BIRLA INSTITUTE OF SCIENCE AND TECHNOLOGY, PILANI

Design Project (CS F376)

A Report on

“Heartbeat Anomaly Detection”

Written by

Manan Agarwal (2016B5A70607P)

Ankita Chakravarty (2016B5A70701P)

Introduction

Heart disease is the leading cause of death globally, resulting in more people dying every year due to cardiovascular diseases (CVDs) compared to any other cause of death [World Health Organization, 2017]. Rural mortality rates have surpassed those of urban areas as 75% of rural primary-care is handled by unqualified practitioners owing to an acute shortage of doctors.

Currently, there are two effective ways to monitor one's heart condition; electrocardiogram (ECG) and echocardiogram. However, both methods are relatively expensive for mass inspection and require technical expertise in using them. Our goal is to develop a reliable, fast and low-cost system that can be used by untrained frontline health workers or anyone with internet access, to help determine whether an individual should be referred for expert diagnosis, particularly in areas where access to clinicians and medical care is limited. This will also help in early diagnosis of CVDs which will drastically decrease the potential risk factors of these deaths [1].

In this work, we take stethoscope sounds and even waveforms recorded using the microphone of a mobile phone as input and apply deep learning to the task of automated cardiac auscultation, i.e. recognizing abnormalities in heart sounds. We describe an automated heart sound classification algorithm that combines the use of time-frequency heat map representations with a deep convolutional neural network (CNN).

Dataset

This dataset was originally used in a Machine Learning challenge for the classification of heartbeat sounds by Mr Peter Bentley [2]. The dataset is divided into 2 sets depending on the sources from where it was collected. Set A (set_a.csv) data was collected from the general public via the iStethoscope Pro iPhone app and Set B (set_b.csv) from a clinical trial in hospitals using the digital stethoscope DigiScope.

In this dataset there are 4 classes of heartbeat sounds:

1. Normal: healthy heart sounds
2. Murmur: extra sounds that occur when there is turbulence in blood flow that causes the extra vibrations that can be heard
3. Extrahls: heartbeats with an additional sound
4. Extrasystoles: are additional heartbeats that occur outside the physiological heart rhythm and can cause unpleasant symptoms

As a part of the project, we were required to collect heartbeat from local clinics and hospitals to test our model. However, the couple of hospitals we talked to denied permission to do so on the grounds of acquiring an IRB approval.

The Institutional Review Board (IRB) deals with the review of behavioural research which involves human participants. The IRB approves and monitors all studies involving human participants to ensure that the experiment or survey is designed to meet legal and ethical concerns and does not involve unnecessary risks for the participants. All proposed research involving ‘human subjects’ must be reviewed and approved by the IRB before recruitment and data collection can start. The IRB ensures that the participants are adequately informed of what their participation entails (e.g.: risks and benefits). This includes written and signed informed consent in most cases.

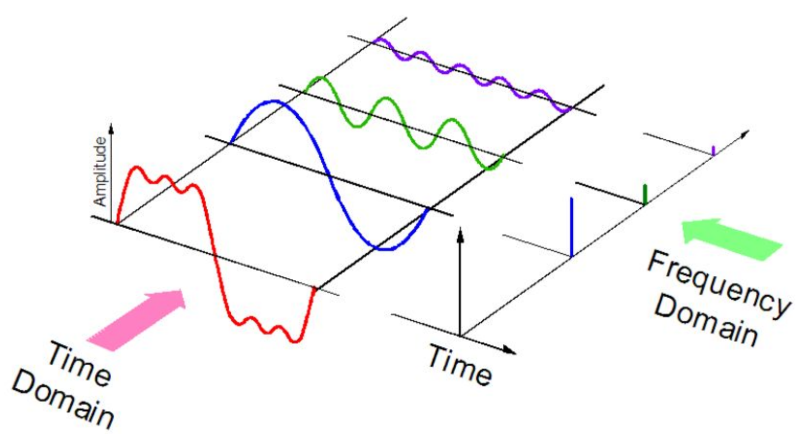
Despite the above-mentioned constraint, we were successfully able to collect 58 data samples from small local clinics for testing our model.

Pre-Processing

The heart sounds recorded by digital stethoscope and the mobile phone microphone often has background noise. The preprocessing of heart sounds is an essential and crucial step for automatic analysis of heartbeat recordings. We have cut out all the audio files that have a duration of fewer than 3 seconds because they do not contain enough data points to accurately classify heartbeats. The recordings have to be converted to some fixed length prior to training, we slice the heart sounds into fixed-length segments of length 3 sec. To increase the size of the dataset we are slicing large files into multiple smaller files while still retaining their original label (i.e. normal or abnormal) We are also cutting about half a second from the start and end of all the audio files because the noise is due to the contact of the microphone with the body.

