored in the SQLite database table, and the auscultation audio file is stored in the mobile application's document area. The SmartScope device and the app communicate via MQTT and TCP/IP protocols.

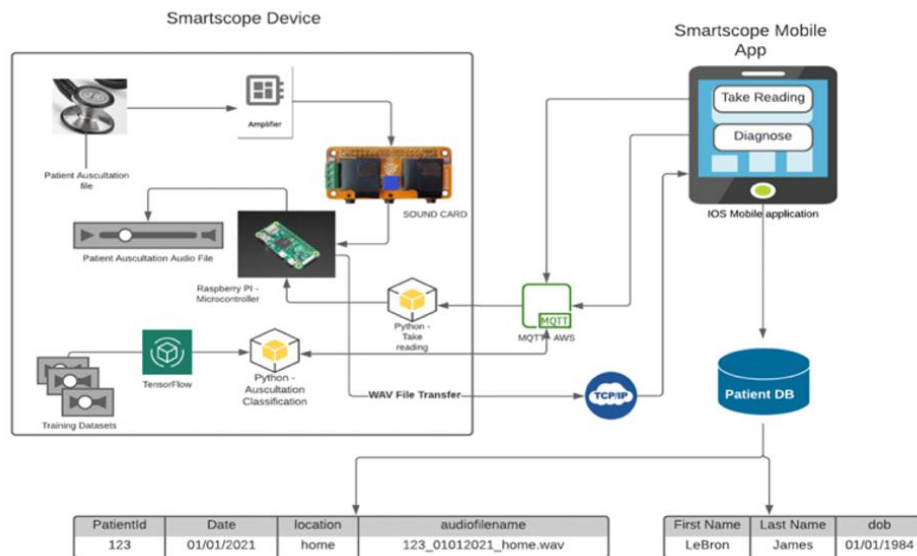


Figure 5: Overall Architecture

2.4 Reinforcement Learning Agent

To help patients select auscultation locations intelligently, the stethoscope employs the reinforcement learning agent: Deep Q Network. In this scenario, the environment is the state space of the auscultation points, and the set of actions is to auscultate another point or return the result. DQN approximates a Q-value function using a neural network whose weights are updated using the Bellman equation:

$$Q(s; a;) = r + \gamma \max_{a^0} Q(s^0; a^0; ; ^0) \quad (1)$$

The neural network consists of five layers, including three hidden dense layers of 256 units with the ReLU activation function. The final layer has 16 units: 12 units for each auscultation point and four units for each lung disease. For heart auscultation, the final layer contains seven units: five units for each auscultation point and two units representing murmurs and clicks. After the auscultation reading at a specific point, the feature vector representing the agent's state describes the characteristics of the audio. The feature vector for heart auscultation: average murmur and clicks probability, duration, and amplitude. Similarly, the feature vector for lung auscultation: average wheeze and crackle probability, duration, and amplitude. The Q-network accepts the feature vector as its input and outputs the Q-values for all possible actions. The action associated with the highest Q-values is chosen. Rewards are given when the predicted analysis of the disease is correct. To constrain the agent from using too many auscultation points, a small penalty is applied for each additional point.

2.5 Real-World Testing

The testing phase was split into two different parts: audio from the datasets and human participants. First, the machine learning algorithm was evaluated to classify cardiopulmonary diseases on test data. Since the test data was relatively similar to the training data, the device was also assessed on diverse and unobserved data from human participants. Most of the subjects were over the age of 30, without any preexisting heart or lung conditions. The machine learning algorithm's effectiveness is judged through the confusion matrix produced after the model has been evaluated on the testing data.

2.6 Data Analysis Methods

The machine learning algorithm's effectiveness is judged through the confusion matrix produced after the model has been evaluated on the testing data. The matrix shows how well the model did to classify respiratory diseases. Based on the results obtained, hyperparameter optimization took place to improve the accuracy of the neural network.

Recall measures how many classes the model has classified correctly. This value should be as high as possible. The formula for recall is below:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Precision measures the positive classes that the model have predicted compared to how many are actually positive. The formula for precision is below:

$$P \text{ precision} = \frac{T \text{ rue P ositive}}{T \text{ rue P ositives} + F \text{ alse P ositives}}$$

F-score measures both recall and precision at the same time. It uses Harmonic Mean instead of Arithmetic Mean to punish the extreme values more (Ghoneim, 2019). The machine learning model is evaluated by the F-Score primarily. The formula for F-Score is below.

