Multiple-track PhonoCardioGraphy (PCG) and Artificial Intelligence (AI) to Detect Heart Defects

1st Saif Khan
Dept. of Electrical and Computer Engineering
Rice University
Houston, Texas, USA
sbk7@rice.edu

2nd Jade Huang
Dept. of Electrical and Computer Engineering
Rice University
Houston, Texas, USA
ch110@rice.edu

3th Ye Zhou
Dept. of Electrical and Computer Engineering
Rice University
Houston, Texas, USA
yz202@rice.edu

4th Rohitaa Ravikumar
Dept. of Electrical and Computer Engineering
Rice Univeristy
Houston, Texas, USA
rr71@rice.edu

***Abstract*—Heart disease is a leading global cause of mortality, and early detection of heart anomalies is crucial for prevention. However, current diagnostic tools such as ECG and PPG have limitations and require specialist interpretation. To address this, we propose Cairdio, a handheld device that utilizes PCG to detect heart anomalies. The device features a 4-track PCG recorder and an AI-powered mobile app that transfers data via Bluetooth Low Energy (BLE) that interprets recorded heartbeats, resulting in high accuracy of identifying abnormal heartbeats. Our goal is to provide an accessible tool for lower trained staff or even skilled professionals to improve early detection of heart disease, ultimately reducing mortality rates.**

I. Introduction

Heart disease is a global health crisis, responsible for an estimated 17.9 million deaths each year as reported by the World Health Organization (WHO). Early detection of heart anomalies is crucial for preventing deaths and improving patient outcomes. The Cairdio device, designed for low-cost health screening, has four tracks that attach to specific areas of the patient's chest, including the aortic valve, pulmonic valve, right ventricular area, and left ventricular area. The device captures the first (S1) and second (S2) heart sounds, which are the two primary sounds produced by a healthy human heart. S1 occurs during ventricular contraction, while S2 arises during diastole with the closure of the aortic and pulmonary valves. These heart sounds are gathered using a microphone attached to an Analog to Digital converter powered by an NRF5340 chip, quantized in 12-bit format.

While the spectral energies of these heart sounds are typically concentrated below 100 Hz, research indicates that patients with cardiovascular diseases exhibit significant spectral components at higher frequencies. Additionally, cardiac murmurs can span up to 500 Hz, causing various cardiac anomalies to overlap with the fundamental heart sounds at certain frequencies. To identify such anomalies, spectrograms are analyzed using Convolutional Neural Networks (CNNs), which have recently emerged as a popular choice for audio and speech processing.

This paper describes our teams effort on building Cairdio's AI and Firmware. The AI Implementation section will discuss detecting abnormalities by analyzing the spectral energies of the heart, and the firmware implementation section will discuss gathering data using Analog to digital converter and transferring the results to a mobile app using bluetooth low energy (BLE).

II. AI Implementation

A. Dataset

The PhysioNet/CinC Challenge 2016 aimed to address the lack of a significant open-source heart sound database available for researchers to train and evaluate automated diagnostics algorithms upon. The dataset used for training consisted of 5 independent sets (sets a through e) of normal and abnormal recordings related to pathological conditions such as Coronary Artery Disease (CAD), valvular diseases, or aortic stenosis. The length of recordings varied from 5 to 120 seconds, providing a range of data for researchers to work with.

The Birdsong Recognition Kaggle competition dataset is also used to develop the envelop method in order to reduce noise. The dataset consists of audio recordings of bird songs and associated metadata, including species, location, and date of recording. The dataset includes a total of 264 species of North American birds, with over 20,000 recordings in total. The recordings range in length from 1 to 30 seconds and were collected from a variety of locations and environments.

B. Data Preprocessing

Digital stethoscopes are used to obtain PCG data for decision-making purposes. However, prolonged recordings can be uncomfortable for patients. To address this issue, we choose a recording duration of 10 seconds, which typically contains

8 to 16 complete cardiac cycles, depending on the patient's heart rate (50 to 100 bpm).

