CLASSIFICATION OF HEART SOUND SIGNALS FOR CARDIAC DISEASE ANALYSIS USING ONE DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK

V. Anantha Natarajan[1*], M. Sunil Kumar[2], Naresh Tangudu[3], Suneetha Konduru[4]

[1]Professor, Department of Computer Science and Engineering,
School of Computing, Mohan Babu University,
(erstwhile Sree Vidyanikethan Engineering College), Tirupati, AP, India.
*Correspondence: vananthanatarajan@vidyanikethan.edu
ORCID: 0000-0002-1577-0708

[2]Professor & Programme Head, Department of Computer Science and Engineering,
School of Computing, Mohan Babu University,
(erstwhile Sree Vidyanikethan Engineering College), Tirupati, AP, India.
sunilmalchi1@gmail.com
ORCID: 0000-0002-1439-2573

[3]Assistant Professor, Department of Information Technology,
Aditya Institute of Technology and Management, Tekkali, K Kotturu, Andhra Pradesh. India.
itsajs@gmail.com
ORCID: 0000-0001-8845-7491

[4]Professor, School of Computer Science & IT, Jain (Deemed-to-be University),
Bangalore, India
keerthisuni.k@gmail.com

ABSTRACT

Cardiovascular disease increasing deaths worldwide also it is affecting young age people at their early stage. Heartbeat analysis of a person can be normal or abnormal heart sounds which can be detected only through trained physician. To reduce the dependency on trained physicians for heart sound detection, proposed system focuses on automatic classification of PCG (Phonocardiogram) signals after removal of noise using Convolution Neural Network. Thus an automated machine/ artificial intelligence based preliminary analysis eliminates the need of trained physicians for detection of abnormalities in heart sounds. This paper proposes noise removal using spectral gating, Deep Learning based classification and energy based segmentation using to detect cardiac disease with efficient accuracy, specificity, and sensitivity. The proposed system becomes a support system for the physicians in automatic classification and diagnosis of the cardiovascular disease fast and efficient.

Keywords: phonocardiogram, Convolution Neural Network, machine learning, noise removal, spectral gating, deep learning, cardiac disease.

1. INTRODUCTION

Heart of a human being consists of a pump with four chambers; two atrial chambers used for collecting the blood from the veins and two ventricular chambers to pump the purified blood in to arteries. The functioning of the heart is regulated by an electrical conduct module. The electrical signals originate in the right atrial chamber and are transferred to the AV-node and then to the ventricles. The electrical action excites the muscle cells and drives the mechanical functioning of the chambers of the cardiac system. During the systolic period of the cardiac cycle the ventricles contract and it is followed by a filling phase termed as diastole. Following mechanical activities are part of the cardiac system functioning namely blood flow, vibrations stimulated by the walls of the cardiac chambers and opening and closing of the valves present in the cardiac system. The systolic period further subdivided in to atrial systole, isovolmic contraction, and ejection period. The diastole period is further sub-divided in to diastasis, isovolmic relaxation, rapid filling.

1.1 Cardiac System Monitoring

Advancements in cardiac system monitoring methods, tools, and techniques have reduced the volume of death tolls due to heart attacks and failure of cardiac system. The waveform representation of the cardiac system functioning using electrocardiogram abbreviated as ECG, pulse rate of the peripheral blood flow, and phonocardiogram abbreviated as PCG has greater potential in monitoring the cardiac system functionality. Phonocardiography is the study of the recordings of the sound generated by the cardiac system and PCG is a visual representation of high fidelity recordings of the same. The machine used for sound recording is called phonocardiograph. The waveforms are consistent between two consecutive cardiac cycles. Also the time duration between two different signals was constant. The opening and closure of the heart valves creates vibrations which results in heart sounds. The sound denoted as S1 is generated at the initial systole period when tricuspid and mitral valves close. The second sound (S2) is generated at the end of the systolic period during which the aortic and pulmonary valves close. In addition the phonocardiography supports the physician to detect and record sub-audible sounds and murmurings of the heart. The stethoscope used by the physician during the physical examination cannot detect such sub-audible sounds and murmurs from the heart. The quantitated heart sounds contains necessary information regarding the effects of drugs on the cardiac system and it is an effective assessment for monitoring the progress of disease on the system. Initially for disease diagnosis or abnormality detection in cardiac system as a pre-processing activity the heart sounds are segmented in to four parts namely S1 considered as the Ist heart sound, systole, s2 the second heart sound and diastole. S2 and S3 are rare sounds that are more generally available in the heart sounds of children. The S3 appears in between S2 and S1 and the S4 appears 100ms ahead of the S1 segment.

