A Multi-Class Cardiac Sound Diagnostic System in Deep Learning Based on PCG Signal

Babna K.

Student, Electronics and Communication Engineering, KMCT College of Engineering, Kozhikode, Kerala, India
ludmeelababi[at]gmail.com

Abstract: ***Heart sound classification plays a vital role in the early discovery of cardiovascular diseases, especially for small primary health care conventions. Despite that important progress has been made for heart sound bracket in recent times, utmost of them are grounded on conventional segmented features and shallow structure-grounded classifiers. These conventional aural representation and classification styles may be inadequate in characterizing heart sound, and generally suffer from a degraded performance due to the complicated and changeable cardiac aural terrain. In this paper, a new heart sound bracket system has been proposed grounded on mongrel features of heart sound signals and convolutional intermittent neural network classifier model. Then uses MFCC and HCQT features for the heart sound spectrograms. And the model categorizes five classes of heart sounds in an effective way using Convolution Neural Network models.***

Keywords: Cardiovascular diseases, Deep Learning, Phonocardiogram signal, Unsegmented heart sounds, Convolution neural network

1. Introduction

As per WHO (World Health Organizations) records that Cardiovascular diseases (CVDs) are the leading cause of death encyclopedically. An estimated17.9 million people died from CVDs in each time, representing 32 of all global deaths. Of these deaths, 85 were due to heart attack and stroke. Over three quarters of CVD deaths take place in low- and middle-income countries. Out of the 17 million unseasonable deaths (under the age of 70) due to non-infectious conditions,

38 were caused by CVDs. Utmost cardiovascular conditions can be averted by addressing behavioral threat factors similar as tobacco use, unhealthy diet and rotundity, physical inactivity and dangerous use of alcohol. It's important to descry cardiovascular complaint as early as possible so that operation with comforting and drugs can begin.

Heart, one of the most important organs of the mortal body produces distinct sounds during its course of mechanical exertion. Since the circumstance of a specific type of complaint alters the heart functionality in a definite manner, the auscultations also change consequently and therefore, they've been historically employed for screening CVDs. Although Lub (S1) and Dub (S2) play vital part for the discovery of cardiac anomalies, occasionally the irregular variants, third heart sound (S3), fourth heart sound (S4) and murmurs are also taken into account while making pathological conclusion using the stethoscope [1]. The simple, non-invasive nature of this auscultation-grounded opinion strategy have made it the most popular and seductive choice of the cardiologists for performing the original disquisition. Besides auscultation, multiple other cardiac signals involving a wide range of advanced styles similar as, electrocardiogram (ECG), myocardial perfusion imaging (MPI), angiography, echocardiography, cardiac reckoned tomography (CCT), cardiovascular magnetic resonance (CMR), carotid pulse graph, apex cardiogram etc. are being employed as ultramodern individual tools for effective screening of CVDs as they vividly reflect the overall transthoracic physiological conditions of the cardiovascular system. Still, in situations where these advanced styles are scarce, lung auscultation stands out as a simple and dependable medium for detecting CVDs [3]. Nonetheless, indeed for an expert physician, it's relatively gruelling to readily descry CVDs just by harkening to the auscultation. This situation is farther aggravated owing to private different interpretation of the same auscultation by the croakers. The disproportionate numbers of professed medical professionals with respect to the total population adds up to the situation and farther slacken the original individual speed. In this script, artificial intelligence-empowered automated cardiac webbing systems on the base of PCG bracket can play a vital part to help the physicians in their decision-making process and can also be used by laypeople in the absence of doctors [5].

Since phonocardiogram (PCG), the visualization of heart sound on graphical waveform, can be fluently reused to prize essential discriminative features for the identification of cardiac anomalies, covering heart condition via PCG is getting a decreasingly popular clinical practice. With the admixture of machine literacy (ML), state-of-the-art networks and advanced audio processing ways, the unpropitious homemade webbing can be replaced with automated bracket fabrics for prompt large-scale prognostications. Thus, multitudinous exploration workshop has explored the sphere of automated PCG classification over the times.

Any system which can help to descry signs of heart complaint could thus have a significant impact on world health. Then we're trying to produce a Deep Literacy Model to identify the sign of heart diseases using the Heartbeat sound [7]. Data is gathered in real-world situations and constantly contains background noise of every conceivable type. Success in classifying this form of data requires extremely robust classifiers

The proposed system is a new heart sound classification model where mongrel feature extractions of MFCC and

HCQT (Hybrid Constant Q Transform) [9] [17] are used in neural network platform. CNN (Convolution Neural Network) uses for the classification of heart rhythm which will help for the diagnosis of cardiovascular conditions.

2. Literature Survey

Diagnose at an early stage is the only way to decrease the death rate due to CVD. There are many invasive and non-invasive methods to diagnose CVD. All Invasive techniques are costly, painful, and readily unavailable at all places, especially in remote areas. Usage of a non-invasive method to diagnose CVD at an early stage is less expensive and painless [1]. ECG and PCG are two such non-invasive ways to diagnose CVD. But their analysis requires an expert doctor of this domain which is not readily available in remote areas [2]. When sounds and murmurs occur during the cardiac cycle are represented diagrammatically, it is called a phonocardiogram. These vibrations generate the wave, which propagates through the chest wall. A stethoscope, a low-cost hand held digital device, is used to record the information generated through acoustic waves. It gives us an estimation of parameters like heart rate, intensity, tone, quality, frequency, and location of various components of the cardiac sound, which helps in the diagnosis of CVD in a non-invasive manner [3]. Machine learning and deep learning algorithms have allowed us to create decision support systems that can help doctors and can also be used by laypeople in the absence of doctors [4].