Feature Extraction

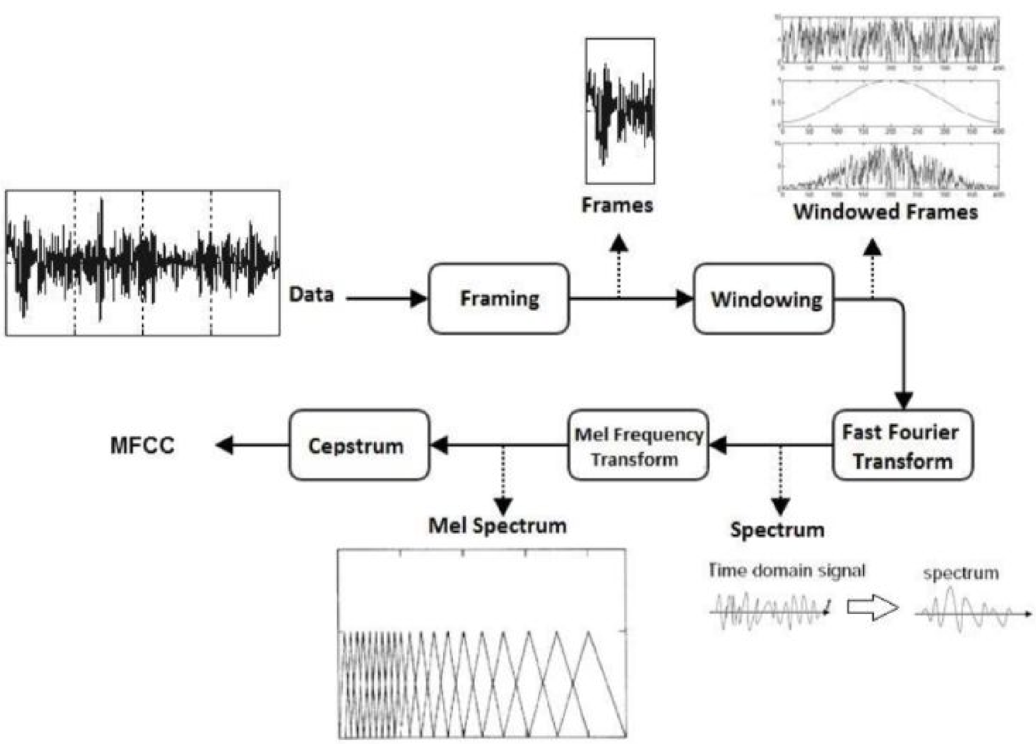
The raw audio data used .wav type (Waveform Audio File Format) was in the amplitude vs time form in the time domain. We transformed this one-dimensional time-series signal into a two-dimensional heat map that captures the time-frequency distribution of the signal. This representation of the spectrum of frequencies of a signal as it varies with time is called a spectrogram. Since the data was collected using different instruments (ie. digital stethoscopes and mobile phone microphone) which results in varying amplitude ranges, converting the data to the frequency domain leads to more accurate results.



In the frequency domain, we have used the following for feature extraction:

- **Melspectrogram**: It represents an acoustic time-frequency representation of a sound. It is basically a spectrogram with the Mel Scale as its y-axis. This Mel Scale is constructed such that sounds of equal distance from each other on the Mel Scale, also “sound” to humans as they are equal in distance from one another.