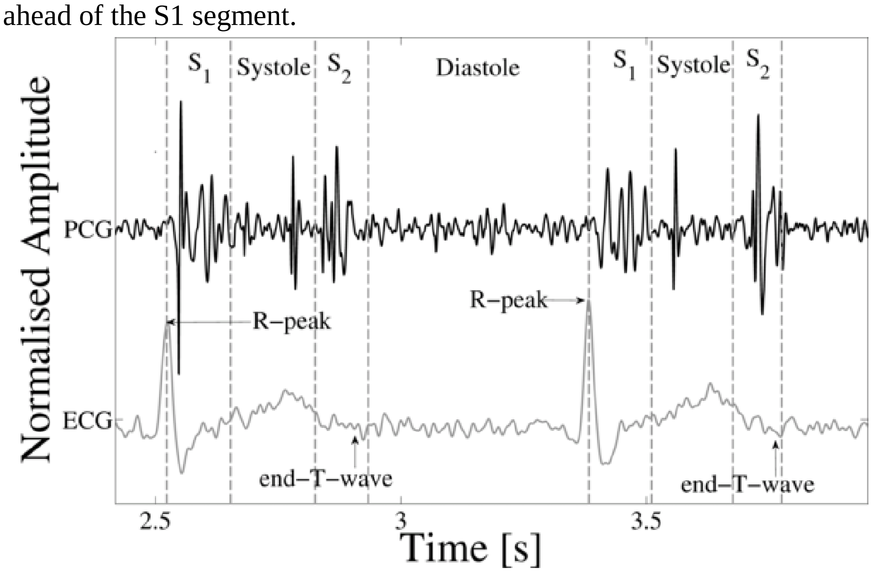

Fig. 1 Schematic view of PCG heart sound signal

[source: https://archive.physionet.org/challenge/2016/]

An arrhythmia is a problem with the heart's rhythm or rate. In case of an arrhythmia, the cardiac system may function fast/ slow, or in an irregular pattern. A state in which the cardiac system functions excessively fast is known as tachycardia. A state in which the cardiac system is beating less efficiently is known as bradycardia. Arrhythmia is caused by modifications in heart tissue and function, as well as the electrical pulses that controls the cardiac function. These changes can occur as a result of disease, damage, or heredity. Some people have an unpredictable heartbeat, despite the fact that there are typically no symptoms. Patient may feel dizzy, bewildered, or have difficulty breathing. Conventional diagnosis method used for cardiac abnormality is ECG and this research paper focus on designing an automated mechanism for detecting abnormality in the cardiac signal using deep learning technique from the heart sounds.

Because ECG signals are non-stationary and nonlinear, transitory illness signs might arise at any moment on the time scale. Hence few approaches of used PCG for automated detection of abnormalities in the cardiac system. [1] suggested a method for automatically detecting cardiac valve abnormalities such as aortic and mitral stenosis abbreviated as (AS) and (MS), and mitral regurgitation (MR) using the captured heart sound. The matrix form of the information estimated from both the frequency and time domain is extracted from the identified cycles of the PCG signal using the wavelet synchrosqueezing transform. The amplitude and spectral features are derived from the time-frequency matrix. The random forest (RF) algorithm method was adopted for classification. The phonocardiogram (PCG) might be a useful tool for examining and identifying arrhythmic heartbeats. The electrical activity of the heart, as well as its valve motions, is directly connected to cardiac rhythm. The data for this article was collected from the Physio-net Challenge. This article is intended to investigate the use of PCG in cardiac diagnostics [28-32].