A. E. F Malik et al. [5] have designed a Scalogram and CNN based model to diagnose cardiovascular diseases from PCG signals. They have applied segmentation method [7] to convert each PCG signals from three cardiac cycles to one cardiac cycle [12]. The used 2-D ConvNet model for final classification [18], [19]. And they achieved high accuracy, but only for three classes of heart sounds. Some authors have used time varying spectral features with different classifier [10]. They used SVM and KNN for binary class classification. S. Patidar et al. [7] have designed a model to detect septal defects by analyzing cardiac sound signals. The authors have used the TQWT based advanced signal processing technique to fetch cycles of the heart beat from cardiac sound signals.

3. Materials and Methods

The proposed system helps to detect the sign of heart diseases. The model uses Artificial Neural Network to classify deasesd and non-deasesed heart sound. The features extraction is difficult in this case because of the nature of data. We need to extract the sound features from the data and use that features to train the model. The basic block diagram of the process is shown inFig.1.

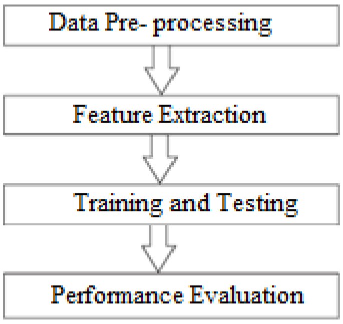


Figure 1: Basic block diagram of the system

A. Data Description

This dataset was originally for a machine learning challenge to classify heart beat sounds. The data was gathered from two sources: (A) from the general public via the iStethoscope Pro iPhone app, and (B) from a clinic trial in hospitals using the digital stethoscope DigiScope.

The audio files are of varying lengths, between 1 second and 30 seconds.

The dataset is split into two sources, **A** and **B**:
set_a.csv-Labels and metadata for heart beats collected from the general public via an iPhoneapp
seta**timing.csv**-contains gold-standard timing information for the "normal" recordings from Set A.
set_b.csv-Labels and metadata for heart beats collected from a clinical trial in hospitals using a digital stethoscope
audio files-Varying lengths, between 1 second and 30 seconds. (some have been clipped to reduce excessive noise and provide the salient fragment of the sound).

B. PCG Dataset of Heart Sound Classifier

In this work we have needed two sets of PCG datasets, first set labelled as different heart sounds and the second set labelled as cardiac diseases for the purpose of implement two classifier systems. The first type of dataset was available as two sets named A & B were generated through the PASCAL heart sound classification challenge. Dataset A contains the variable-length (varying from 1 to 30 seconds) sounds recorded through a digital stethoscope in a real-time situation having background noise. Dataset A was partitioned into four classes named normal, extra heart sound, murmur, and artifact, while dataset B was partitioned into three classes: normal, extra-systole, and murmur. The authors have merged both datasets into a single dataset consisting of all five classes in this work. The number of phonocardiogram signals in normal, murmur, artifact, extra-systole, and extrahls classes are 255, 114, 40, 37, and 16. Fig 2 shows the waveforms of the heart sound classes and their corresponding spectrograms.

1) Normal Category

In the Normal category there are normal, healthy heart sounds. These may contain noise in the final second of the recording as the device is removed from the body. They may contain a variety of background noises (from traffic to radios). They may also contain occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. A normal heart sound has a clear “lub dub, lub dub” pattern, with the time from “lub” to “dub” shorter than the time from “dub” to the next “lub” (when the heart rate is less than 140 beats per minute).

In medicine we call the lub sound "S1" and the dub sound "S2". Most normal heart rates at rest will be between about 60 and 100 beats (‘lubdub’s) per minute. However, note that since the data may have been collected from children or adult sincalmor excited states, the heart rates in the data may vary from 40 to 140 beats or higher per minute. Each of the classes are described below. The waveforms and corresponding spectrograms of each of the classes are shown in Fig 2.

2) Murmur Category

Heart murmurs sound as though there is a “whooshing, roaring, rumbling, or turbulent fluid” noise in one of two temporal locations: (1) between “lub” and “dub”, or (2) between “dub” and “lub”. They can be a symptom of many heart disorders, some serious. There will still be a “lub” and a “dub”.

3) Extra Heart Sound Category

Extra heart sounds can be identified because there is an additional sound, e.g. a “lub-lubdub” or a “lubdub-dub”. An extra heart sound may not be a sign of disease. However, in some situations it is an important sign of disease, which if detected early could help a person. The extra heart sound is important to be able to detect as it cannot be detected by ultrasound very well.