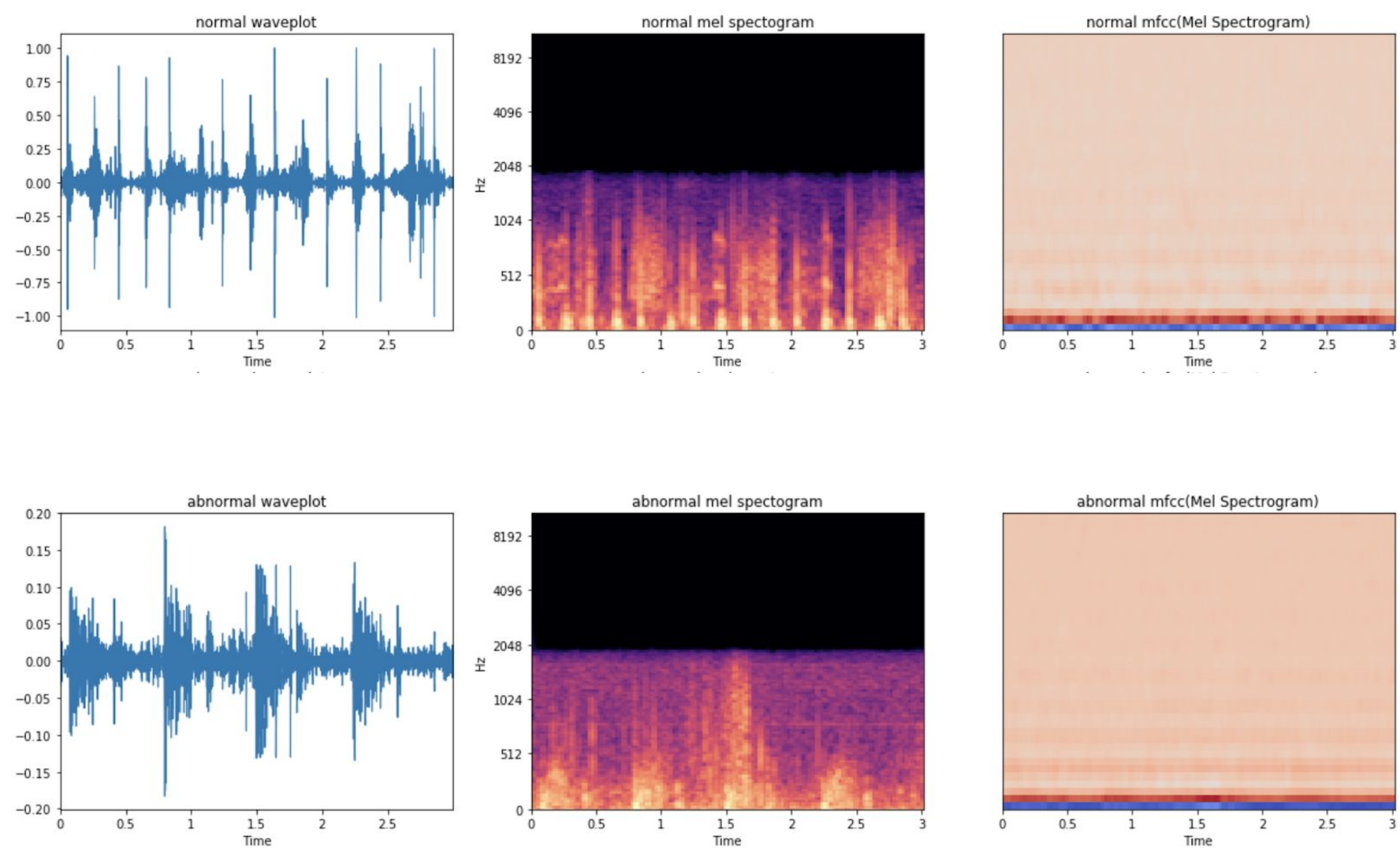
- **MFCC (Mel Frequency Cepstral Components)** [3]: We chose to use Mel Frequency Cepstral Coefficients to perform this transformation, as MFCCs capture features from audio data that more closely resembles how human beings perceive loudness and pitch. MFCCs are commonly used as a feature type in automatic speech recognition. [4]



Process Flow to get the informations in Frequency Domain from Time Domain

From librosa library in python, we used the inbuilt function for the above as follows:

1. The preprocessed sound files are transformed into magnitude spectrogram and then mapped onto the mel-scale by using the inbuilt feature method of librosa to get the mel-scaled spectrogram.
2. A cepstral analysis is performed on the Mel-Spectrum to obtain Mel-Frequency Cepstral Coefficients (MFCC) by passing the log-power Melspectogram as an argument to the MFCC function.
3. Thus the audio file is now represented as a sequence of Cepstral vectors. These Cepstral vectors are then given to the model for anomaly detection



Model

The original one-dimensional time series data is transformed into a two-dimensional time-frequency representation (i.e. spectrogram), which allows each heart sound segment to be processed as an image. The Convolutional Neural Network (CNN) is one of the neural network architecture specifically used for image classification. Just like other neural network methods, CNN is also inspired by human brain tissue. Convolution neural network is mainly composed of two parts, feature extraction, and classification.

The network architecture of a convolutional neural network that accepts as input a single channel 40x130 MFCC heat map and outputs a binary classification, predicting whether the input segment represents a normal or abnormal heart sound.

