Bridging Auscultation and Tiny Machine Learning: A Digital Stethoscope Leveraging Convolutional Neural Networks on an Embedded Device for Organ Sound Analysis

Eray Mutlu , Valid Hüseyin , Görkem Serbes*

Department of Biomedical Engineering, Faculty of Electrical and Electronics Engineering, Yildiz Technical University, Istanbul 34220, Turkey

Corresponding Author Email: gserbes@yildiz.edu.tr

ABSTRACT

Traditional auscultation is used to determine certain pathological conditions related to internal organs utilizing cardiac, pulmonary, and intestinal sounds. However, this method relies heavily on the experience of the physician, which leads to non-repeatable subjective diagnosis. Automated analysis can be implemented by digitally recording organ sounds to address this limitation. The proposed system employs a convolutional neural network (CNN) model to determine the auscultated organ and subsequently applies digital filtering to the recorded raw signals based on the organ-specific frequency range. Additionally, the de-noised signals obtained can be transmitted to other smart devices via Bluetooth for further analysis. All the data acquisition, signal processing and learning steps were carried out in an embedded system, the Raspberry Pi 4 board. To achieve organ determination, the input of CNNs is obtained from the raw digital signals in the form of Mel-Spectrograms using the short time Fourier transform (STFT). The obtained time-frequency representations were fed into several pre-trained CNN architectures and compared in performance to a new CNN model derived from FISC-Net. The concept of tiny machine learning was employed in learning to enable real-time, low-power auscultation analysis on a portable and cost-efficient device, ensuring immediate feedback and enhanced patient privacy. The results showed that FISC-Netv1 surpassed other pretrained models by achieving a 90% accuracy rate demonstrating the effectiveness of the proposed system. Furthermore, the application of quantization awareness training reduced the learning model size by 4x without significantly compromising its performance.

1. INTRODUCTION

Conventional auscultation is the process of listening to the sounds originating from the organs of the body by means of an analog stethoscope [1]. The stethoscope is an acoustic medical device designed for the purpose of auscultation. It is typically utilized to listen to pulmonary and cardiac sounds, as well as to examine the intestines and blood flow in arteries and veins. Cardiac, pulmonary, and intestinal sounds serve to provide valuable information to medical practitioners, and thus assist in the diagnosis of various pathological conditions [2, 3].

The human body organs produce sounds at different intensities and frequency ranges. The heart mainly produces two sounds termed S1 and S2 in addition to S3 and S4 which tend to be less audible. The heart sounds occur within the frequency range 50-600 Hz [4]. They are produced by the opening and closing of the Atrioventricular and Semilunar valves which serve to regulate the direction of blood flow. The appearance of heart sound signals, visualized through Phonocardiograms, assist physicians in the diagnosis of certain diseases such as Aortic Stenosis, Mitral Stenosis, and Mitral Regurgitation [5]. The respiratory system produces three types of sounds: breath sound, voice sound, and adventitious sounds, each of which corresponds to a different frequency range. However, the overall frequency range in which pulmonary sound occurs is 50-2500 Hz [6]. Adventitious sounds, including crackles and wheezes, are considered abnormal and are a major indicator of pulmonary pathological conditions [7]. Intestinal sounds are produced by the movement of liquids and gases as a result of the contraction of muscles that comprise the walls of the gastrointestinal tract. These sounds occur in the range 100-200 Hz [8]. During the process of auscultation, medical practitioners rely on characteristics such as the duration, pitch, and frequency of bowel sounds in order to gather diagnostic information regarding the presence or absence of gastrointestinal diseases. Although there is no consensus on what constitutes abnormal bowel sounds, as a convention, the absence of bowel sound is considered to be abnormal [9].

The great interest in the evaluation of bodily sounds goes back to the times of Hippocrates and even to the seventeenth century B. C. in ancient Egypt. Back then, auscultation was done by placing one's ear on the patient's chest. This was the case till Laennec invented the stethoscope in 1817 which was quite different from the currently used conventional stethoscope. Laennec's stethoscope was a rather simple apparat consisting of two hollow tubes connected together. The currently used analog stethoscope is composed of a

binaural earpiece and a chest piece. The latter contains a diaphragm that is mostly used for cardiac auscultation and a bell that is mostly used for pulmonary auscultation [10]. Conventional analog stethoscopes attenuate sound signals at higher frequencies portraying lowpass filtering characteristics and thus causing loss of valuable information [11]. Additionally, such stethoscopes only allow for instant analysis of sound signals making the diagnosis process dependent on the experience of the medical practitioner [1]. Electronic stethoscopes surpass their analog counterparts in providing the ability to amplify and bandpass-filter the analog sound signals while converting them to digital ones. Digital stethoscopes, on the other hand, differ in the sense of being "smarter" providing visualization capabilities and the option of telecommunicating with other devices [12]. Additionally, they provide more complex signal processing techniques such as signal extraction and segmentation in addition to diagnosis algorithms [13].

There are many commercial digital stethoscopes that incorporate different mechanisms for noise cancellation in addition to the ability to select between different modes of frequency response. A good example is the Thinklabs One Digital Stethoscope which provides five different frequency ranges for filtering. It can also amplify organ sounds by a factor of 100 by virtue of capacitive transduction [14]. Similarly, 3M Littmann® Range can amplify organ sounds by a factor of 24 by virtue of adaptive noise cancellation [15]. Another example is the Welch-Allyn® Elite Electronic Stethoscope which provides a bell mode with a frequency range of 20 to 420 Hz and a diaphragm mode with frequency range of 350 to 1900 Hz [16]. The majority of commercial digital stethoscopes allow for the transmission of sound signals to other devices such as smartphones or personal computers thus allowing for further processing and analysis. The use of digital stethoscopes was reported to improve the diagnostic capabilities of auscultation compared to it done using analog stethoscopes [17, 18].

Machine learning techniques have recently been incorporated into applications related to embedded edge devices paving the way to the rise of the Internet of Things (IoT) technology. This concept is referred to as TinyML (Tiny Machine Learning). It unleashes the ability for implementing machine learning algorithms on hardware components with low processing power and memory. It also results in consuming less energy and preventing data security issues, thus averting the drawbacks of cloud-based computations [19]. Microcontroller units are a good example of hardware with low computational power that can be used for running machine learning algorithms. This is achieved by means of manual optimizations. Such optimizations are necessary since different hardware vendors have different architectures and use various frameworks limiting the portability of machine learning models. One way of tackling this problem is through TensorFlow Lite, a library that serves to deploy portable optimized machine learning models on edge devices [20].

The experience of the physician has a major effect on the success of traditional auscultation. This may result in non-reproducible diagnoses. Therefore, recording of organ sound signals for further analysis would be of major benefit. The system proposed in this article satisfies the aforementioned advantage. It is intended to record sound signals, classify them to the organ classes they correspond to, apply digital filtering, and transmit the signals to other devices. Thus, the system eliminates the need for servers by analyzing the signals on the edge device in real-time. Furthermore, the system will identify the auscultated organ by means of TinyML. Consequently, the system will allow for filtering sound signals according to frequency ranges that correspond to the targeted organ. Thus, the proposed system will be effective in noisy hospital environments.

This paper introduces a novel methodology that integrates traditional auscultation techniques with TinyML for analysis of organ sounds. The proposed convolutional neural network (CNN), named FISC-Netv1, demonstrates advancements by effectively classifying sounds originating from three distinct organs with exceptional precision and efficiency. Also, the proposed system stands out for its capability to classify sounds from three different organs, namely the heart, lungs, and bowel, while executing all processing tasks entirely on an embedded system, thus demonstrating a comprehensive integration of traditional auscultation methods with cutting-edge TinyML technology.

In this article, first a literature review on the use of machine learning based stethoscopes in the classification of physiological sound signals is given in Section 2. Then, an overview of the used materials and methods are included in Section 3. After that, the obtained results are shown in Section 4. Finally, conclusion and future studies are included in Section 5.

2. RELATED WORK

Numerous studies have investigated the use of machine learning techniques in digital stethoscopes for the purpose of diagnosing potential pathological conditions from bodily sound signals. Among these studies, some were focused on heart sounds. In a study by Fattah et al. [21], a smart phone based digital stethoscope was developed. The proposed system was intended to automatically analyze phonocardiograms, and to facilitate follow-ups by providing accessibility to doctors from remote areas. The design included a conventional chest piece with a piezoelectric sensor. Two probabilistic models, namely Markov model and k-nearest neighbor, were applied to classify phonocardiograms into healthy/unhealthy [21]. In a study by Chowdhury et al. [22], a system consisting of two units communicating through Bluetooth Low Energy (BLE) was developed. One unit is a digital stethoscope that performs sound acquisition, filtering, and transmission of signals to the other unit. The other is a personal computer that extracts features and classifies signals into normal/abnormal. It uses ensemble learning fed by Mel frequency cepstral coefficients (MFCC) in addition to time and frequency domain features [22]. The advantage of the system proposed in our study over the systems in the aforementioned studies is that it does not require an additional processing unit. That is, signal acquisition, filtering, and inference are performed on the same embedded system. In a study by Alqudah et al. [23], four different classes of heart disease were classified using a convolutional neural network (CNN) architecture, namely AOCTNet, fed by bispectrum images. The model in question achieved accuracy and F1-score of 99.47% and 99.48%, respectively. This result showcases the potential of CNNs for audio classification [23].