2. RELATED WORK

In majority of the approaches the heart sound classification is carried out using the machine learning models based on the spatial and spectral/ frequency features extracted. More recently, the authors in [5] segmented data into distinct cardiac phases, and resampled the segments to 1000 Hz to match the criteria of the developed framework, which requires each specific cardiac cycle as input. In another approach [3], the authors have used un-segmented PCG for classification. Other approaches for segmentation have been utilised, such as a Hidden Markov Model (HMM) and its variant architecture termed as HSMM – Hidden Semi-Markov Model [6]. The efficiency of the pre-trained 2D CNN models were studied in [4] which utilizes the image representations of the sound signals for classification[19][21].

In [7], Pedrosa segmented the normalised data into segments with duration of 1.5s, ensuring that each fragment had a cycle of the cardiac functioning, and then discarded excessive magnitude noise in portions[20]. The autocorrelation function (ACF) is used to determine the residual cyclic components, and the length of systole is calculated based on the ACF peaks. Last, heart sounds are categorised as S1 or S2 based on systole length. Following the segmentation step, features are retrieved from the fragmented PCG waveform, and a classifier is used. For PCG signal analysis, many scholars have designed different approaches such as time domain and statistical-frequency domain [8].

For anomaly identification, researchers and signal processing specialists generally utilise machine learning-based classifier model [9-11]. Researchers in [9,10,12] used the support vector machine (SVM) algorithm for diagnosing the cardiac sound signals using computers/ machine in an automated way. In various research contributions the authors have used deep neural architectures such as convolutional neural networks (CNN), recurrent neural architectures - LSTM, and feed forward multi-layer artificial neural networks (FF-ML-ANN) in addition to traditional machine learning techniques for diagnosis [13, 11]. Aside from the fact that SVM [10] and LSTM [14] have higher accuracy in classification (78.64 percent and 74.9 percent, respectively), experts have conducted necessary research using various characteristics and classification algorithms to improve accuracy[15].

With the prevalence of abnormalities and diseases of the cardiovascular system throughout the world, building machine intelligence for automated diagnosis has become a top priority and, as a result, a hot research topic for bio-medical engineering and digital signal processing. Experts are working to develop effective and dependable methodologies to aid physicians and other healthcare practitioners in their diagnosis and treatment planning[16][18]. For increasing the accuracy of the classification, this work proposes automated classification of heart sounds utilising several classification methods and variable characteristics. PCGs are analysed and several classification models are trained using a publically available dataset including over 3000 sound recordings of the heart functioning from the PhysioNet Challenge hosted in 2016 related to the field of computational science in cardiology. In this work, novel classifier techniques and parameters are studied, and fascinating findings are achieved, which are contrasted to other existing techniques using the standard heart sound recordings database[21-27].

3. DATASET

The audio files range in duration from 1 second to 30 seconds, with some being clipped to remove excessive noise and offer the most important portion of the sound. The low frequency components of heart sounds carry the majority of the information, whereas the higher frequency components include noise. PhysioNet Challenge data was collected in real-world circumstances and typically incorporates various types of background noise. Differences in heart sounds that correlate to different cardiac ailments might be exceedingly subtle and difficult to distinguish. Classifying this type of data successfully necessitates the use of incredibly robust classifiers. Despite its medicinal importance, this is a machine learning application that has yet to be fully investigated.