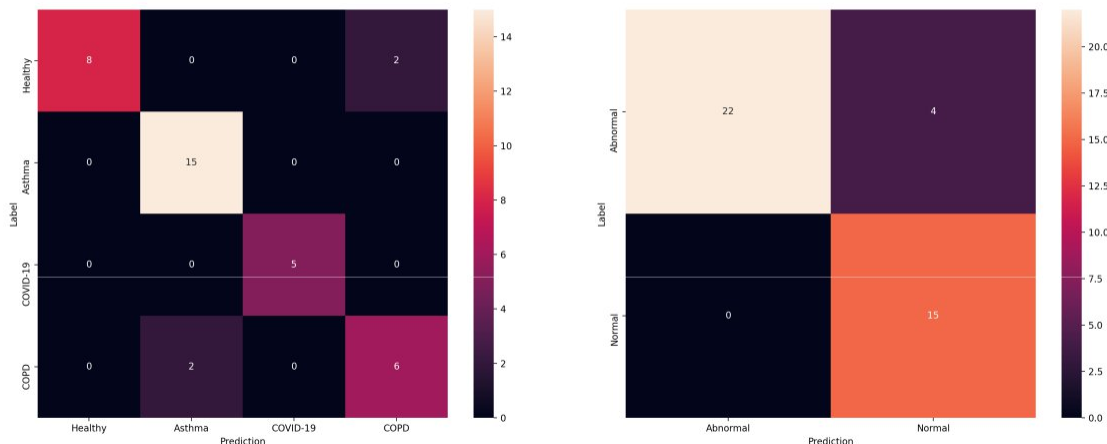
$$F \text{ Score} = \frac{2 \text{ Recall P recision}}{\text{Recall} + \text{P recision}}$$

3. Results

This project aimed to develop a digital auscultation device and a machine learning model to detect cardiopulmonary diseases with high accuracy. The prototype of SmartScope met engineering goals because it was designed to the correct speci cations, and the machine learning model acquired su cient accuracy to diagnose the diseases after it was fully optimized. Below are the classi cation results evaluated using metrics, like Precision, Recall, and F-Score. During the testing process, the machine learning algorithm’s ability to classify cardiopulmonary diseases on the training, validation, and testing data were evaluated.

3.1 CRNN Results

The machine learning model was trained on healthy, asthma, COPD, and COVID-19 audio les from various online datasets. The cardiovascular dataset was comprised of normal and abnormal sounds, which include heart murmurs.



(a) Respiratory Diseases

(b) Cardiovascular Diseases

Figure 6: Confusion Matrices

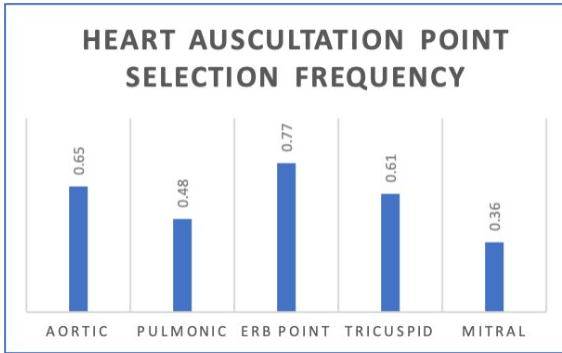
Class	Acc.	Precision	F1
Healthy	94.74%	0.80	0.89
Asthma	94.74%	1.0	0.94
COVID-19	100%	1.0	1.0
COPD	89.47%	0.75	0.75

Table 1: Respiratory Classification

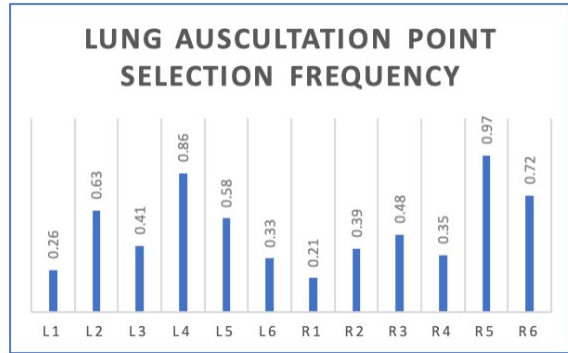
Class	Acc.	Precision	F1
Abnormal	90.24%	0.85	0.92
Healthy	90.24%	1.0	0.88

Table 2: Cardiovascular Classification

3.2 Reinforcement Learning Agent Results



(a)



(b)

Figure 7: Point Selection Frequency For Cardiopulmonary Auscultation

Technique	Acc.	F-Score	APs
Conventional	92.35%	0.93	5
Interactive	90.24%	0.90	1.6

Table 3: Cardiovascular Auscultation

Technique	Acc.	F-Score	APs
Conventional	95.67%	1.0	12
Interactive	93.33%	0.95	4.3