Similar to heart sound signals, respiratory sounds were investigated by many studies for their diagnostic potential. This was done by means of incorporating classification algorithms with feature extraction techniques. In a study carried out by Bahoura [24], pulmonary sounds were classified

into normal and wheezes. Various feature extraction techniques were used in combination with different machine learning methods. The best performance was obtained using MFCC with Gaussian Mixture Models (GMM), achieving an F1-score of 92.92% [24]. A snore detection system was developed in a study by Azarbarzin et al. [25] using K-means clustering. The algorithm was fed using a 2-D feature space obtained by applying principal component analysis (PCA) to 10-dimensional energy features. The algorithm achieved an accuracy of 98.2% and an F1-score of 96.67%. In a study by Sankur et al. [26], lung sounds were classified into healthy/pathological using Autoregressive (AR) models for feature extraction. Two classifiers, namely k-nearest neighbor (k-NN) with Itakura distance measure and quadratic classifier, were compared for different orders for both of them. K-NN performed better with an accuracy of 93.75% and an F1-score of 83.31%.

Gastrointestinal sound signals were also investigated for digital auscultation applications. In a study by Ficek et al. [27], a deep learning model that can classify sound into intestinal/non-intestinal was developed. Bowel sounds were collected using a contact piezoelectric microphone. Subjects were asked to follow a procedure before data acquisition like not eating a large meal after 18:00 or answering a clinical questionnaire. The model consisted of a combination of CNN and RNN (Recurrent Neural Network) also known as CRNN. It achieved an accuracy of 98.10% and an F1-score of 91.57% [27]. In a study conducted by Burne et al. [28], a method was developed to detect peristalsis from neonatal bowel sound signals. Bilinear feature fusion was used to incorporate 1-D and 2-D MFCC features in an ensemble of CNNs. The model yielded an accuracy of 95.1% and an F1-score of 97.3%.

Also, the classification of heart, bowel and lung sounds were achieved on an edge device in a previous research of ours [29]. In this study, three datasets were used namely ICBHI 2017 [30], Gastrointestinal Acoustic Activity [27] and PhysioNet/Computing in Cardiology (CinC) Challenge 2016 [31]. These datasets were picked for solving the new classification problem as classifying three organ sounds. Signals in these datasets are gathered and resampled to 41 kHz

and Mel-Spectroms were extracted to be used as features. Then, FISC-Net, a 2-D CNN was proposed for the classification task and it had achieved 99% accuracy at the Raspberry Pi 4 with a 0.05 seconds to predict.

3. MATERIAL AND METHODS

3.1 Materials

3.1.1 MAX9814 Electret microphone
The MAX9814 is a microphone amplifier that is both affordable and high-quality. Figure 1 shows a block diagram corresponding to its constituents whereas the physical appearance of the microphone is depicted in Figure 2. It contains circuitry for automatic gain control (AGC), a low-noise preamplifier with a gain of 12 dB. Additionally, it contains a variable gain amplifier (VGA) which can be adjusted to 20 dB or 0 dB. The Microphone also contains an output amplifier that can provide gains of 8 dB, 18 dB, or 28 dB. Consequently, the component can amplify sound with a gain between 40 dB and 60 dB. Trilevel digital input can be used for programming the gain of the output amplifier in addition to the ratio of attack-to-release time whereas the hold time of the AGC is fixed at 30 ms. MAX9814 comes in a compact 14-pin TDFN package and is rated for the extended temperature range of -40°C to +85°C [32].

3.1.2 Analog to digital convertor (MCP3008)
The MCP3008 ADC employed in this study is a 10-bit successive approximation ADC that includes a built-in sample-and-hold circuit. It can be programmed to function in differential or single ended mode providing four or eight inputs, respectively. The device communicates through a straightforward serial peripheral interface (SPI) protocol. The device can provide variable conversion rate that can reach up to 200 kilo samples per second (ksps). Furthermore, the MCP3008 operates effectively within a wide voltage range spanning from 2.7 V to 5.5 V [33]. The pinout of the ADC is shown in Figure 3.
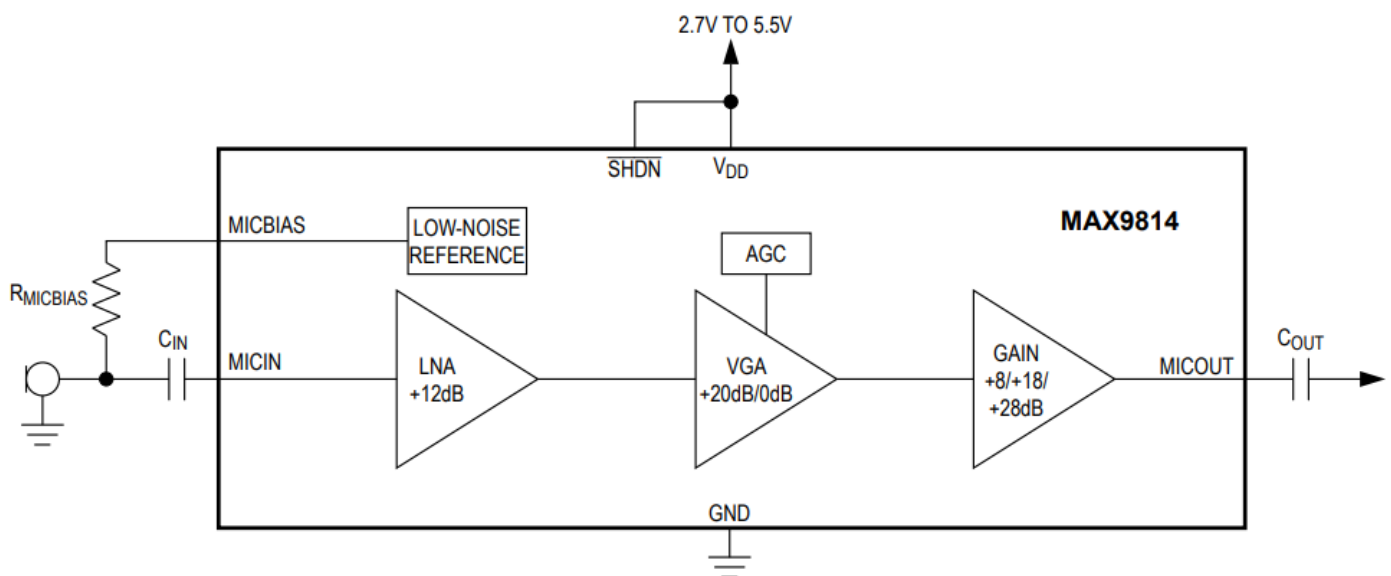

Figure 1. Simplified block diagram of the MAX9814 electret microphone [32]
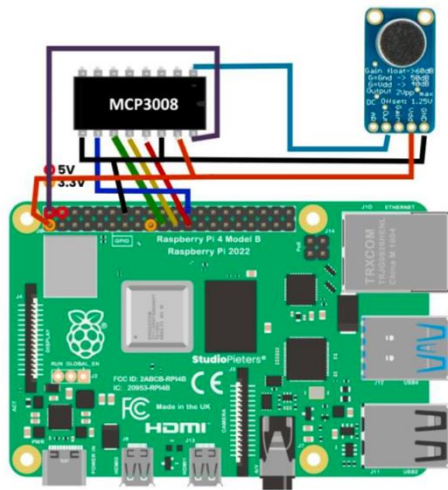
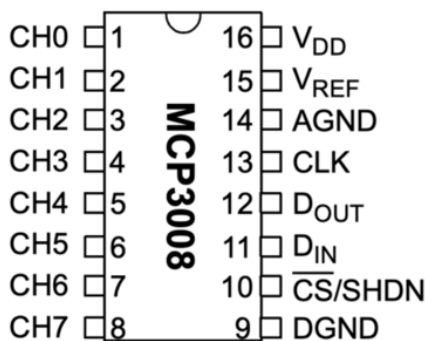**Figure 2.** Schematic of electronic components of the proposed system

Figure 3. MCP3008 pinout [34]

3.1.3 Raspberry Pi 4

The Raspberry Pi 4 board was connected to the ADC converter by virtue of the SPI 0 interface. As shown in Figure 2, SPI_MOSI, SPI_MISO, SPI_CLK, and SPI_CE0 were connected to MCP3008 pins 11, 12, 13, and 10, respectively. Furthermore, pins 9 and 15 of the ADC were connected to ground, whereas pins 15 and 16 were fed by 3.3 V and 5 V, respectively. The ADC was adjusted to receive input from the MAX9814 microphone via channel 0. The ADC was soldered to a prototyping plate that maintained firm connections with the raspberry pi via female header pins. This was done in order to ensure firm connections between components which is crucial for avoiding signal artifacts. Lastly, the physical appearance of the digital stethoscope is shown in Figure 4.