4. PRE_PROCESSING

Acquisition and pre-processing of PCG signals are critical activities in the process of diagnosing heart diseases and other applications. De-noising of the cardiac sound signal, identifying the boundaries of the I^{st} and II^{nd} heart sounds (S1, S2), and other components of the cardiac sound from the PCG, and classification are among the heart sound pre-processing techniques. The moving average filter becomes a prime choice in order to decrease random noise thus yielding a sharp step response, which is a typical task and hence it is considered as the best filter for analysis of signals in time domain.

A high-frequency signal is used to describe a general cardiac sound. Specific aspects, such as S1 and S2, are determined by the envelope value. The envelope identified by the Hilbert transform comprises numerous minor variations that might have an influence on particular feature extraction based on the inherent properties of heart sounds. In feature extraction, these oscillations are also considered high frequency noise, thus a moving-average filter (MAF) was utilised for envelope smoothing considering it as kind of low pass filter. Duration of the MAF was 0.40 seconds (cut-off frequency was 3 Hz).

Initially the heart sound is normalized as per the following mathematical expression;

$x_{norm}(t) = \left(\frac{x(t)}{max|x(t)|}\right)^2$ where $x(t)$ is the sound signal and considering the square of the signal weakens the noise components and makes the signal peaks more prominent. The signal envelope can be detected by squaring the signal which makes the S1 and S2 peak detection accurate and simple. After normalization of the signal, the sound signal was decomposed into multiple frequency bands using empirical wavelet transforms.

4.1 Recurssive Moving Average Filter

Because of its ease of understanding and application, the running sum or the moving average (MA) is the frequently used filter in DSP. Despite its less complexity, the MA filter is appropriate for a frequent activities: lowering randomly generated noise while keeping a clear step response. As a consequence, it's the finest signal filter which process the given input signals in the time domain. The MA, on the other hand, is the poorest filter for signals that are represented in the frequency domain, having poor ability in separating signals of one frequency from the other. The moving average filter is related to the Gaussian, Blackman, and multiple pass moving average filters. These have enhanced frequency domain performance by a little margin at the cost of more processing. A moving average filter generates the output by taking average of a sequence of points from the given input. It may be stated numerically as:

$$y[i] = \frac{1}{M}\sum_{j=0}^{M-1} x[i+j] \qquad \text{Eq. 4.1}$$

$$y[i] = y[i-1] + x[i+p] - x[i-q] \quad \left[p = \frac{(M-1)}{2}; q = p+1\right] \qquad \text{Eq. 4.2}$$

where x denotes the given input, the filter output is represented as y, and M represents the length of the moving window. This equation calculates each point in the output using two sources of data: values from the current input window and previously estimated values from the output. A recursive equation is one in which the output of one estimation is employed in subsequent estimations. The term "recursive" has a variety of connotations, particularly in the field of computer science. The recursive MA filter differs from other filters that follow recursive procedure in several ways. Majority of recursive type of filters employs an infinitely long impulse response (IIR) made up of exponentials and sinusoids. The moving average's response is in a rectangular shape.