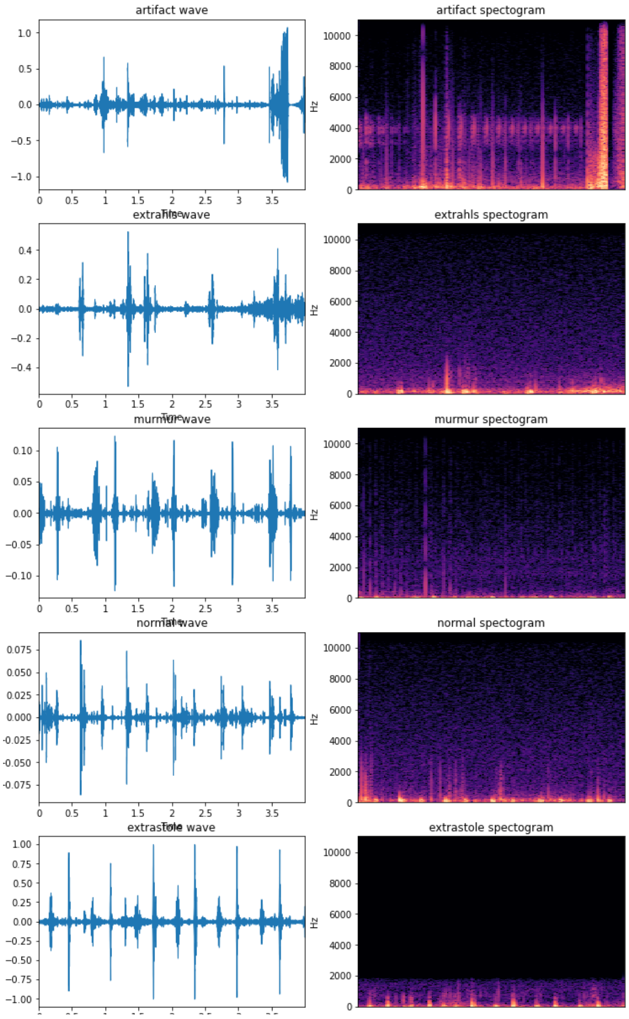


Figure 2: Waveforms and Corresponding Spectrograms of five classes of heart sounds

4) Extra systole Category

Extra systole sounds may appear occasionally and can be identified because there is a heart sound that is out of rhythm involving extra or skipped heart beats, e.g. a “lub-lub dub” or a “lub dub-dub”. (This is not the same as an extra heart sound as the event is not regularly occurring.) An extra systole may not be a sign of disease. It can happen normally in an adult and can be very common in children. However, in some situations extra systoles can because by heart diseases. If these diseases are detected earlier, then treatment is likely to be more effective.

5) Artifact Category

In the Artifact category there are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernable heart sounds, and thus little or no temporal periodicity at frequencies below 195Hz. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again.

C. PCG Dataset of Disease Classifier

The second set of data for cardio vascular disease classification was taken from PCG recordings used in the article [18]. The recordings were collected from various sources like books and websites and contained a total number of 1000 PCG recordings in. wav format in five different classes i.e., Normal (N), Aortic stenosis (AS), Mitral regurgitation (MR), Mitral stenosis (MS), Mitral valve prolapse (MVP). Each of the classes has 200 recordings for roughly 3s. All the recordings are sampled at 8 kHz. Since the lowest signal length present in this dataset is 1.125s, all the recordings are truncated from the start of the recording up to 1.125s. Figure 3 illustrates the waveform for each of the disease classes.

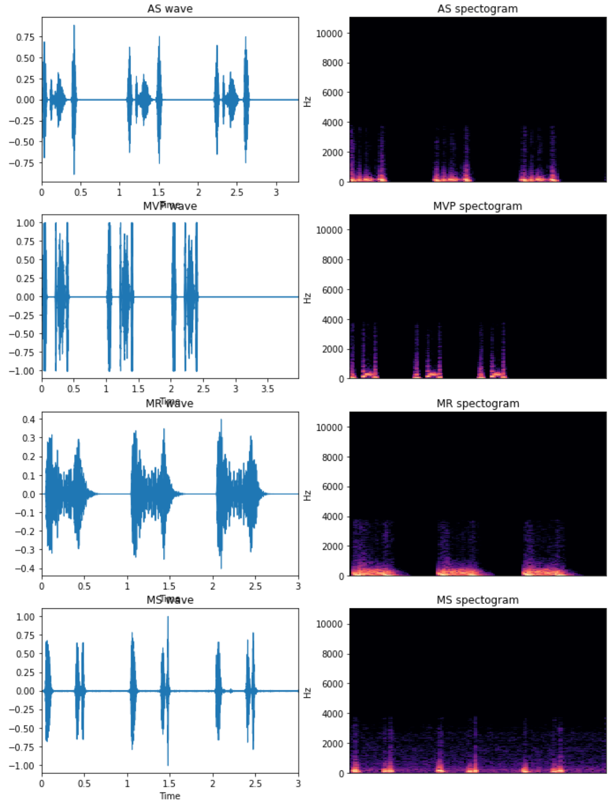


Figure 3: Waveforms of Disease Classes and Corresponding Spectrograms

2-D convolution works by applying the convolution filter on the input image. The filter passes over several pixels, which is called a stride. At each spatial location, the convolution between the part of the image and _lter is attained. The outcome is a 2-D array which is called a feature map. Softmax, Rectified Linear Unit (ReLU), Randomized Leaky ReLU, and other non-linear activation layers are used to pass this feature map. The pooling layer, also known as the subsampling layer, is another major component of ConvNet. Its purpose is to reduce the spatial size of the activation map to reduce the number of parameters needed for further processing. It applies to all feature maps on its own. the result of the last pooling layer is received by a fully connected layer and utilized to categorize images into labels. It is the component of ConvNet where discriminative learning is performed. It behaves like a multi-layer perceptron model which can learn weights & identify image classes.