Figure 4. Digital stethoscope

3.1.4 Dataset

The collected dataset consists of three types of organ sounds which are heart, lungs, and intestines. It included 242 sound records pertaining to 22 adult and healthy subjects equally distributed among different genders. 11 records of analog signals (3 heart, 4 lungs, and 4 bowel) were obtained from each subject by virtue of the designed stethoscope. There was no exceptional procedure as in the bowel sounds research [27]. Upon aural inspection, it seemed that in some sound records cardiac, pulmonary, or intestinal sounds were either muffled, indistinguishable, or completely absent. The reason for this can be attributed to the limited experience of the staff conducting the recording process in addition to the mediocre quality of the hardware used. Therefore, only the samples in which clear cardiac, pulmonary, or intestinal sound is present were selected to be used for training the model. Consequently, the total number of records dropped to 146. For conducting this study which includes experimentation with subjects, an ethics allowance was acquired from the Academic Ethics Committee of Yildiz Technical University. The subjects have been informed about the objective of this study and their personal information will be kept anonymous. Information regarding the subjects included in the study is given in Table 1. Ranges for both age [35] and body-mass index (BMI) [36] were determined according to the literature.

To enhance the model's generalizability and robustness, various data augmentation techniques can be employed. These techniques included time stretching and adding background noise to augment the dataset, ensuring exposure to a wide range of acoustic variations commonly encountered in clinical settings. Additionally, variations in recording conditions, patient demographics such as age, sex and body-mass index (BMI) incorporated to enrich the dataset's diversity, thereby bolstering the CNN's ability to accurately classify organ sounds across diverse scenarios.

Table 1. Dataset description

Gender	Age	BMI
Female 10	Young-Adult 13	Healthy 13
Male 12	Middle-aged 7	Overweight 5
	Old Adults 2	Obesity 4

Young-Adult: 18-35, Middle-aged Adult: 36-55, Old Adults: 55+; Healthy: 18.5-25, Overweight: 25-30, Obesity: 30+

3.2 Methods

3.2.1 Organ sound acquisition via digital stethoscope

The applied approach is to place the microphone inside a ready-made stethoscope head. Electret microphones to be used are widely utilized in medical applications, such as fetal phonocardiography, due to their compact size, low power consumption and high efficiency [37]. The preferred model in this study is the Max9814 which has a wide frequency range (20 Hz – 20 kHz) and is, therefore, quite sufficient for this application.

3.2.2 Tiny machine learning

The preferred embedded system for this study was determined by considering factors such as processing power and sampling frequency. Additionally, the compatibility with the TensorFlow Lite library and the Python language were also considered. The latter supports a library widely used for music and audio analysis known as Librosa. Accordingly, Raspberry Pi 4 was chosen due to its high computing capabilities (Quad-

core Cortex A72 (ARM v8) 64-bit. SoC 1.5 GHz processor). The Raspberry Pi board does not contain analog input pins, so an additional integrated circuit is needed to perform Analog-to-Digital conversion.

Feature extraction. The feature set was extracted from the audio signals as a time-frequency representation known as Mel-Spectrograms. The latter term refers to a type of spectrogram that utilizes a nonlinear frequency scale consistent with human perception of sound. It is represented as Mel bands on the frequency axis and the energy within each band calculated using Fast Fourier Transform (FFT) or other spectral analysis techniques [38]. The Mel-Spectrograms were computed by first applying the short-time Fourier transform (STFT). Then, Mel filter-bank was applied to the absolute value of the obtained complex STFT coefficients. Thus, Mel-Spectrograms were calculated and saved as RGB images to be fed to the classification model. This was carried out by virtue of the Python Librosa library.

Classification model. Mobile-Net and Efficient-Net are well known pre-trained networks that are created exactly for edge device applications, and they are used for this reason in this study. The selection of pre-trained CNN architectures, including Mobile-Net and Efficient-Net, was based on their suitability for edge device applications. These models are recognized for their efficiency in terms of both computational requirements and memory usage. Also, FISC-Net was deployed as it was created with a relevant dataset and achieved an accuracy of 99%.

CNN architecture. The proposed 2-D CNN architecture called FISC-Netv1 is a conventional CNN architecture and derived from FISC-Net [29] with determined hyperparameters shown in Table 2. FISC-Net is a CNN that was designed for classification of auscultation sounds for edge device applications. Since it has achieved higher accuracy than Mobile-Net and Efficient-Net, FISC-Net was chosen for existing classification task. Then to improve the classification performance, existing architecture was altered for optimization. The new architecture of the proposed CNN is illustrated in Figure 5, whereas the parameters pertaining to the classification model are shown in Table 2.

Table 2. FISC-Netv1 architecture details and hyperparameters

Convolutional Layers	Pooling Layer	Dropout Layer
16 filters (3×3) stride (1×1)	Max-pooling (2×2)	0.5 rate
Output Layer	**Loss**	**Optimizer**
ReLU and Softmax	Categorical Cross-Entropy	Adam

One of the key reasons for the improved performance of FISC-Netv1 compared to other pre-trained CNN models in this specific application is its streamlined architecture. With only five convolutional layers, FISC-Netv1 strikes a balance between model complexity and efficiency, allowing it to efficiently capture and process relevant features for organ sound classification. The simplicity of the architecture ensures that the model can focus on extracting optimal feature maps specifically tailored to distinguish between different organ sounds. This focused approach not only reduces computational overhead but also mitigates the risk of overfitting, thereby enhancing the model's generalization capabilities.
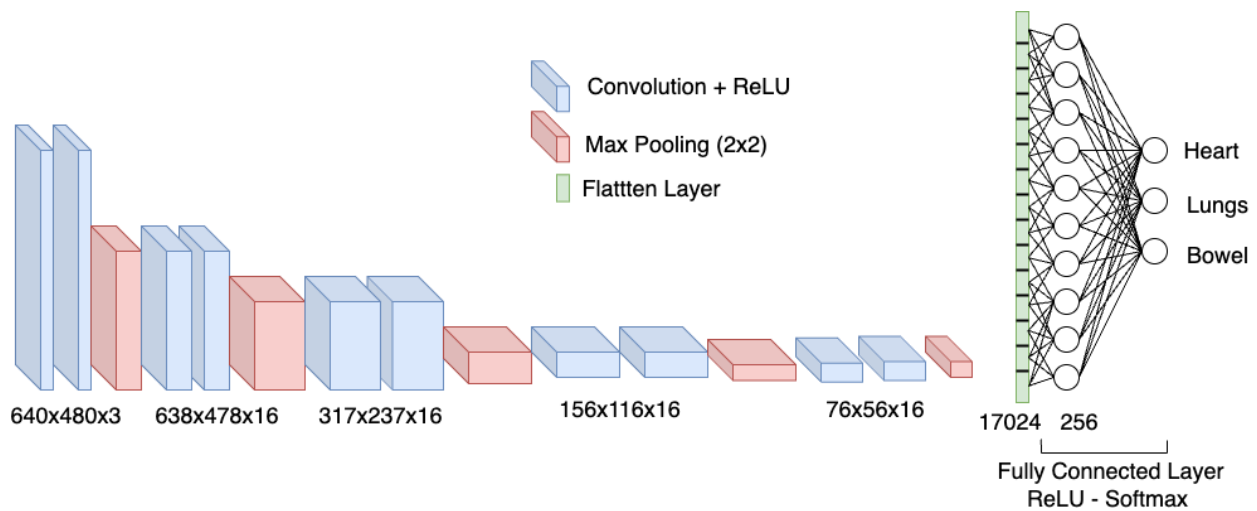

Figure 5. FISC-Netv1 architecture

Quantization aware training. For edge device applications, the size of the model and its inference time are key factors for deploying classification models. With quantization, both model size and inference time are reduced without any architectural change. Quantization can be done after the training phase, which means turning weights to 8-bit integers in common [39]. But there is another way to reduce the required memory size, that is quantization-aware training. In this method, weights of the model are quantized while the training process is continued so that layers become quantized too. This method is more prone to yield reduced accuracy though it shrinks the model by 4x. While quantization reduces

the model size, it introduces a slight trade-off in accuracy. The benefits of reduced model size and faster inference time far outweigh the minimal sacrifice in accuracy, making quantization-aware training a viable strategy for edge device deployment. To compare model size and accuracy results both methods were deployed using TensorFlow libraries.