To reduce the computational load of the network, we first apply a low pass filter with a cut-off frequency of 500 Hz to the raw signals, followed by down-sampling of the 2000 Hz audio signals to 1000 Hz. We then compute a logarithmic spectrogram using a Tukey window of length 256 (256 ms long) with 38% overlapping [1]. This results in a spectrogram with 64 time and 129 frequency bins. We discard the first frequency bin (corresponding to the DC term) to obtain a matrix of 64 x 128 x 1. The matrix is then normalized between 0 and 1 using min-max normalization.

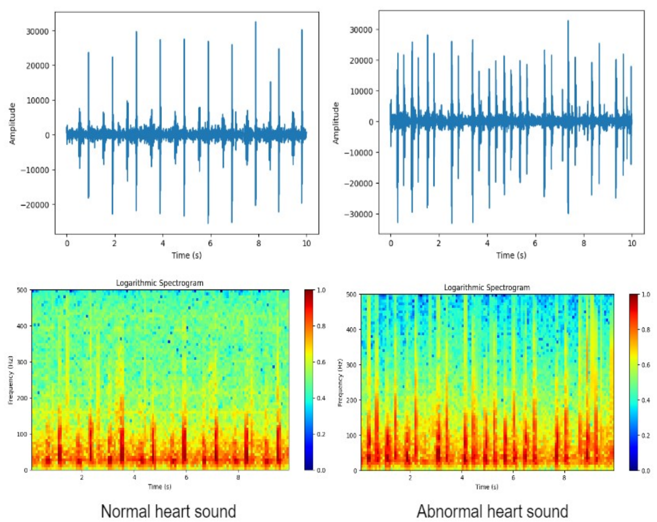

Fig. 1. Time vs Amplitude and Log-spectrogram plots of normal and abnormal heart sounds

Figure 1 depicts the raw signal and log-spectrogram plot of normal and abnormal heart sounds. The plot illustrates the distinguishing patterns in the spectrogram of murmurs at higher frequencies, which can aid in identifying cardiac anomalies.

After converting the signals to spectrograms, the image dataset is further split into two sets. The training set consists of 80% of dataset and the rest of 20% is assigned for validation.

C. *Deep Learning Models*

1) *Binary Classifier*

A binary classifier is used for predicting normal and abnormal heart sounds using supervised learning. In this paper, we propose a Convolutional Neural Network structure for efficient feature extraction from the spectrogram of heart sounds. The model contains a total of 5 convolutional layers, each of them is associated with a maxpool layer for dimensionality reduction. The kernel size is taken as 3 x 3 and filters ranging from 32 to 256 are used in the convolutional layers. The pool size is selected as 2 x 2. The output of the final maxpool layer is flattened. Rectified Linear Unit (ReLU) is used as the non-linear activation function in the convolutional layers of the CNN except the final layer, which takes a sigmoid activation function. The model trains on images of both normal and abnormal heart signals and uses binary crossentropy loss function. During training, the objective function of the CNN is optimized using an Adam optimizer.

2) *Autoencoder*

Autoencoder is a type of neural network used for unsupervised data encoding. It can be utilized for anomaly detection by computing an anomaly score for an unknown test instance. A score exceeding a pre-determined threshold, suggests the presence of an anomaly.

The proposed architecture is shown in Figure 2, comprises of an encoder and a decoder. The encoder consists of three convolutional layers, each with a 3x3 kernel size and 16, 8, and 8 filters, respectively, followed by max-pooling layers with a 2x2 pool size. The output of the last max-pooling layer represents the latent space representation of the input. The decoder mirrors the encoder architecture with three convolutional layers. However, the maxpool layers are replaced by upsample layers of similar size to gradually enhance the encoded vector dimension to the original space. The output of the final upsample layer is applied to an additional convolutional layer, having a single filter for the reconstruction of the original spectrogram. The decoder is optimized to reconstruct the original input image, while the encoder learns to capture the salient features of the input in the compressed latent space representation.