The raw data provided is in Waveform Audio File Format, encoding phonocardiogram signals. To pass these sounds. waves to ConvNet model, these phonocardiogram signals are converted into an image, i.e., 2-D spectrogram. Spectrograms are convenient for representing these heartbeat recordings because they capture the intensity of the frequencies throughout a given sound. Thus, these spectrograms are effective representations of an audio recording. In this work, the authors have proposed the use MFCC and HCQT [17] based spectrograms for phonocardiogram signal classification. At last, the result of the last pooling layer is received by a fully connected layer and utilized to categorize images into labels.

D. Data Pre-Processing

The heart sounds recorded by digital stethoscope and the mobile phone microphone often has background noise. The pre-processing 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 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. Since the number of heartbeat signals in each class is very low, audio augmentation is performed over raw audio signals. I have applied noise injection, shifting time, varying pitch, and speed to generate augmented data for phonocardiogram signals. After audio augmentation, the number of phonocardiogram signals in normal, murmur, artifact, extra-systole, and extrahls classes are 2555, 1146, 400, 378, and 158, respectively. The augmented dataset is partitioned into training and testing datasets with an 80: 20 ratio.

E. 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. Time frequency domain transformation depicted in Fig 4.

The 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 (i.e. digital stethoscopes and mobile phone microphone) which results in varying amplitude ranges, converting the data to the frequency domain leads to more accurate results.

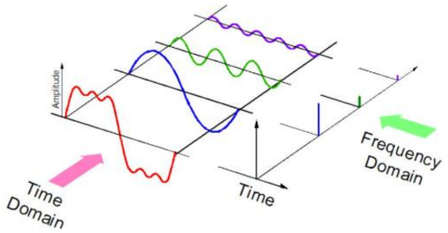


Figure 4: Time to Frequency domain Transformation

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. Spectrograms of five classes of heart sounds are illustrated in the Fig 2. Also, the Fig.3 shows the waveforms and corresponding spectrograms four classes of heart diseases. In this work, I have used MFCC (Mel Frequency Cepstral Components) and Hybrid-CQT (Hybrid Constant Q Transform) for feature extraction in frequency domain. More details about these feature extraction methods and ConvNet classifier model will explain in next chapter.

1) Mel Frequency Cepstral Coefficients (MFCC)

First, 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. Firstly Load metadata (set_a.csv) using pandas. Also extract features from each audio data in data_a folder. Then we're rooting MFFC sound feature using librosa.

The MFCC feature extraction technique principally includes windowing the signal, applying the DFT, taking the log of the magnitude, and also warping the frequentness on a Mel scale, followed by applying the inverse DCT. The reason for choosing MFFC feature is, it’s observed that rooting features from the audio signal and using it as input to the base model will produce much better performance than directly considering raw audio signal as input [13], [14].

The point birth process of MFCC is composed of the following way.

- Pre-emphasis It amplifies high frequentness by passing phonocardiogram signals from a high pass sludge.
- Framing Phonocardiogram signals are separated into overlapping frames. It's implemented to cost local spectral-properties.
- Windowing It's enforced on frames for the minimization of discontinuities around edges. An illustration of a extensively used fashion is Hamming windowing.
- Separate Fourier Transformation DFT is applied to the sound signal after the third step to gain the frequence sphere signal from the time sphere.
- Frequency Warping It's used to calculate the volume of energy that occurs in colorful locales of a frequence sphere. Melin this case is a pitch unit. A pitch of 1000 Mels is a pure tone at 1000Hzwitha40dBstrengthoverthelistener's threshold. Mel-scale is used to determine this non-linear frequency result. Then, the frequence term is denoted by f, while the Mel-scale frequency is denoted by M (f).
- Discrete Cosine Transform and Log Compression In this step, the logarithmic function IFFT is applied on, altered bank powers entered in step 5. The DCT follows it.

The visualization of MFCC series of five classes of heart sounds and four disease categories are illustrated in Fig 5 and Fig 6 respectively.