3.2.3 Signal processing

Since acquired signals were noisy, a clear listening of auscultation sounds is needed for assessing organ conditions. Classified signals were filtered digitally according to the organ specific characteristics.
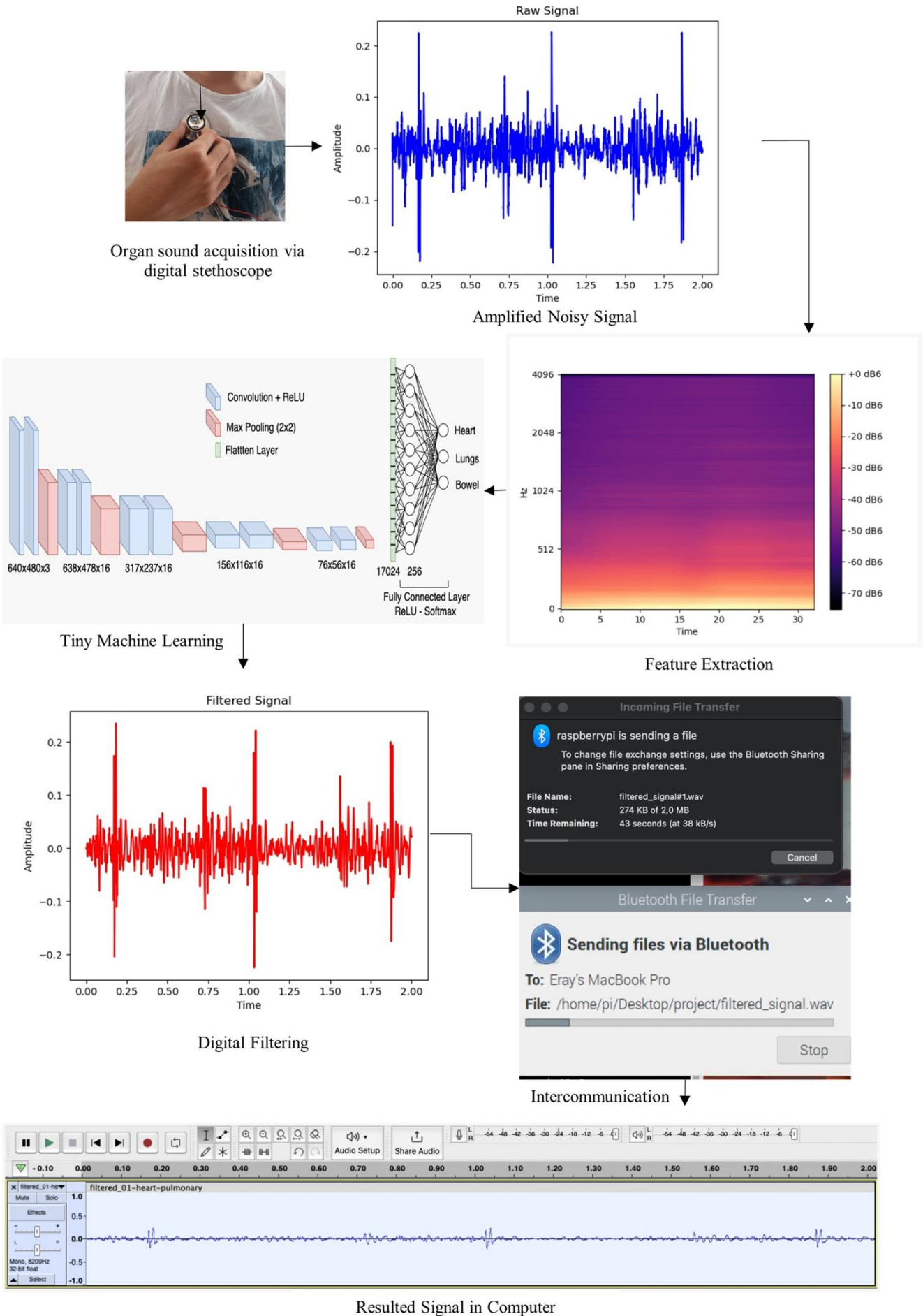
Figure 6. Flowchart of the proposed system

Heart sound filtering. Raw predicted heart sounds were 10th order Butterworth band-pass filtered to mitigate undesired high and low frequency components. Although the frequency range corresponding to heart sounds is from 50 to 600 Hz [4], the range was selected as 20-260 Hz according to frequency spectrums of acquired heart sound signals.

Lung sound filtering. Even though respiratory sounds are found between 50-2500 Hz [6], the acquired signals did not reach up to 600 Hz. Therefore, the picked range was between 20-500 Hz of 14th order Butterworth band-pass filter.

Bowel sound filtering. Most of the bowel sounds appear in the range of 100-200 Hz [8]. 6th order Butterworth band-pass filter with the range of 20-150 Hz was used to reduce noise in bowel sound signals.

3.2.4 Intercommunication between devices

There are two common options for intercommunication between devices as Bluetooth and internet. Bluetooth is an approach for short-scope and wireless data communication. The main reasons for this are low energy consumption, low cost, and safety issues [40]. Thanks to the Bluetooth module in the selected embedded system, it will be able to provide data communication between the patient and the doctor. In order to meet the communication requirement, organ-specific clean sounds obtained will be made ready for the Bluetooth use on the Bluetooth module in Raspberry Pi 4.

Another option is using Virtual Network Computing (VNC) technology embedded in Raspberry Pi 4. With this technology, accessing and controlling a device environment over an internet connection is possible. So, to compare the speed of file transmission between devices, VNC servers were used.

The proposed design can be summarized as follows. First, sound is captured by the stethoscope and digitized for further processing. Then, Mel-Spectrogram is extracted as features and fed to the classification model. After the classification, the raw signal is filtered according to the classified organ with the aim of obtaining clean signals without noise components. Then, filtered signal is transmitted to the computer both by internet and Bluetooth. The flowchart of the design is illustrated in Figure 6.

4. RESULTS

In the results section, feature extraction, classification, filtering, and intercommunication results are given.

4.1 Feature extraction

It is essential to carry out a feature extraction stage while preparing input features for our model. The Mel-Spectrogram is the method of choice in this situation for extracting relevant time-frequency features from the sounds of each organ. Mel-Spectrogram was chosen because it is well known for being a fundamental and popular feature extraction method for acoustic signal classification, notably for auscultation sounds. This preference is attributed to its efficacy in capturing and representing the intrinsic spectral content of acoustic signals, which is vital for discriminating subtle nuances between different organ sounds. The parameters of Mel-Spectrogram were determined by calculating the average difference between extracted Mel-Spectrograms of heart and lung sounds as the distinction between them is more difficult. Among the

various parameters, only the type of windowing function was altered, while other parameters, such as the number of FFT points and hop length remain unchanged. The aim of this method was to determine which windowing function performed best in terms of distinction between heart and lung sounds. The unchanged parameters are shown in Table 3 and the independent variable with the related results are shown in Table 4.

Figure 7 illustrates examples of the resulting Mel-Spectrograms generated using the selected Tukey window. This windowing technique was chosen based on the observation of the highest difference between spectrograms, as indicated in Table 4.

Table 3. Unchanged parameters of Mel-Spectrogram

Number of FFT	32800 (sampling rate × 4 seconds)
Hop Length	64

Table 4. Different windowing in Mel-Spectrogram

Window Type	Difference
Tukey	25×10^6
Cosine	19×10^6
Taylor	14×10^6
Lanczos	13×10^6

4.2 Classification performance

By means of the trained models, sound signals were classified into three classes representing the organs heart, lung, and bowel. A standard train-test split technique was used to train all models, with 80% of the data used for training and 20% used for testing.

As shown in Table 5, the models performed well in the training phase whereas only FISC-Net performed well in the testing phase achieving 88% accuracy. According to these results, FISC-Net was chosen to move further with. By adjusting parameters of FISC-Net, higher accuracy was achieved with both quantization methods as shown in Table 6 and Table 7. Confusion matrices of two quantization results are also depicted in Figure 8 and Figure 9.

As shown in Figure 8 and Figure 9, the implementation of quantization-aware training resulted in a notable enhancement in the classification performance of bowel and lung sounds. However, it is worth noting that while the F1-scores for bowel and lung sounds increased, there was a minor decrease in the performance of heart sound prediction.

Table 5. Performance of pre-trained networks

Models	Train Accuracy	Test Accuracy
Mobile-Net	96.7%	66.6%
Efficient-Net	90.3%	66.6%
FISC-Net	96%	88%

Table 6. Classification results for FISC-Netv1 without quantization-aware training

	Precision	Recall	F1 Score	Accuracy
Heart	1	0.87	0.93	100%
Lung	0.88	0.77	0.82	88%
Bowel	0.79	1	0.88	79%
Average	0.89	0.88	0.88	88%

While there are differences for predicting different organs in two methods of quantization, in total, F1-score did not change and the size of the model was decreased to 4.4 MB from 17.5 MB, which indicates a lower computational cost.