The autoencoder model was trained using the Adam optimizer and Mean Squared Error (MSE) loss function. The training was conducted with a batch size of 32. The model was evaluated using the validation data. The training aimed to minimize the difference between the original and reconstructed input signals, as captured by the MSE loss function.

D. *Noise reduction*

1) *Bandpass filter*

A bandpass filter works by selecting a specific center frequency and a range of frequencies around that center frequency, known as the passband. Frequencies within the passband are allowed to pass through the filter with minimal attenuation, while frequencies outside of the passband are blocked or attenuated. To preprocess the PhysioNet/CinC Challenge 2016 dataset, a bandpass filter is applied. And both the low cutoff bandpass value and the high cutoff bandpass value are adjusted in the Matlab code.

2) *Envelop method*

The sound envelope is a concept in music and audio processing that describes the shape of a sound waveform over time. It refers to how the amplitude, or volume, of a sound changes over time, and can be broken down into four distinct phases: attack, decay, sustain, and release (ADSR). For example, the original waveform and its corresponding envelope are shown in Fig. 3 and Fig. 4 respectively.

In this project, sound envelope is used as a metric to distinguish between desired voice (pure birdcall) from unwanted noise (other background voice).

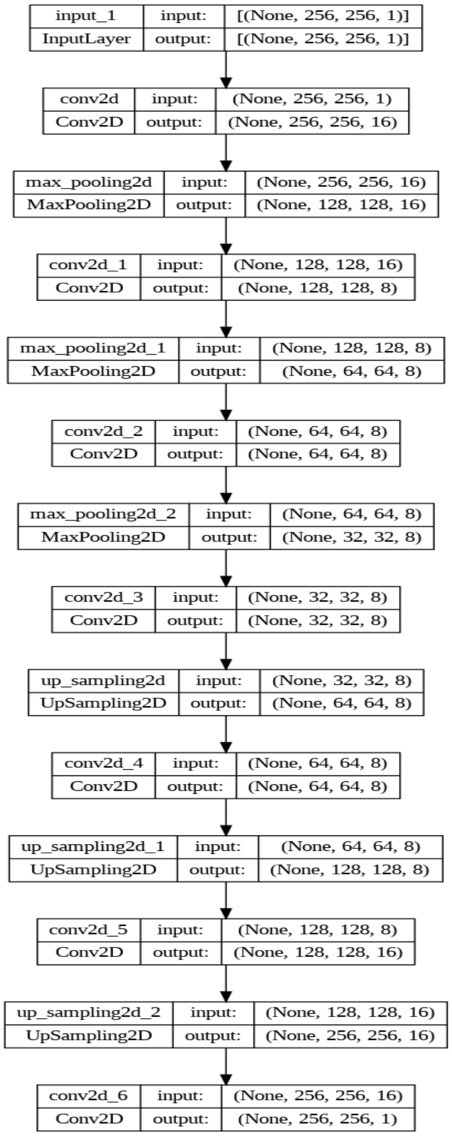

Fig. 2. Model Architecture for Autoencoder

Rolling window is used to calculate the mean of absolute amplitude value of the waveform and a threshold value is defined in advance. Then a mask array is calculated for the waveform. If the mean is below the threshold, false value is applied to the mask, which means this waveform is unwanted noise. Similarly, if the mean is above the threshold, true value is applied to the mask, which means the waveform is desired voice. A processed waveform is shown in Fig. 5.

III. Firmware

The Cairdio firmware comprises of several components, including the nRF5340 chip, a custom PCB board, four stethoscope headers with microphones, and a rechargeable power supply. The stethoscope headers are affixed to the custom PCB using a 3D-printed case to record phonocardiogram (PCG) signals from the upper and lower atria and ventricles.