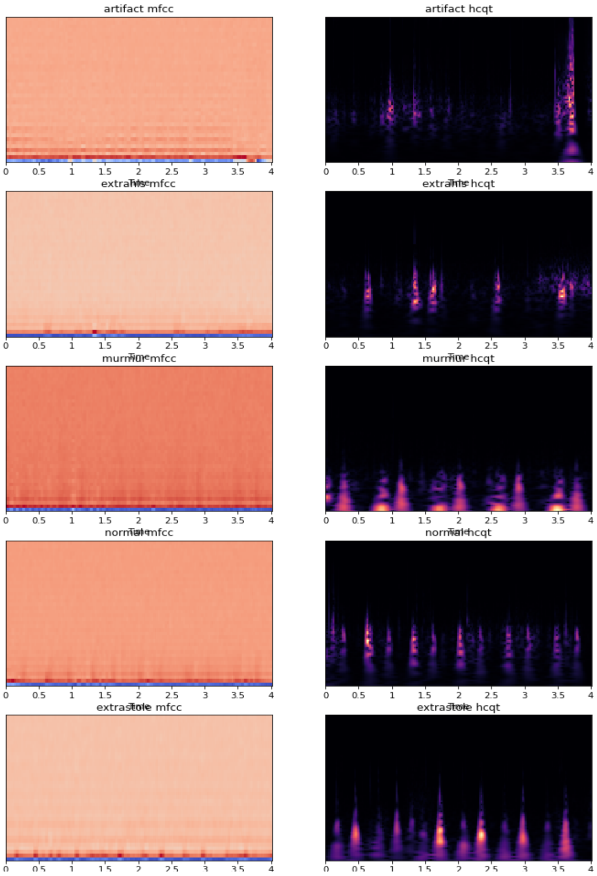


Figure 5: MFCC and HCQT Heatmap Visualisation of Five classes of Heart Sound.

2) Hybrid Constant-Q Transform (HCQT)

J. C. Brown, in 1988 has introduced CQT. It refers to a approach that transforms a signal from time to frequence sphere. Still, it's different from Fourier transformation as central frequencies are geometrically spaced, and corresponding Q-factors are equal. CQT is defined as a1/24octave, lter bank, but it isn't confined to 24 only; it can be varied to 12, 36, or 48 binsperoctave also. Unlike DFT, central frequencies of analysis aren't slightly distributed but aligned with inversely tempered scale notes; this makes CQT suitable for the processing of sound. Likewise, the frequence resolution of CQT has a constant Q-factor, which effectively improves resolution delicacy in low-frequence regions. Under the N-th frame of CQT, the frequence element of the K-th semitone can be stated in (1)

$$X_n^{cqt}(k) = \frac{1}{N} \sum_{m=0}^{N_k-1} x(m) w_{N_k}(m) e^{-j2\pi mQ/N_k} \qquad \text{……… (1)}$$

where Q is a constant whose value depends on the number of spectral lines of a single octave (β).

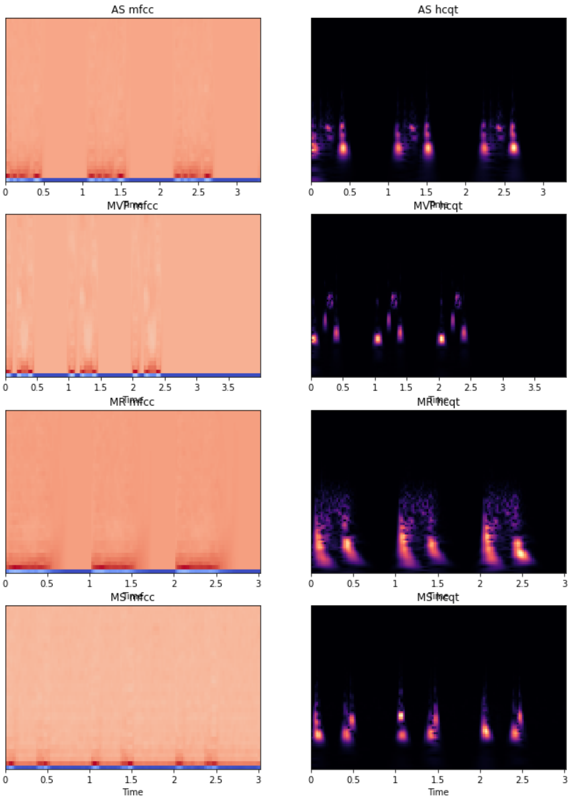


Figure 6: MFCC and HCQT Heat map Visualisation of Heart Disease Classes.

The capability of the constant-Q transform to give equal frequency support to all semitones and a variable number of bins among them is its main advantage. Still, it has downsides, one of which being the absence of harmonious temporal resolution at lower frequentness. This trade-off can be soothed by introducing variants of CQT i.e., VQT and HCQT. When compared to the CQT transformation, the VQT transformation provides better temporal resolution at lower frequencies [15], [16]. A new parameter is introduced to allow for an equitable drop of the bins' Q-factors as they approach low frequencies

High frequencies are those that exceed f_kc, whereas low frequencies are those that are less than f_kc. The high frequency section of hybrid CQT uses the filter bank of the high-frequency part of CQT to filter the short-term Fourier transform-based spectrogram. The regular CQT is used directly for the low-frequency section of HCQT. In compared to CQT, HCQT is more computationally capable. The process plow to get the information in frequency domain from time domain is illustrated in Fig 7.

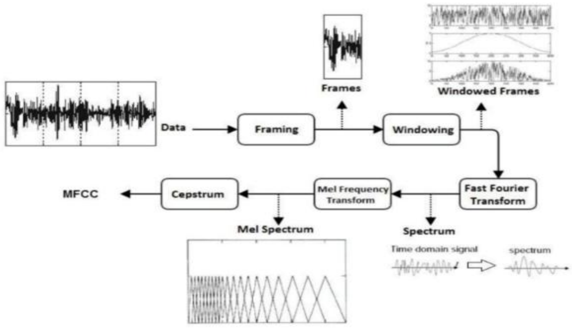


Figure 7: The Process flow to Get the Information in Frequency Domain from Time Domain

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

- 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.
- Compute the hybrid constant-Q transform of an audio signal. Here, the hybrid CQT uses the pseudo CQT for higher frequencies where the hop length is longer than half the filter length and the full CQT for lower frequencies.
- 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.