Table 7. Classification results for FISC-Netv1 with quantization-aware training

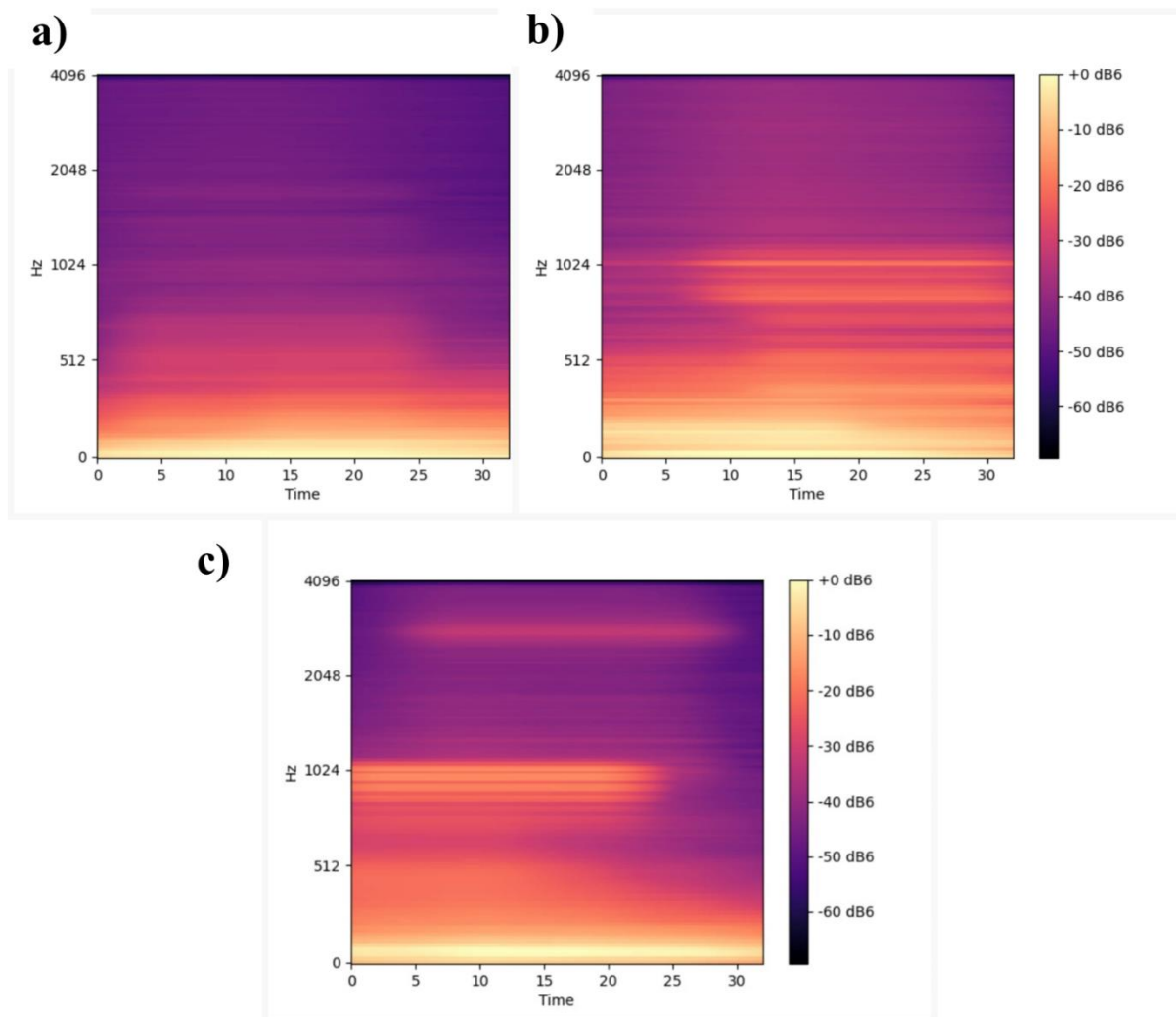|	Precision	Recall	F1 Score	Accuracy
Heart	0.96	0.83	0.89	96%
Lung	0.87	0.87	0.87	87%
Bowel	0.88	1	0.94	88%
Average	0.9	0.9	0.9	90%

Figure 7. Mel-Spectrograms of heart (a), lungs (b) and bowel (c)
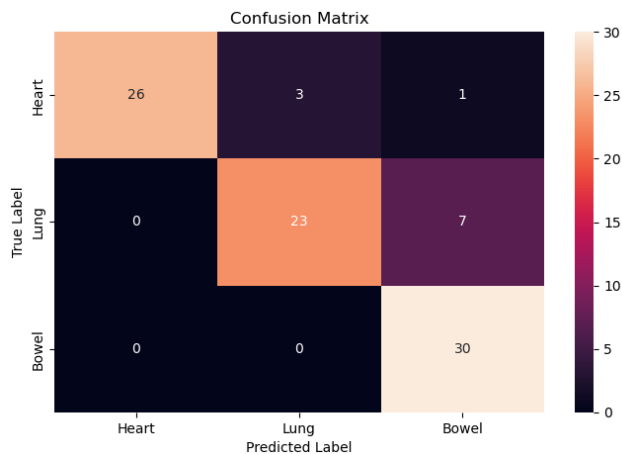

Figure 8. Confusion matrix of FISC-Netv1 without quantization-aware training
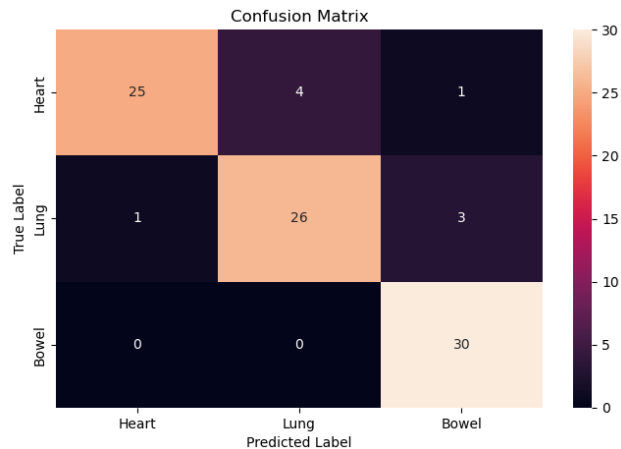

Figure 9. Confusion matrix of FISC-Netv1 with quantization-aware training

4.3 Filtered signals

There are two reasons for filtering the raw signal. First, filtering helps isolate and highlight the distinctive traits that are unique to each organ. The specific characteristics that can help with accurate analysis by carefully keeping frequency components relevant to the targeted organ sounds. Second, filtering is essential for reducing the impact of noise and undesirable artifacts that could exist in the raw data. Reduction of the effect of noise by digital filtering, raising the overall quality of the acoustic data in the procedure.

Organ-specific filtering was done according to predicted class. Examples of filtered signals are given in Figures 10, 11, and 12. Although there is no clear distinction between signal and noise in example plots, differences between raw and filtered signals were obvious in practical listening.
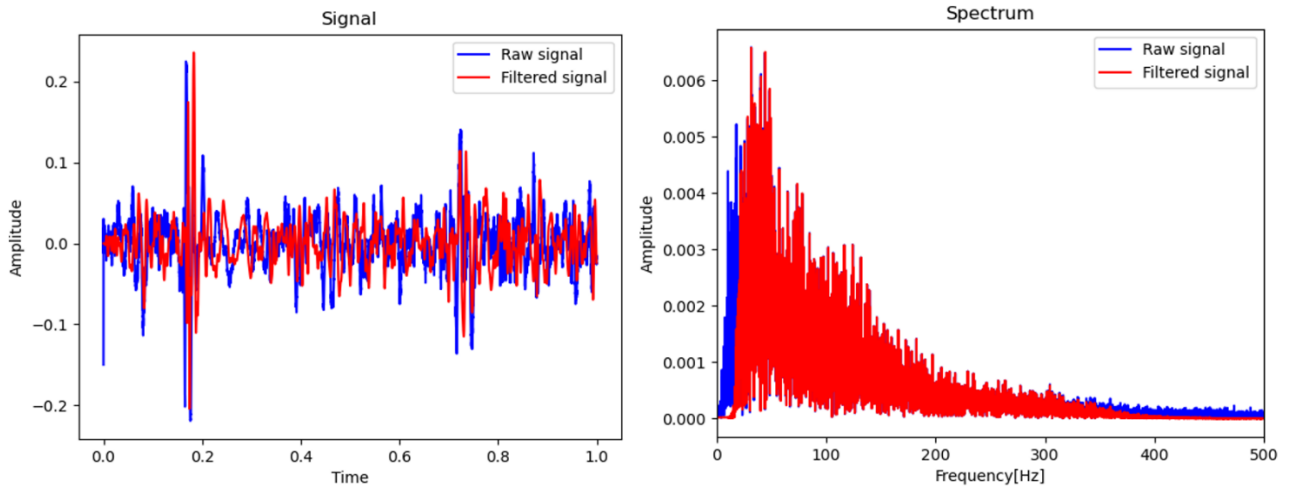