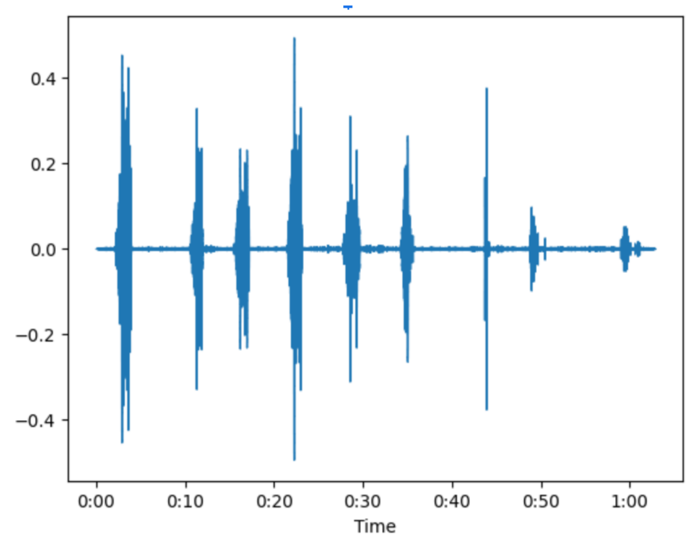

Fig. 3. Original waveform of a birdcall recording

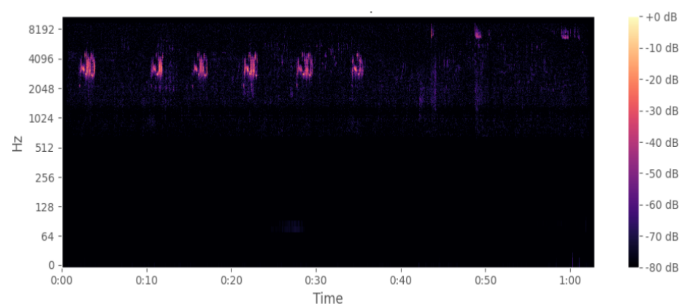

Fig. 4. Sound envelope of a birdcall recording

Once the heart sounds are captured by the four microphones, an analog-to-digital converter (ADC) attached to the PCB board is used to convert the analog signals into digital format. When sound waves are detected by a microphone, they are initially in analog format. The ADC samples the incoming analog signal at a high rate, typically thousands of times per second, and converts each sample into a digital value. The resulting digital signal can then be stored, edited, and processed by digital audio equipment, leading to high-quality recordings. The recordings can then be transmitted via built-in Bluetooth Low Energy (BLE) to a mobile application for further AI processing.

A. Bluetooth Low Energy (BLE)

The nRF5340 DK is a powerful, dual-core System-on-Chip (SoC) designed by Nordic Semiconductor. It is built to deliver exceptional performance and versatility, making it an ideal choice for a wide range of applications, including advanced wearables, smart home devices, and industrial automation. One of its key features is its Bluetooth Low Energy (BLE) functionality, which enables low-power, wireless connectivity for IoT devices. The nRF5340 BLE implementation provides several advanced features, including high throughput, long-range, and secure connections, making it an excellent choice

Fig. 5. Processed audio

Fig. 6. Ideal Firmware Architecture

for developers looking to build high-performance and reliable IoT solutions.

The nRF5340 DK that are resource constrained and require low power are usually the peripheral device in a Bluetooth LE connection, while the central device is like a mobile phone, which has more power [3]. Currently, we demonstrated a sample on how to establish a BLE connection with a central device (using nRF Connect mobile App), send and receive data over UART, and handle errors and events.