F. Classifier 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 architectures 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. Architecture of the classifier model of the proposed model is shown in Fig 8.

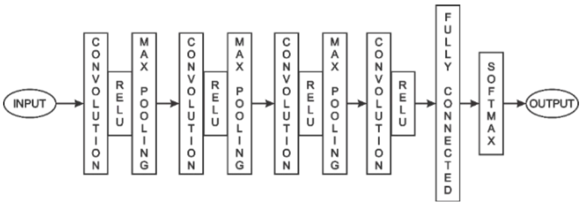


Figure 8: Architecture of the Classifier Model of the Proposed Method

CNN has brought the revolution in the domain of computer vision. It has remarkably achieved better results than the traditional classification algorithms. Deep learning is a sub-class of machine learning which is based on Deep Neural Networks (DNNs). Word deep indicates the presence of greater than such type of deep neural network, which is also known as the ConvNet model. It is made up of primarily three layers: a convolution layer, a pooling layer, and a dense layer (fully connected layer) [17], [18], [19]. The first layer i.e., the convolutional layer is an essential building block of ConvNet. This layer performs the mathematical operation convolution. In a continuous domain, the convolution of two functions f and g is given as in (2):

$$(f \ * \ g)(t) = \int_{-\infty}^{+\infty} f\,(\tau) g\,(t \ - \ \tau) d\tau$$
$$= \int_{-\infty}^{+\infty} f\,(\mathrm{t} - \tau) g\,(\tau) d\tau \ldots\ldots\ldots\ldots. \ (2)$$

In the discrete case, the same is expressed ain (3):

$$(f \ * \ g)(n) = \sum_{m=-\infty}^{\infty} f(m) g\,(n - m)\mathrm{s} \ \ldots\ldots\ldots \ (3)$$

2-D convolution for a digital image can be extended as in (4):

$$(f * g)(x, y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} f \begin{pmatrix} mx - n, yn \\ -m \end{pmatrix} g(n, m) \ \ (4)$$

The function g represents a filter that is applied to the input image f in this case. 2-D convolution works by applying the convolution filter on the input image. The filter passes over several pixels, which is called a stride. At each spatial location, the convolution between the part of the image and filter is attained.

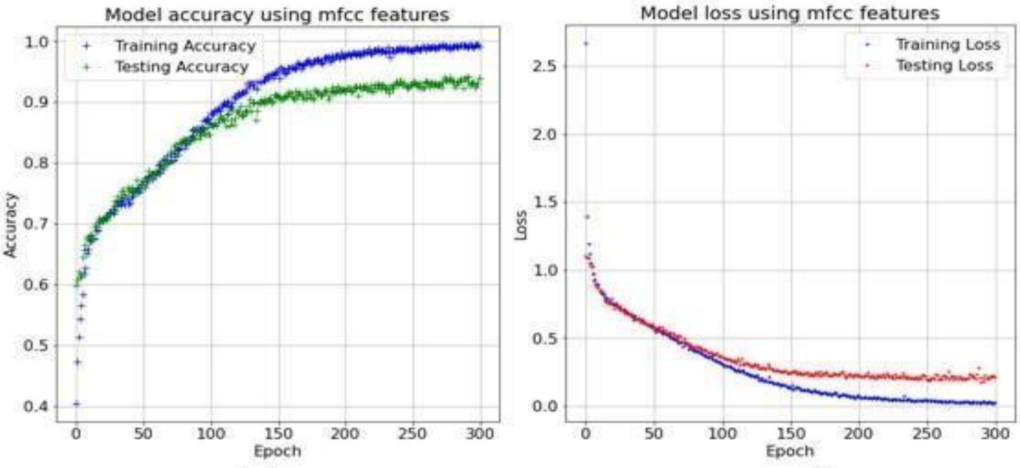
Architecture of the classifier model of the proposed method is depicted in the Fig 4.4. The outcome is a 2-D array which is called a feature map. Softmax, Rectified Linear Unit (ReLU), Randomized Leaky ReLU, and other non-linear activation layers are used to pass this feature map. The pooling layer, also known as the subsampling layer, is another major component of ConvNet. Its purpose is to reduce the spatial size of the activation map to reduce the number of parameters needed for further processing. It applies to all feature maps on its own. Max pooling is the most effective method for the implementation of pooling. At last, the result of the last pooling layer is received by a fully connected layer and utilized to categorize images into labels. It is the component of ConvNet where discriminative learning is performed. It behaves like a multi-layer perceptron model which can learn weights & identify image classes.

4. Results and Discussions

Classifier models on CNN based on MFCC and HCQT features are designed for the multi class classification of cardiac sounds. To build the proposed classifier models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like Tensor Flow. In addition, the Librosa library in Python is used for generating MFCC and HCQT spectrograms.