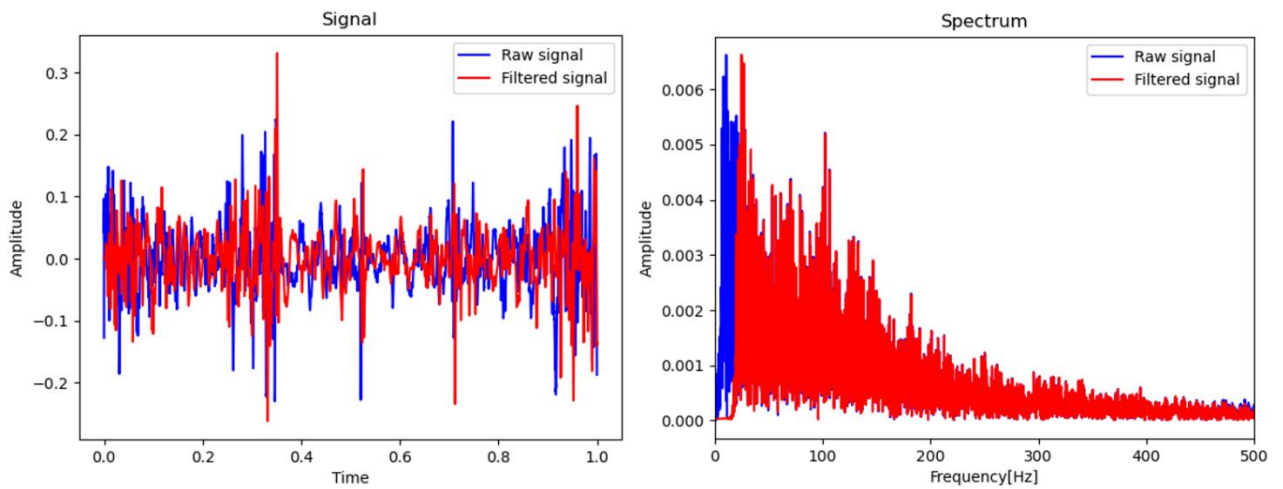**Figure 10.** Band-pass filtered heart sound & frequency spectrum

Figure 11. Band-pass filtered lung sound & frequency spectrum
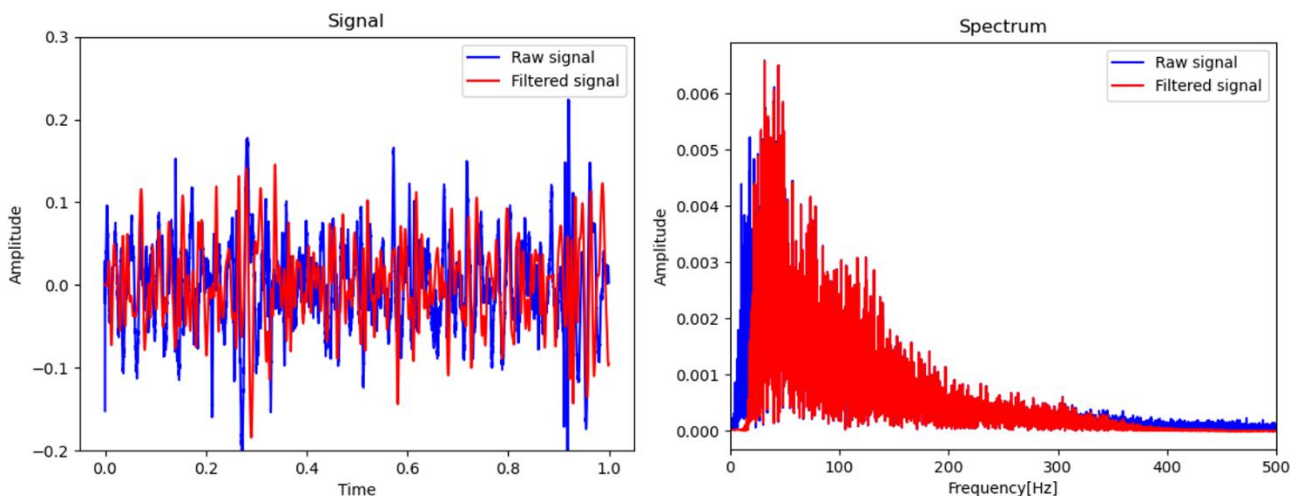

Figure 12. Band-pass filtered bowel sound & frequency spectrum

4.4 Communication

Intercommunication is essential to the proposed concept since it enables the transfer of filtered auscultation signals to a platform or device intended for further investigation. This approach combines both auditory and visual modes of analysis. Additionally, it allows for efficient data storage and retrieval for archival and reference purposes as well as real-time analysis in medical examinations and remote monitoring settings.

In the Raspberry Pi 4, Bluetooth communication was supplied by Blueman and Bluez packages. By using GUI, filtered signals can be transferred via Bluetooth to any device that has a Bluetooth module.

For comparison, VNC technology was used for intercommunication between digital stethoscope and computer. VNC is embedded in Raspberry Pi 4 and graphical user interface (GUI) was serviced the file transmission.

Time from the data acquisition to transmission of clean signal to the computer was calculated and shown in Table 8.

Table 8. Required time for processes in the design

Communication Method	Time (in Seconds)
Bluetooth	91
VNC	24

VNC has higher speed due to the nature of internet and Bluetooth. Therefore, if communication speed was the primary concern, VNC would be picked as an intercommunication method between devices.

5. DISCUSSION AND CONCLUSION

Existing literature was able to classify pathologies in heart and lungs separately or detect gastrointestinal sounds with a program that is a must to follow. Compared to the results in Table 9, the proposed model performed relatively well considering the limitations in the design. In this study, CNN was used as a classification model, being the most successful method in the literature related to the classification of auscultation signals. Similarly, feature extraction via Mel-frequency cepstral coefficients (MFCCs) yielded the best results in the related literature. However, feature extraction via Mel-Spectrogram was preferred for this study. The reason is that although being quite similar to MFCCs, Mel-Spectrograms require fewer processing steps which makes them more appropriate for embedded applications. In addition, the used microphone was of the electret type as it provided good quality relative to its price.

While previous studies were aimed at the classification of pathologies, the proposed design differs in the objective of classifying the source organ of sound. This will allow for the system to identify the targeted range of frequency. The proposed system, thus, performs digital filtering specific to the auscultated organs and transmits the filtered signals to other smart devices for further examination. This is expected to contribute to the diagnostic potential of the obtained signals and reduce related errors. Also, the proposed system has the ability to classify three different organs in the same system and all data acquisition is supposed to happen in daily-life conditions. Another novel aspect of the proposed system is running machine learning algorithms on an edge device. This feature makes the digital stethoscope more independent, safe, and power-efficient. As a result, the system proposed in this study allows healthcare professionals to further listen to, and process, sound records, thus yielding more robust and objective diagnoses. In addition, such a device can be incorporated into clinical workflows. It can be used in cases of pandemic outbreaks to limit the contact between medical practitioners and patients. For instance, instead of a single doctor having to auscultate a single patient, a nurse wearing protective shields can record bodily sound from multiple patients. The obtained records can be broadcasted to multiple devices and assessed later by doctors from different departments (e.g., emergency and intensive care unit). Thus, it assists in maintaining the safety of medical doctors and controlling the outbreak of the disease. This approach can be particularly relevant in the last couple of years due to the technical challenges that were brought about by the Covid-19 outbreak. An example of such challenges includes the inconvenience of using conventional stethoscopes with protective clothing on. In addition, it can be utilized by non-expert health-care professionals such as emergency crews who can use it in the field in remote locations without internet access [41, 42]. Also, the proposed project could serve telemedicine technology as a continuation of the teleradiology system. By integrating Internet of Things (IoT) into the system, patients can be followed by the physician. While telemedicine increases the service to be provided by the health sector, it can also be used to save time and cost. By eliminating the necessity of going to metropoles for diagnosis and treatment, delivering many health services to people living in areas far from the town would be possible via telemedicine technology. Regular and continuous follow-up of patients provides taking precautions without the need for emergency interventions. Also, maintaining the treatment in a planned manner is better for the patient, as well as reducing treatment and care costs. At the same time, it will be possible to use it in the follow-up after the operation and/or treatment. For example, heart sound monitoring after heart valve insertion can show the physician whether there is an anomaly in the patient's heart function [43]. Furthermore, the system can be used for educational purposes serving as means for medical students and trainees to listen to abnormal bodily sounds that can be associated with different medical conditions.