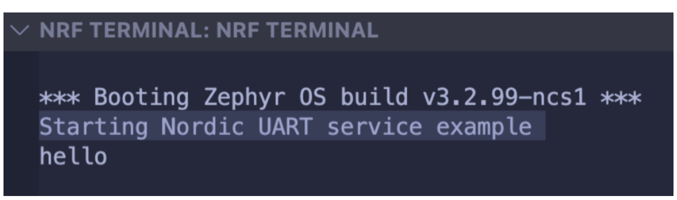

Fig. 7. Received Data via BLE

Next step for BLE is to increase throughput by using a larger Attribute Protocol (ATT) Maximum Transmission Unit (MTU). The MTU size is negotiated between the central and peripheral devices during the initial connection setup, and it directly impacts the throughput and compatibility of the data transfer. Increasing the MTU size can result in faster data transfer rates, but it also requires more memory and can lead to compatibility issues [4]. We have to optimize the MTU size is crucial for achieving the best balance between throughput and compatibility in our BLE communications in order to transmit audio files within reliable connection.

An example code can be found in the private github repository Cairdio BLE Sample Programming Guide.

B. Acquiring Heartbeat Data using ADC's

The NRF5340 chip [5] houses 6 ADC channels that can sample up to 12-bit of data and 14-bit with oversampling. For our purposes we only need 4 channels and only one of those channels will be used for demonstrating the results.

In order to acquire the heartbeat sounds, an acoustic microphone [6] is connected to a first order passive high pass filter and then connected to a LM386 Audio Amplifier [7] with a gain of 200. The low pass filter used was a first order filter. The frequency to be blocked was , f = 400 Hz ,and the resistor chosen was 2.2kohms as per the microphone datasheet [6] so the capacitor calculated using the formula , C=1/(2*pi*400Hz*2.2kohms) = 0.18uF . Since a 0.18uF was not available a 0.1uF non-polarized capacitor was used instead which allows frequencies up until 723Hz to pass through. The Schematic is shown in Figure 8 below:

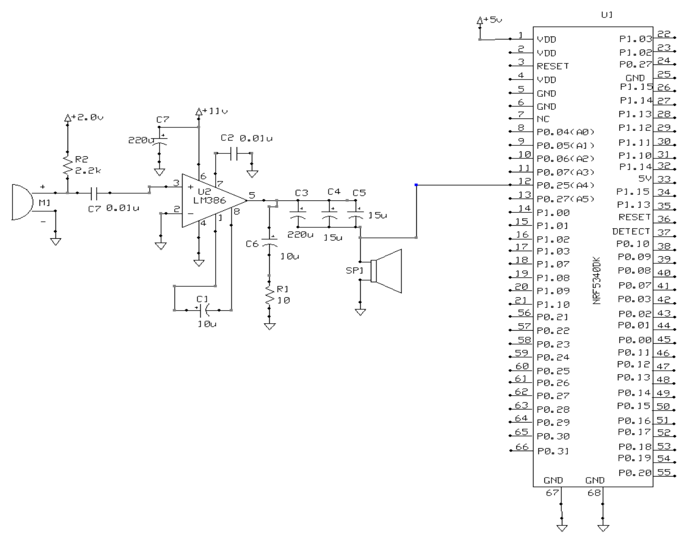

Fig. 8. ADC Circuit Setup Schematic.

Below is an image of the setup in real time:

The image below shows an artificial heartbeat sound [8] was sent to only test the circuit and the frequency of this signal is unknown but it is well within the 723Hz for testing the circuit:

The image below is of same amplified heart sound but zoomed in:

From Figures 10 11, CH1 1 (Yellow) CH2 (Blue) it can be seen that the heartbeat sounds are amplified as the amplitude of the signal at the output is clearly larger , the high gain is causing the voltage reading to exceed oscilloscope display. This means that a lower gain is required to display the artificial heartbeat sound on the oscilloscope and since this is just used to test ADC channel it fits our purpose.