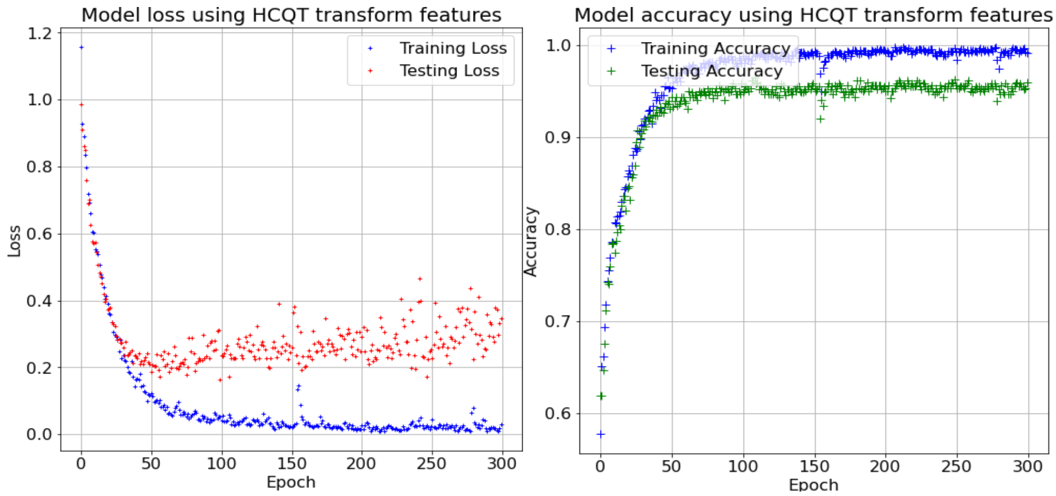
The dataset was divided into training, validation and testing parts with 70% of the PCG signal used for training the model and 10% of the signal used for validation and remaining 20% data used for testing. 10-fold cross validation was performed randomly on the dataset for generalization of the obtained results.

Two separate classifier models are designed in Convolution Neural network with two different features. MFCC and HCQT features are taken to get the two classifier models. The accuracy curves of the two classifiers are shown in Fig 9. and Fig 10. It is clear from the curve that classifier model with HCQT feature is much better than those of MFCC model.



(a) (b)

Figure 9 (A): Evolution of Classification Accuracy With The Training & Validation Image Datasets Throughout the Training of MFCC Model, (B): Evolution of Classification Loss with the training & Validation Image Dataset of Convnet-MFCCModel.



(a) (b)

Figure 10 (a): Accuracy with Training and Validation Image Datasets throughout the Training of HCQT Model. (b):

Evolution of Classification Loss with Training and Validation Image Datasets throughout the Training of HCQT Model.

To build the proposed ConvNet models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like TensorFlow. In addition, the Librosa library in Python is used for generating MFCC and HCQT spectrograms. ConvNet models used in this phonocardiogram signal classification model using these spectrograms have four convolutional layers. The first convolution layer has a size of 32-5×5, the second convolution layer has a size of 64-5 × 5, the third convolution layer has a size of 64-5 × 5, and the last layer has a size of 32-5 × 5. A subsampling layer using max-pooling follows the first two convolution layers. The size of these max-pooling layers is 2 × 2 with a stride of size 2 × 2. The final layer of the ConvNet model is a fully connected layer with a softmax non-linear activation function with five units. These five units in the last layer are essential for this five-class phonocardiogram signal classification problem. Fig. 9 and 10 shows the accuracy and loss curves for the train and test set during the training of ConvNet models.

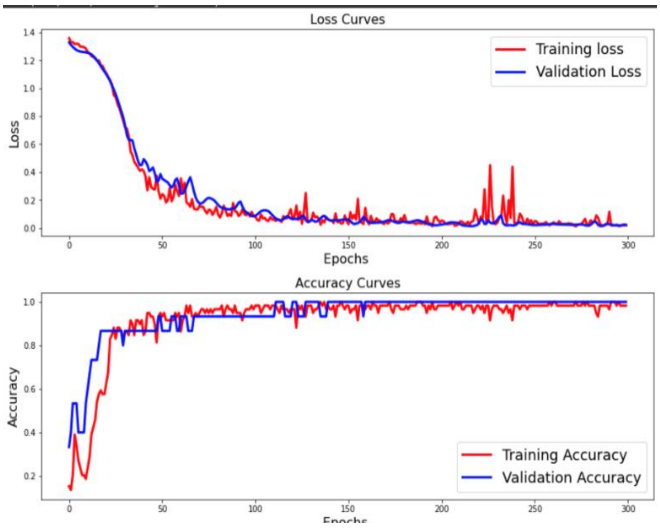


Figure 12: Accuracy-Loss Curve of Disease ClassifierModel

To build the proposed ConvNet models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like TensorFlow. In addition, the Librosa library in Python is used for generating MFCC and HCQT spectrograms. ConvNet models used in this phonocardiogram signal classification model using these spectrograms have four convolutional layers. The first convolution layer has a size of 32-5×5, the second convolution layer has a size of 64-5 × 5, the third convolution layer has a size of 64-5 × 5, and the last layer has a size of 32-5 × 5. A subsampling layer using max-pooling follows the first two convolution layers. The size of these max-pooling layers is 2 × 2 with a stride of size 2 × 2. The final layer of the ConvNet model is a fully connected layer with a softmax non-linear activation function with five units. These five units in the last layer are essential for this five-class phonocardiogram signal classification problem. Fig. 5.1 and 5.2 shows the accuracy and loss curves for the train and test set during the training of ConvNet models. The shape and dynamics of these learning curves are studied to diagnose the behavior of a ConvNet model. Three common dynamics observed in these learning curves are under-fitting, over fitting, and optimal fitting. From these plots, it can be verified that the ConvNet-HCQT model has offered optimal fit in comparison to other models. Accuracy-loss curve of the disease classifier model is depicted in Fig. 12. Classification report of disease classifier model is illustrated in Table 2. which shows that the model shows 99% of average accuracy. Precision and recall of all disease classes are high. This shows the best performance of the proposed disease classifier system.