Convolutional Neural Network in this study uses 4 convolution layers, 4 max-pooling layers, 4 dropout layers, 1 global average pooling layer and finally a dense layer. The activation function in convolution layers uses Rectifier Linear Unit (ReLU) algorithm. The ReLU algorithm has advantages in time efficiency for training and testing.

The dropout layer: The term "dropout" refers to dropping out units (both hidden and visible) in a neural network. It is a very efficient way of performing model averaging with neural networks. Model averaging is a natural response to model uncertainty. The dropout layer allows for regularization by randomly setting some neurons in previous layers to zero during training.

Max Pooling: The objective of Max pooling is to down-sample an input representation. It helps in reducing the dimensionality and alleviate feature extraction. It reduces the computational cost-reducing the number of parameters to learned.

Dense layer: Here every input is connected to every output by weight. And we are using softmax as the non-linear activation function after this layer.

Adam method is used for the optimization process to update the weight on the Convolutional Neural Network. This method has efficient computation (memory and time), invariant to gradient scaling and suitable when applied to large data or parameters.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 39, 129, 16)       80
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 19, 64, 16)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 19, 64, 16)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 18, 63, 32)        2080
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 9, 31, 32)         0
_________________________________________________________________
dropout_2 (Dropout)          (None, 9, 31, 32)         0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 8, 30, 64)         8256
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 4, 15, 64)         0
_________________________________________________________________
dropout_3 (Dropout)          (None, 4, 15, 64)         0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 3, 14, 128)        32896
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 1, 7, 128)         0
_________________________________________________________________
dropout_4 (Dropout)          (None, 1, 7, 128)         0
_________________________________________________________________
global_average_pooling2d_1 ( (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 258
=================================================================
Total params: 43,570
Trainable params: 43,570
Non-trainable params: 0
_________________________________________________________________
```

Conclusion

This study was designed to detect an abnormal heartbeat sound using Stethoscope sounds and heartbeats recorded through the microphone of a mobile phone. Classification of heartbeat sounds was conducted using a Convolutional Neural Network. We did not use any other time sequence based Neural Networks such as RNNs since the temporal behavior of the heartbeat was repeated within the window of observation and different sequential patterns were not needed to be learnt.

The work presented here is one of the first to apply deep convolutional neural networks to the task of automated heart sound classification of heartbeat sound recorded through a stethoscope. We developed a novel algorithm first transforms the one-dimensional time-series input into a two-dimensional time-frequency Melspectrogram. It then trains a 4-layer CNN architecture on the MFCC obtained from the Melspectrogram. The trained network automatically distinguish between normal and abnormal heartbeat sound inputs. The epoch values used were 100, 150, 200, 250 and 300. The best results were obtained with 300 epoch at 0.001 learning rate applied on batch size of 128. The training accuracy is 89.73, while the testing accuracy rate is 84.04.

	precision	recall	f1-score	support
abnormal	0.78	0.80	0.79	97
normal	0.88	0.86	0.87	160
accuracy			0.84	257
macro avg	0.83	0.83	0.83	257
weighted avg	0.84	0.84	0.84	257

References

[1] [Z.-J. Yang, J. Liu, J.-P. Ge, L. Chen, Z.-G. Zhao, and W.-Y. Yang, 2011] “Prevalence of cardiovascular disease risk factor in the Chinese population: the 2007–2008 china national diabetes and metabolic disorders study,” European heart journal, vol. 33, no. 2, pp. 213–220, 2011.

[2] [Bentley P. and Nordehn G. *et al.*,2011] The Classifying Heart Sounds Challenge 2011 http://www.peterjbentley.com/heartchallenge/index.html"

[3] [Davis and Mermelstein, 1980] Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE transactions on acoustics, speech, and signal processing, 28(4):357–366, 1980.

[4] [Luigi Bungaro, 2018] How to detect anomalies in Audio Signal Processing of the heart with the sound coming from mobile phone https://medium.com/@luigi.bungaro/how-to-detect-anomalies-in-audio-signal-processing-of-the-heart-with-sound-coming-from-mobile-e034e8fd709b

[Jonathan Rubin, *et al.* 2017] Recognizing Abnormal Heart Sounds Using Deep Learning arXiv:1707.04642v2

[Gaurang Jungare *et al,* 2018] Heart Anomaly Detection using Deep Learning approach based on PCG signal analysis. IRAJ International Conference, 2018, Pune, India

[Siddique Lati *et al*, 2018] Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection arXiv:1801.08322v3, IEEE