The image below shows the output on the console when reading the ADC channel as shown in Figures 8 9:

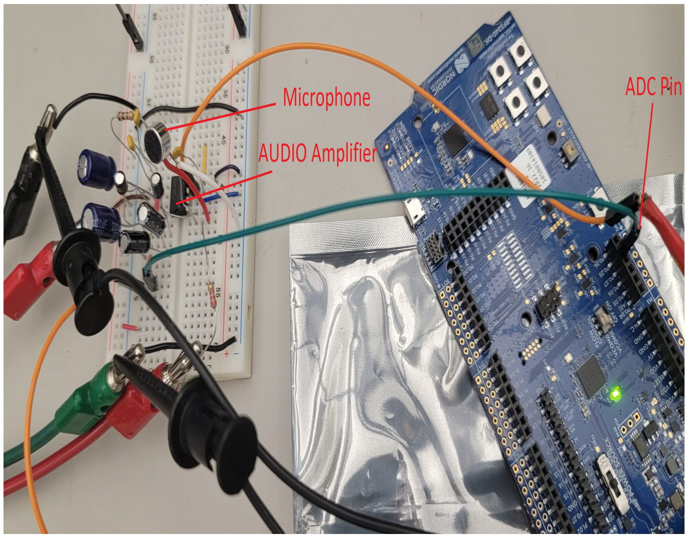

Fig. 9. ADC Circuit Setup.