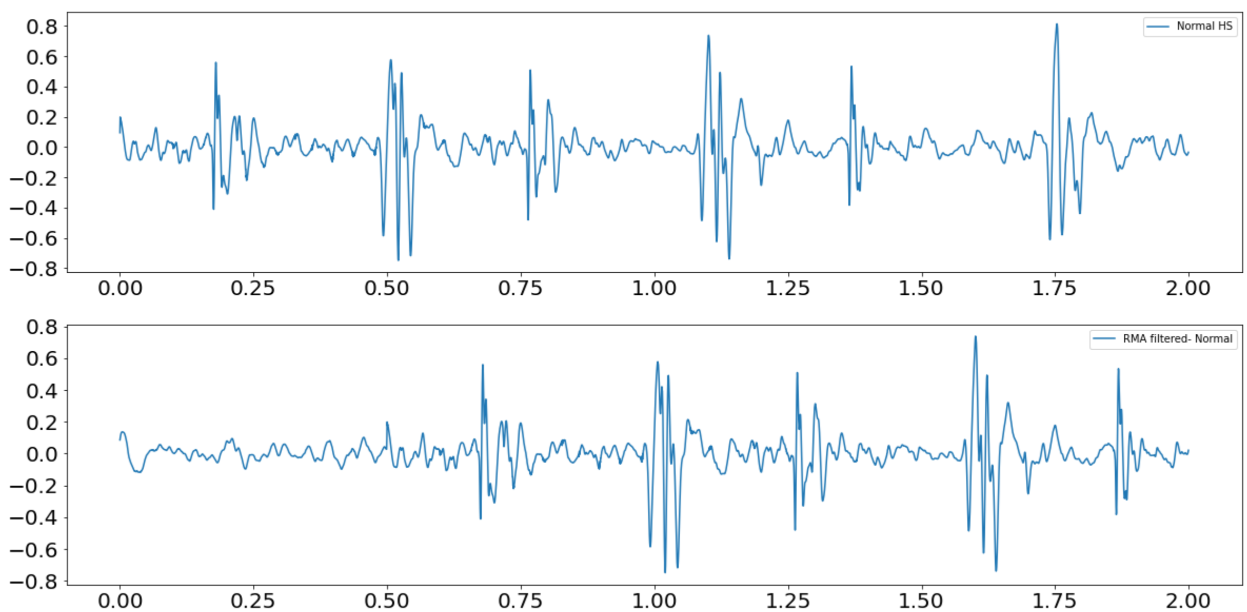

Fig. 2. Normal (Raw) PCG signal and RMA filtered PCG signal waveform

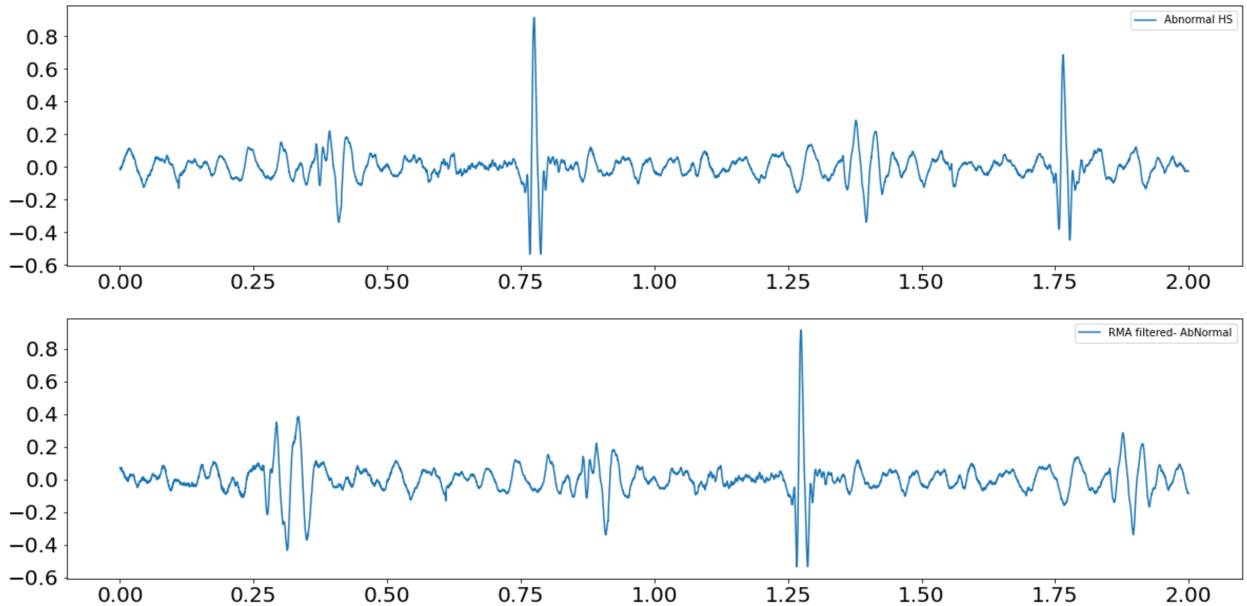

Fig. 3. Abnormal (Raw) PCG signal and RMA filtered PCG signal waveform

4.2 Noise Reduction using Spectral Gating

In general, background noise removal techniques for audio can be categorized as stationary or non-stationary noise reduction technique based on the frequency and duration of the noise present in the signal. The background noise removal technique used in this study is based on a technique known as "spectral gating," which is a type of Noise Gate. It works by calculating a spectrogram of a signal and, optionally, a noise signal, and then measuring a noise cut-off point for each frequency range of that signal or noise. This threshold is used to generate a mask that gates noise underneath the frequency varying threshold. In a non-stationary noise reduction method the noise cut-off frequency varies over time.

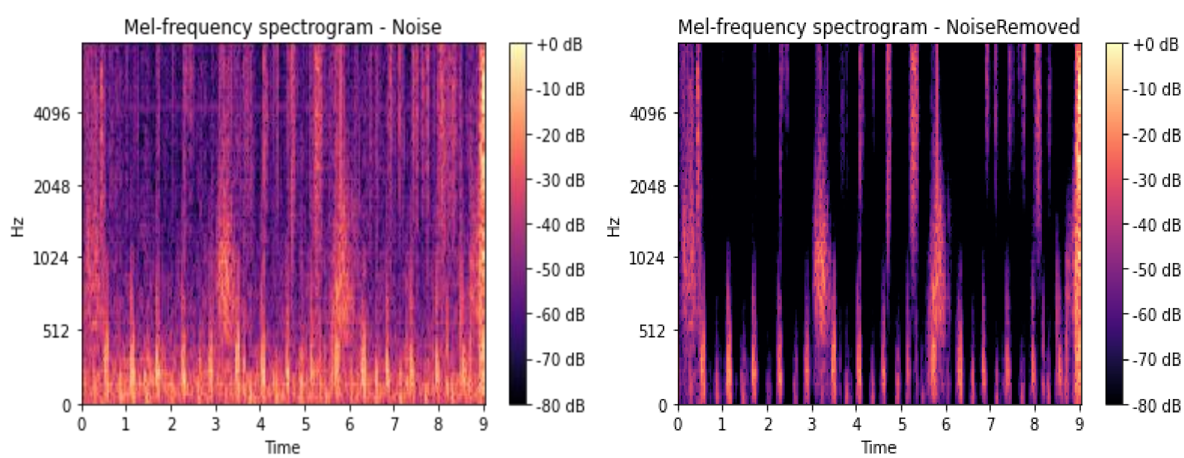

Fig. 4. Spectrogram of Noisy PCG signal and Noise removed PCG signal

Step-by-step procedure for noise reduction:

1. Over the signal, a spectrogram is computed.
2. An IIR filter applied forward and backward on each frequency band is used to generate a time-smoothed copy of the spectrogram.
3. Focusing on that time-smoothed spectrogram, a mask is derived.

4. A filter is used to smooth the mask over time and frequency.
5. The mask is applied to the signal's spectrogram snd h.

4.3 Segmentation using Moving Average

The envelope of the heart sound is extracted using the moving average algorithm as follows:

$$E_{FFT(n)} = \frac{1}{L_{F+} + n + 1} \sum_{k=0}^{n+L_F} FFT_H(k)\,, 0 \leq n \leq L_F - 1; n = 0,1, \ldots., N-1 \qquad \text{Eq. 4.3}$$

where N is the count of samples and L_f is the count of neighboring points. In our experiments the value of L_f is set to 6. The primary peak will be located in the region having lower frequency and greatest amplitude, while the secondary peak is at a comparatively at a high frequency region with a considerably lesser magnitude. Thus, the majority of the energy in cardiac sounds is contained inside the primary peak's frequencies, and eliminating out the side peak has little effect on the structure of heart sounds. As a result, the information about heart sound components is preserved by the major peak. The primary peak's terminating point is found between 20 and 200 Hz and is the minimum value in-between two consecutive spikes following the peak (primary) with an FFT value less than 0.2. The threshold of 0.2 was arrived at by trial and error. The cut-off frequency is fixed as 200 Hz if there is no such terminating point between 20 and 200 Hz.