The comparison of the performance heart sound classifier models is shown in the table 5.1.

.

Table 2: Classification Report of Disease Classifier

Class/Metric	Precision	Recall	F1 Score
AS	1	0.99	0.99
MR	0.92	1	0.96
MS	0.99	0.99	0.99
MVP	1	1	1
Accuracy	0.99		
Macro average	0.98	0.99	0.98
Weighted average	0.99	0.99	0.99

Table 3: Comparison of Performance the Heart Sound Classifier Models

	MFCC-CNN	HCQT-CNN
Accuracy	.93	.96
Sensitivity	.93	.95
Precision	.96	.98

5. Conclusions and Future Scopes

Diagnose at an early stage is the only way to decrease the mortality rate occurring due to CVD. However, due to a lack of awareness for routine health checkups and unavailability of all resources at low cost, there are major hurdles in the early diagnosis of CVD. The situation worsens in developing countries where population density is high, and a doctor is not available in remote locations. To target these issues, the authors have offered a design of a decision support system that utilizes the PCG signals for the early diagnosis of CVD. PCG signals can be captured by a small, low-cost handheld device called a stethoscope.

In this work, a multi-class phonocardiogram signal database with five classes, namely, extra heart sound, artifact, extra-systole, normal, and murmur heartbeat, are used to design the phonocardiogram signal, classification model. The model we are using with five classes of cardiovascular diseases also. In this work, a CNN classifier network has been proposed with two efficient acoustic features MFCC and HCQT. To build the proposed CNN models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like TensorFlow. In addition, the Librosalibraryin Pythonis used for generating MFCC and HCQT spectrograms. And the proposed work achieved an accuracy of96% with the heart sound datasets and 99% accuracy with the disease classifier model.

In future, more features can ensemble to get discriminative stacked features and there by the performance of the system may increase. Also, ECG signals can use with PCG signals using these acoustic features to get a multi-modality model.

PCG signals can be captured by a small, low-cost handheld device called a stethoscope. In this work, a multi-class phonocardiogram signal database with _ve classes, namely, extra heart sound, artifact, extra-systole, normal, and murmur heartbeat, are used to design the phonocardiogram signal, classification model. In this work, a CNN classifier network has been proposed with two efficient acoustic features MFCC and HCQT. To build the proposed CNN models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like Tensor Flow. In addition, the Librosalibraryin Python is used for generating MFCC and HCQT spectrograms. And the proposed work achieved an accuracy of 98%. The model we are using with five classes of cardiovascular diseases also.

References

[1] (Jun.2021). CVDDataasCitedon27th. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Jain, S. Tiwari, and V. Sapra, ``Two-phase heart disease diagnosis system using deep learning, '' *Int. J. Control Autom.*, vol. 12, no. 5, pp. 558_573, 2019.

[3] K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, ``Algorithms for automatic analysis and classi_cation of heart sounds_A systematic review, '' *IEEE Access*, vol. 7, pp. 8316_8345, 2019.

[4] P. Rani, R. Kumar, N. M. S. Ahmed, and A. Jain, ``A decision support system for heart disease prediction based upon machine learning, '' *J. Reliable Intell. Environ*, pp. 1_13, Jan.2021.

[5] E. F. Malik, S. Barin, and M. E. Yuksel, "Accurate classification of heart sound signals for cardiovascular disease diagnosis by wavelet analysis and convolutional neural network: Preliminary results, '' in *Proc. 28th Signal Process. Commun. Appl. Conf. (SIU)*, Oct. 2020, pp. 1_4.

[6] G. Y. Sonand S. Kwon, ``Classification of heart sound signal using multiple features, '' *Appl. Sci.*, vol. 8, no. 12, pp. 2344_2358, 2018.

[7] S. Patidar, R. B. Pachori, and N. Garg, "Automatic diagnosis of septal defects based on tunable-Q wavelet transform of cardiac sound signals, '' *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3315_3326, 2015.

[8] Gharehbaghi, A. A. Sepehri, A. Kocharian, and M. Linden, ``An intelligent method for discrimination between aortic and pulmonary stenosis using phonocardiogram, '' in *Proc. WorldCongr. Med. Phys. Biomed. Eng.*, Toronto, ON, Canada, 2015, pp.100_1013.

[9] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, and A. K. Sangaiah, ``Diagnosis of heart diseases by a secure Internet of Health Things system based on autoencoder deep neural network, '' *Comput. Commun.*, vol. 162, pp. 31_50, Oct.2020.

[10] M. Banerjee and S. Majhi, ``Multi-class heart sounds classification using 2D-convolutional neural network, '' in *Proc. 5th Int. Conf. Comput., Com-mun. Secur. (ICCCS)*, Oct. 2020, pp.1_6. G. Redlarski, D. Gradolewski, and A. Palkowski, ``A system for heartsounds