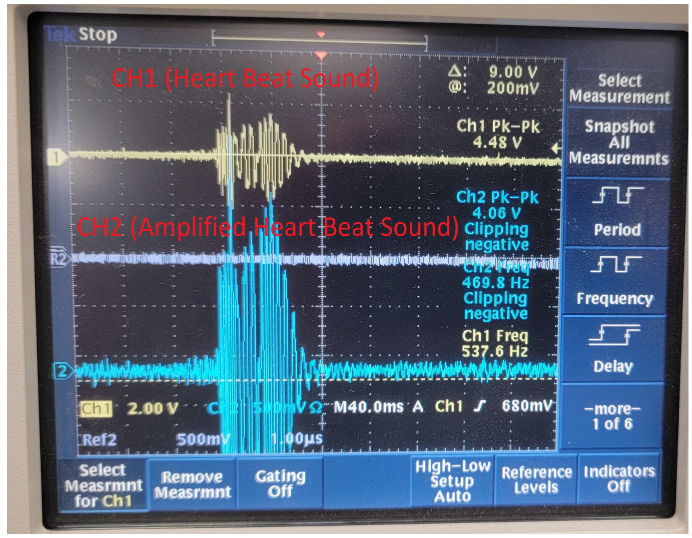

Fig. 11. Amplified Heartbeat sound on a oscilloscope zoomed out

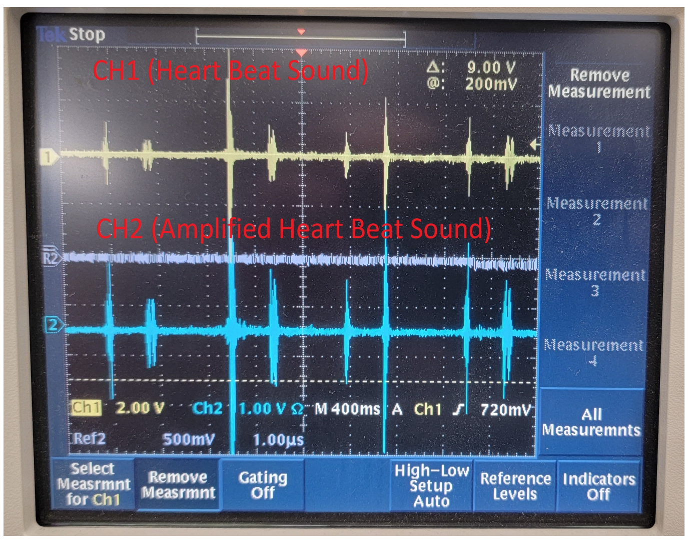

Fig. 10. Amplified Heartbeat sound on a oscilloscope zoomed out

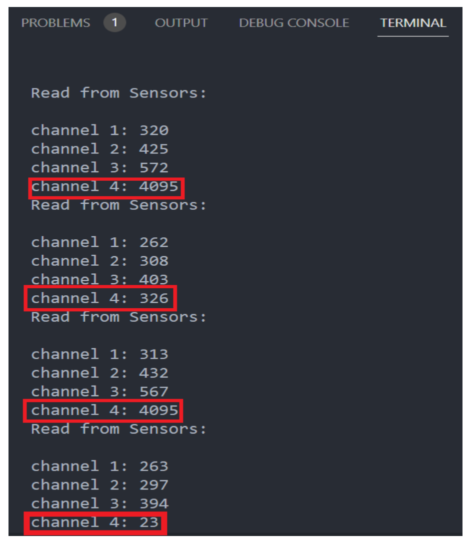

Fig. 12. Amplified Heartbeat sound on a oscilloscope zoomed out.

Figures 12 and 13, the rest of the channels are idle and are not connected except for channel 4 ,highlighted in red. As the amplitude of the waveform increases the voltage reading increase up until 4095 (12-bit) and then decreases down to values like 23 or -103 and then goes back up to 4095, this is clearly expected behavior since the heartbeat sound also increases and then decrease on every cycle as shown in figures 10 and 11. This also shows the 12-bit is not enough to represent this heartbeat sound at this gain and requires increased sampling such as 14-bit instead.

IV. Results

A. AI Results

The proposed networks were trained in Google Colab, without using a graphics processing unit. The implementation is done in Python using TensorFlow. The training of the CNN model was observed to have 90.5% training accuracy and 92.5% validation accuracy. The model demonstrated an overall accuracy of 84.14%, with a specificity of 89.04%. The ability of the model to classify between normal and abnormal heart sounds is evident.

The prediction results of auto encoder model were in terms of reconstruction error and it was observed that for few noisy but normal recordings, the overall reconstruction error is found high, causing false positive in prediction of normal recordings.

B. Noise Reduction

In this project, SNR (signal to noise ratio) is used as the metric to determine the performance of the envelope method. For the experiment dataset, a subset is extracted to illustrate the experiment results. For each audio file in this subset, SNR

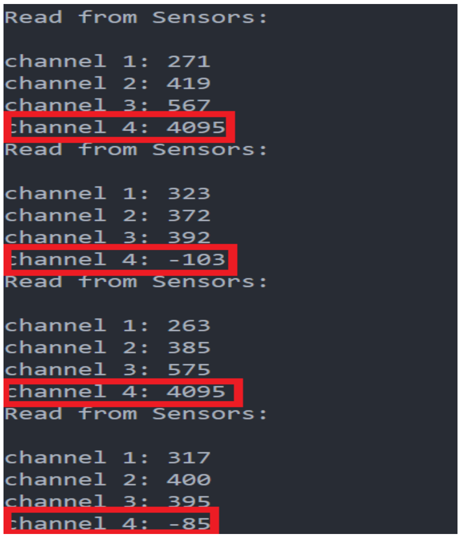

Fig. 13. Amplified Heartbeat sound on a oscilloscope zoomed out.

value is calculated after the noise is separated from the signal as shown in Figure 7.

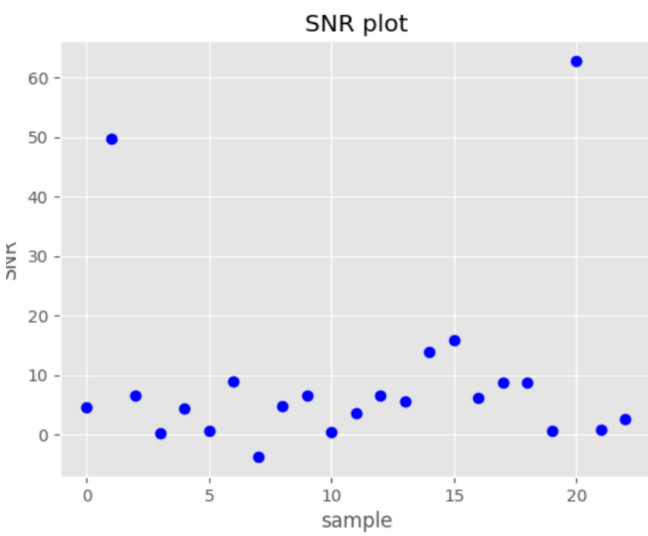

Fig. 14. SNR plot

The average SNR achieved for this subset is 9.502.