Despite how useful the proposed system is, certain limitations might hinder its utility. Since it imposes a paradigm shift in how clinical trials are conducted, it can require intensive training that might be costly to some healthcare institutions. Also, technical aspects such as the transfer and storage of the obtained signals may not be compatible with the present clinical data management systems. In addition, as is the case with other types of patient data, sound signals might be susceptible to malicious attacks and misuse. Therefore, some institutions that are not proficient in protecting their patient data might be reluctant to incorporate such systems into their clinical workflow. Sound acquisition is one of the main parts of this study. As the aim of this study is creating an end-user friendly product, collecting signals in a daily life environment is essential. Due to this reason, noise coming from many sources, such as talking, should be reduced. The effect of such noise is more prominent for bowel sound than it is for the other sound signals, due to their relatively low amplitudes. The comparatively low accuracy for the Bowel class in Table 6 can be attributed to this effect. Also, recall pertaining to the Heart class is low compared to accuracy

associated with the same class, as evident in Table 7. This indicates a high false positive rate caused by non-heart samples being classified as heart. Introducing the incoming signals to analog filtering can help in resolving the shortcoming of the model in certain classes. Another limitation is a communication problem, it stems from the fact that Bluetooth technology works with close distances. In addition, internet connection requires access to Wi-Fi and consumes high power.

As the current design does not make pathological detection, working with patient data that can be obtained following the similar principles can improve the detection of pathological conditions such as interstitial lung diseases, asthma, COPD and heart failure. After that, different designs and communication with the internet instead of Bluetooth can be considered in order to disseminate and facilitate the procedure like point-of-care designs. Another point of advancement is relating to additive manufacturing. Stethoscopes can be designed and printed via 3-D printing which makes the design cheaper. Moreover, the entire system should be designed or manufactured to transform to a product from scientific research.

Bodily sound signals have different characteristics. Namely, heart sounds being more transient whereas lung sounds have more eminent oscillatory behavior. While spectrograms are well-suited for oscillatory signals, scalograms are suited for both types of characteristics [44]. For this reason, and in order to enhance the performance of the current model, Wavelet Transform can be considered in future studies as a feature extraction technique while taking computational complexity into consideration. Also, in order to yield a more robust model, data augmentation methods can also be considered. Basic time-domain augmentation techniques such as Time Stretching, Noise Addition, and Pitch Shifting, seem to offer laudable results [45]. In addition, Horizontal Flipping, Principal Component Analysis (PCA) Color Augmentation, Saturation-Value (SV) Perturbations, and Spectrum Correction are among the most promising techniques when augmenting data in the form of Mel-Spectrogram as seen in related studies [46, 47].

Table 9. Summary of related work

Reference	Analysed Sound	Classes	Sensor Type	# of Subjects	Feature Extraction	Classification Method	Performance
[21]	Heart sounds	Healthy & Unhealthy	Piezo disc vibration sensor	100	Homomorphic Envelogram, Hilbert Envelope, Discrete Wavelet Transform Envelope, Power Spectral Density Envelope. PCG signal segmentation with Hidden Semi Markov Model and annotated as S1, Systole, S2 and Diastole.	k-NN	Accuracy = 82% F1-score = 83.57%
[22]	Heart sounds	Normal & Abnormal	Custom made sensor	3126 sample	t-domain, f-domain, and Mel frequency cepstral coefficients (MFCC).	Ensemble Learning	Accuracy = 97% and 88%
[23]	Heart sounds	Normal & Abnormal	Electret microphone	1000 sample	High order spectral features.	Bispectrum images with CNN	Accuracy = 99.47% F1-score = 99.48%
[24]	Lung sounds	normal & wheeze sounds	Electret microphone	24	Fourier Transform (STFT), Linear Predictive Coding, Wavelet Transform, Cepstral Analysis (MFCC).	Vector Quantization, GMM, and ANN	F1-score = 92.92%
[25]	Lung sounds	Snore & No-snore	2 ECM-77B Electret microphone	20	0-5000 Hz frequency range-500Hz subbands. Short-time Fourier transform (STFT) of each episode was calculated using 50ms windows with 50% overlap.	k-means clustering	Accuracy = 98.2% F1-score = 96.67%
[26]	Lung sounds	Healthy & Pathological	Electret microphone (ECM44, Sony)	69	Autoregressive Model.	k-NN classifier	Accuracy = 93.75% F1-score = 83.31%
[27]	Bowel sounds	Intestinal & Non-intestinal Sound	contact microphone	19	Spectrogram { frame length = 441 frame step = 110 freq. range = 0-1500 Hz Smoothing = Hann window}.	CRNN	Accuracy = 98.10% F1-score = 91.57%
[28]	Bowel sounds	Peristaltic & Non-Peristaltic Sound	--	49	1-D and 2-D MFCC features.	CNN	Accuracy = 95.1% F1-score = 97.3%

Furthermore, running the same or a similar model on edge devices other than Raspberry Pi 4 can also be considered in future studies. By doing so, notable performance can be acquired using devices with lower computational power. Also, to make our design more extensive and less power-consuming, other means of communication, such as the Long Range Wide Area Network (LoRaWAN), can be considered. Additionally, the performance of the model can be enhanced further by including a greater amount of data. Moreover, including real-time disease diagnosis in the proposed system can be considered in further research.

ACKNOWLEDGMENT

REFERENCES

[1] Leng, S., Tan, R.S., Chai, K.T.C., Wang, C., Ghista, D., Zhong, L. (2015). The electronic stethoscope. Biomedical Engineering Online, 14(1): 1-37. https://doi.org/10.1186/s12938-015-0056-y

[2] Sarkar, M., Madabhavi, I., Niranjan, N., Dogra, M. (2015). Auscultation of the respiratory system. Annals of Thoracic Medicine, 10(3): 158-168. https://doi.org/10.4103/1817-1737.160831

[3] Felder, S., Margel, D., Murrell, Z., Fleshner, P. (2014). Usefulness of bowel sound auscultation: A prospective evaluation. Journal of Surgical Education, 71(5): 768-773. https://doi.org/10.1016/j.jsurg.2014.02.003

[4] Coast, D.A., Cano, G.G., Briller, S.A. (1990). Use of hidden markov models for electrocardiographic signal analysis. Journal of Electrocardiology, 23: 184-191. https://doi.org/10.1016/0022-0736(90)90099-N

[5] Hall, J.E., Hall M.E. (2020). Guyton and Hall Textbook of Medical Physiology E-Book. Elsevier Health Sciences, pp. 283-286.

[6] Sovijarvi, A.R.A., Vanderschoot, J., Earis, J.E. (2000). Standardization of computerized respiratory sound analysis. European Respiratory Review, 10(77): 585-585.

[7] Bohadana, A., Izbicki, G., Kraman, S.S. (2014). Fundamentals of lung auscultation. New England Journal of Medicine, 370(8): 744-751. https://doi.org/10.1056/NEJMra1302901

[8] Ulusar, U.D., Canpolat, M., Yaprak, M., Kazanir, S., Ogunc, G. (2013). Real-time monitoring for recovery of gastrointestinal tract motility detection after abdominal surgery. In 2013 7th International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, pp. 1-4. https://doi.org/10.1109/ICAICT.2013.6722654

[9] Baid, H. (2009). A critical review of auscultating bowel sounds. British Journal of Nursing, 18(18): 1125-1129. https://doi.org/10.12968/bjon.2009.18.18.44555

[10] Ferns, T. (2007). Respiratory auscultation: How to use a stethoscope. Nursing Times, 103(24): 28.

[11] Grenier, M.C., Gagnon, K., Genest, J., Durand, J., Durand, L.G. (1998). Clinical comparison of acoustic and electronic stethoscopes and design of a new electronic stethoscope. American Journal of Cardiology, 81(5): 653-656. https://doi.org/10.1016/S0002-9149(97)00977-6

[12] Seah, J.J., Zhao, J., Wang, D.Y., Lee, H.P. (2023). Review on the advancements of stethoscope types in chest auscultation. Diagnostics, 13(9): 1545. https://doi.org/10.3390/diagnostics13091545

[13] Myint, W.W., Dillard, B. (2001). An electronic stethoscope with diagnosis capability. In Proceedings of the 33rd Southeastern Symposium on System Theory (Cat. No. 01EX460), Athens, OH, USA, pp. 133-137. https://doi.org/10.1109/SSST.2001.918505

[14] Thinklabs Medical LLC. Thinklabs One Digital Stethoscope. http://www.thinklabs.com/, accessed on Aug. 11, 2023.

[15] 3M Littmann. 3M Littmann Range. http://www.littmann.com/3M/en_US/littmann-stethoscopes/, accessed on Aug. 11, 2023.

[16] Welch Allyn. Welch-Allyn Elite Electronic Stethoscope. http://www.welchallyn.com/content/welchallyn/americas/en/products/categories/discontinued-products/physical-exam/stethoscopes/elite_electronic_stethoscope/, accessed on Aug. 11, 2023.