Table 4: Respiratory Auscultation

4. Discussion

As stated above, the machine learning model is primarily evaluated based on the F-Score metric since it is the most important metric to evaluate a neural network's effectiveness. As displayed in Table 1, both the respiratory classes obtained an average F-Score of 0.903. More importantly, the machine learning model correctly classified all of the COVID-19 audio files, providing patients a baseline test for COVID-19. As displayed in Table 2, the average F-Score for cardiovascular diseases was 0.90. Although it is slightly lower than those of respiratory diseases, the cardiovascular auscultation dataset was larger than the respiratory dataset, adding more variability. Since limitations exist with both datasets as they contain some low-quality auscultation sounds, acquiring higher-quality data will make the models even more accurate. For this research, it is hypothesized machine learning model will be more than 90% accurate in classifying various cardiopulmonary diseases. Since respiratory

and cardiovascular disease classification had an average testing accuracy and F-Score greater than 0.90, the results support the hypothesis.

The reinforcement learning agent, Deep Q Network, significantly reduced the auscultation time for both types. For cardiovascular disease classification, the number of points auscultated reduced by more than 3-fold while still maintaining an accuracy and a F-Score greater than 0.90. For respiratory disease classification, the number of points auscultated reduced by slightly less than 3-fold while still maintaining an accuracy and a F-Score greater than 0.90. Additionally, the agent preferred certain auscultation points compared to others. It favored the ERB point for cardiovascular disease classification and the L4 and R5 points for respiratory disease classification. These points were chosen with high frequency because they most likely provided high-quality and information-rich data. Using interactive auscultation not only reduces the number of points auscultated than conventional auscultation but does so without compromising accuracy.

5. Conclusion

Auscultation is a vital part of the patient examination for the early detection of heart and lung diseases. However, the subjectivity in identifying and analyzing auscultation is a challenge for physicians and can lead to misdiagnosis and inaccurate treatment. Since SmartScope classifies both respiratory and cardiovascular diseases like COPD, COVID-19 and Heart Murmurs with a combined accuracy greater than 90%, it is an effective solution to eliminate the subjectivity in diagnosis. The disease's diagnosis and progression can be readily transferred using SmartScope between doctors and specialists for faster evaluation and treatments. It also helps to avoid further evaluation with expensive procedures like echo, MRI, and CT scans. Additionally, the reinforcement learning agent, Deep Q Network, creates an interactive auscultation process where users can receive accurate results in the shortest period of time. Since SmartScope is a very affordable, user-friendly device, costing less than \$50, it can be used as a telemedicine solution for patients, especially during the COVID-19 pandemic, as people tended to avoid essential doctor appointments. It can also be beneficial in rural and low-income communities that lack both affordable and advanced technologies and easy access to health care. Future research for this device will include adding more diseases to classify, such as asthma and bronchitis. With all the features and applications, SmartScope will be revolutionary to the medical community as it provides accurate diagnosis instantaneously since time is a critical factor in saving lives.

References

- [1] Honorata Hafke-Dys, Anna Breborowicz, Pawel Kleka, Jędrzej Kocinski, and Adam Biniakowski. The accuracy of lung auscultation in the practice of physicians and medical students. PLOS ONE, 14(8), 2019.
- [2] Wen-Shuai Hu, Heng-Chao Li, Lei Pan, Wei Li, Ran Tao, and Qian Du. Feature extraction and classification based on spatial-spectral convlstm neural network for hyper-spectral images. CoRR, abs/1905.03577, 2019.

- [3] Tyler Lee and Frederic Theunissen. A single microphone noise reduction algorithm based on the detection and reconstruction of spectro-temporal features. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2184):20150309, 2015.
- [4] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, Jose Millet Roig, Ikaro Silva, and Alistair E Johnson. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37(12):2181{2213, 2016.
- [5] Niall McCarthy. These were the leading causes of death worldwide in 2019.
- [6] Bruno Machado Rocha, Diogo Pessoa, Alda Marques, Paulo Carvalho, and Rui Pedro Paiva. Automatic classification of adventitious respiratory sounds: A (un)solved problem? *Sensors*, 21(1):57, 2020.
- [7] David Springer, Lionel Tarassenko, and Gari Clifton. Logistic regression-hsmm-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, page 1{1, 2015.
- [8] Glaciele Nascimento Xavier, Antonio Carlos Duarte, Cesar Augusto Melo-Silva, Carlos Eduardo dos Santos, and Veronica Moreira Amado. Accuracy of pulmonary auscultation to detect abnormal respiratory mechanics: A cross-sectional diagnostic study. *Medical Hypotheses*, 83(6):733{734, 2014.