C. ADC Results

The result was successful , Figures 10 11 show the amplified heartbeat sound using the LM386 audio amplifier on CH2 (Blue) of the oscilloscope. When the amplified sound is quantized using an ADC on channel 4 of the nrf5340 chip (Figure 12 13) we can see that the analog voltage values are increases and decreasing proportional to the waveforms in Figures 10 11. Since this is an artificial heartbeat and the gain was high the analog voltage went up to 4095. This worked for the our purpose of testing the ADC's to acquire heartbeat data using NRF5340 chip. Source code explanation and additional testing is added to the github repository, however this is private.

V. Conclusion

A. AI Implementation

Automatic detection of abnormal heart sounds is a popular research problem. In this paper, we propose a supervised and unsupervised approach using Convolution Neural Network and Auto encoder respectively to solve the problem. The proposed approach is successfully evaluated on dataset with cardiac abnormalities. The code can be found in the private github repository under Cairdio ADC Programming Guide.

B. Acquiring Heartbeat Data using ADC's

The ADC channels are able to acquire data from the microphone after the signal is amplified using an audio amplifier. As shown in the results the output audio signal from the amplifier is has a much larger amplitude and the ADC analog voltage values are proportional to changes shown in the heartbeat sound. Since actual heartbeat data is a low frequency and low amplitude signal, the gain may need to be higher than what LM386 amplifier can offer and an active bandpass filter to filter out the noise frequencies between 20Hz and 400Hz will be needed.

C. Bluetooth low energy example

It initializes the LEDs and user button, sets up the Lightbulb Service, and starts advertising the device. When a remote device connects to the peripheral, the connected callback is called and an LED is turned on to indicate the connection status. When the device disconnects, the disconnected callback is called and the LED is turned off.

Acknowledgment

We would like to thank prof.Joseph Young , Rick jones, Ethan peck, Darya Yelshyna , Arnab Saha ,Carlos Alejano and the entire healthseers team for their support.

References

[1] R. Banerjee and A. Ghose, "A Semi-Supervised Approach For Identifying Abnormal Heart Sounds Using Variational Autoencoder," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 1249-1253, doi: 10.1109/ICASSP40776.2020.9054632.

[2] S. Li, K. Tian and R. Wang, "Unsupervised Heart Abnormality Detection Based on Phonocardiogram Analysis with Beta Variational Auto-Encoders," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 8353-8357, doi: 10.1109/ICASSP39728.2021.9414165.

[3] "Advertising process," Nordic Developer Academy. [Online]. [Accessed: 25-Apr-2023].

[4] N. 7, "Maximizing ble throughput part 2: Use larger ATT MTU," Punch Through, 25-Apr-2023. [Online]. [Accessed: 25-Apr-2023].

[5] Semiconductor, Nordic. "NRF5340 DK - Development Kit for the NRF5340, a Dual-Core Bluetooth 5.2 SOC Supporting Bluetooth Low Energy, Bluetooth Mesh, NFC, Thread and Zigbee." Nordic Semiconductor, www.nordicsemi.com/Products/Development-hardware/nRF5340-DK. Accessed 27 Apr. 2023.

[6] "Microphones: AUM-4537L-HD-R." Puiaudio.Com, 12 Sept. 2022, puiaudio.com/product/microphones/AUM-4537L-HD-R.

[7] Instruments, Texas. “LM386.” LM386 Data Sheet, Product Information and Support — TI.Com, www.ti.com/product/LM386. Accessed 27 Apr. 2023.
[8] Sound, Heartbeat. “Heartbeat Sound.” YouTube, 2 Feb. 2015, www.youtube.com/watch?v=HUxS21-cf80t=11548s.