[17] Kevat, A.C., Kalirajah, A., Roseby, R. (2017). Digital stethoscopes compared to standard auscultation for detecting abnormal paediatric breath sounds. European Journal of Pediatrics, 176: 989-992. https://doi.org/10.1007/s00431-017-2929-5

[18] Silverman, B., Balk, M. (2019). Digital stethoscope—improved auscultation at the bedside. The American Journal of Cardiology, 123(6): 984-985. https://doi.org/10.1016/j.amjcard.2018.12.022

[19] Ray, P.P. (2022). A review on TinyML: State-of-the-art and prospects. Journal of King Saud University-Computer and Information Sciences, 34(4): 1595-1623. https://doi.org/10.1016/j.jksuci.2021.11.019

[20] David, R., Duke, J., Jain, A., Janapa Reddi, V., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Wang, T., Warden, P., Rhodes, R. (2021). Tensorflow lite micro: Embedded machine learning for TinyML systems. In Proceedings of the 4th MLSys Conference, San Jose, CA, USA, pp. 800-811. https://proceedings.mlsys.org/paper_files/paper/2021/file/6c44dc73014d66ba49b28d483a8f8b0d-Paper.pdf.

[21] Fattah, S.A., Rahman, N.M., Maksud, A., Foysal, S.I., Chowdhury, R.I., Chowdhury, S.S., Shahanaz, C. (2017). Stetho-phone: Low-cost digital stethoscope for remote personalized healthcare. In 2017 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, USA, pp. 1-7. https://doi.org/10.1109/GHTC.2017.8239325

[22] Chowdhury, M.E.H., Khandakar, A., Alzoubi, K., Mansoor, S., Tahir, A.M., Reaz, M.B.I., Al-Emadi, N. (2019). Real-time smart-digital stethoscope system for heart diseases monitoring. Sensors, 19(12): 2781. https://doi.org/10.3390/s19122781

[23] Alqudah, A.M., Alquran, H., Qasmieh, I.A. (2020). Classification of heart sound short records using bispectrum analysis approach images and deep learning. Network Modeling Analysis in Health Informatics and Bioinformatics, 9: 1-16. https://doi.org/10.1007/s13721-020-00272-5

[24] Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and

wheeze classes. Computers in Biology and Medicine, 39(9): 824-843. https://doi.org/10.1016/j.compbiomed.2009.06.011

[25] Azarbarzin, A., Moussavi, Z. (2010). Unsupervised classification of respiratory sound signal into snore/no-snore classes. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, pp. 3666-3669. https://doi.org/10.1109/IEMBS.2010.5627650

[26] Sankur, B., Kahya, Y.P., Güler, E.Ç., Engin, T. (1994). Comparison of AR-based algorithms for respiratory sounds classification. Computers in Biology and Medicine, 24(1): 67-76. https://doi.org/10.1016/0010-4825(94)90038-8

[27] Ficek, J., Radzikowski, K., Nowak, J.K., Yoshie, O., Walkowiak, J., Nowak, R. (2021). Analysis of gastrointestinal acoustic activity using deep neural networks. Sensors, 21(22): 7602. https://doi.org/10.3390/s21227602

[28] Burne, L., Sitaula, C., Priyadarshi, A., Tracy, M., Kavehei, O., Hinder, M., Withana, A., McEwan, A., Marzbanrad, F. (2022). Ensemble approach on deep and handcrafted features for neonatal bowel sound detection. IEEE Journal of Biomedical and Health Informatics, 27(6): 2603-2613. https://doi.org/10.1109/JBHI.2022.3217559

[29] Mutlu, E., Serbes, G. (2023). FISC-Net: Classification of auscultation sounds on an edge device. In 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, pp. 174-180. https://doi.org/10.1109/IMSA58542.2023.10217510

[30] Rocha, B.M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E., Kaimakamis, E., Natsiavas, P., Oliveira, A., Jácome, C., Marques, A., Paiva, R.P., Chouvarda, I., Carvalho, P., Maglaveras, N. (2018). A respiratory sound database for the development of automated classification. Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017. Springer, Singapore, pp. 33-37. https://doi.org/10.1007/978-981-10-7419-6_6

[31] Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E.W., Syed, Z., Schmidt, S.E., Papadaniil, C.D., Hadjileontiadis, L., Naseri, H., Moukadem, A., Dieterlen, A., Brandt, C., Tang, H., Samieinasab, M., Samieinasab, M.R., Sameni, R., Mark, R.G., Clifford, G.D. (2016). An open access database for the evaluation of heart sound algorithms. Physiological Measurement, 37(12): 2181. https://doi.org/10.1088/0967-3334/37/12/2181

[32] Analog Devices. MAX9814 Microphone Amplifier with AGC and Low-Noise Microphone Bias. http://www.analog.com/media/en/technical-documentation/data-sheets/MAX9814.pdf/, accessed on Aug. 20, 2023.

[33] Gupta, A., Jain, R., Joshi, R., Saxena, R. (2017). Real time remote solar monitoring system. In 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA)(Fall), Dehradun, India, pp. 1-5. https://doi.org/10.1109/ICACCAF.2017.8344723

[34] Microchip Technology Inc. MCP3004/3008 2.7V 4-Channel/8-Channel 10-Bit A/D Converters with SPI Serial Interface.

http://datasheet.octopart.com/MCP3008-I/P-Microchip-datasheet-8326659.pdf/, accessed on Aug. 20, 2023.

[35] Petry, N.M. (2002). A comparison of young, middle-aged, and older adult treatment-seeking pathological gamblers. The Gerontologist, 42(1): 92-99. https://doi.org/10.1093/geront/42.1.92

[36] Yiengprugsawan, V., Horta, B.L., Motta, J.V.S., Gigante, D., Seubsman, S.A., Sleigh, A. (2016). Body size dynamics in young adults: 8-year follow up of cohorts in Brazil and Thailand. Nutrition & Diabetes, 6(7): e219-e219. https://doi.org/10.1038/nutd.2016.24

[37] Kao, K.C. (2004). Dielectric Phenomena in Solids. Elsevier, pp. 321.

[38] Zhou, Q., Shan, J., Ding, W., Wang, C., Yuan, S., Sun, F., Li, H., Fang, B. (2021). Cough recognition based on mel-spectrogram and convolutional neural network. Frontiers in Robotics and AI, 8: 580080. https://doi.org/10.3389/frobt.2021.580080

[39] Tailor, S.A., Fernandez-Marques, J., Lane, N.D. (2020). Degree-quant: Quantization-aware training for graph neural networks. arXiv preprint arXiv:2008.05000. https://doi.org/10.48550/arXiv.2008.05000

[40] Toulson, R. Wilmshurst, T. (2017). Chapter 11-Wireless communication-bluetooth and zigbee. In Fast and Effective Embedded Systems Design (Second Edition). Newnes, pp. 257-290. https://doi.org/10.1016/B978-0-08-100880-5.00011-6

[41] Jain, A., Sahu, R., Jain, A., Gaumnitz, T., Sethi, P., Lodha, R. (2021). Development and validation of a low-cost electronic stethoscope: DIY digital stethoscope. BMJ Innovations, 7(4): 45. https://doi.org/10.1136/bmjinnov-2021-000715

[42] Yang, C., Zhang, W., Pang, Z., Zhang, J., Zou, D., Zhang, X., Guo, S., Wan, J., Wang, K., Pang, W. (2021). A low-cost, ear-contactless electronic stethoscope powered by Raspberry Pi for auscultation of patients with COVID-19: Prototype development and feasibility study. JMIR Medical Informatics, 9(1): e22753. https://doi.org/10.2196/22753

[43] Bailoor, S., Seo, J.H., Schena, S., Mittal, R. (2021). Detecting aortic valve anomaly from induced murmurs: Insights from computational hemodynamic models. Frontiers in Physiology, 12: 734224. https://doi.org/10.3389/fphys.2021.734224

[44] Unser, M., Aldroubi, A. (1996). A review of wavelets in biomedical applications. Proceedings of the IEEE, 84(4): 626-638. https://doi.org/10.1109/5.488704

[45] Tariq, Z., Shah, S.K., Lee, Y. (2022). Feature-based fusion using CNN for lung and heart sound classification. Sensors, 22(4): 1521. https://doi.org/10.3390/s22041521

[46] Zhou, G., Chen, Y., Chien, C. (2022). On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks. BMC Medical Informatics and Decision Making, 22(1): 226. https://doi.org/10.1186/s12911-022-01942-2

[47] Nguyen, T., Pernkopf, F. (2022). Lung sound classification using co-tuning and stochastic normalization. IEEE Transactions on Biomedical Engineering, 69(9): 2872-2882. https://doi.org/10.1109/TBME.2022.3156293