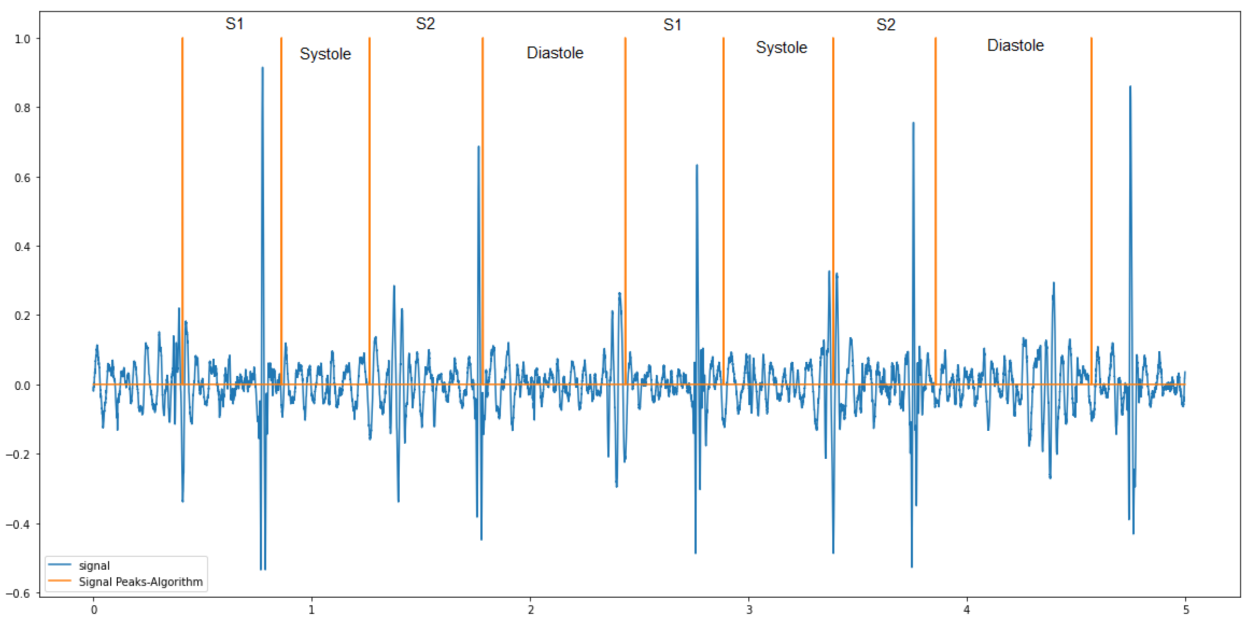

Fig. 5 Segmented PCG signalusing Running-Sum (Moving average) algorithm

5. CLASSIFICATION USING 1D-CNN

A typical heart sound or the digital version termed as phonocardiogram (PCG) contains two main parts namely S1 and S2 (Ist and 2nd heart sound). Due to few abnormalities or cardiac diseases murmur noise will be present as part of the S1 or S2 signal. But the murmur which arises within end of S1 and beginning of S2 is common. For certain diseases an interval within the single cycle of a cardiac sound does contain the relevant information regarding the disease. Thus fragmentation or segmentation of the cardiac sound in to various segments/ components becomes essential. The boundary of the s1 and s2 components of the heart sound becomes unclear because of the presence of murmur noise. Thus it makes the inference from the auscultation complex. The main objective this study is to analyze the cardiac sound to identify the PCG sound signal boundaries, extract relevant features using the convolutional layers and classify using a fully connected neural network.

When compared to the conventional artificial neural networks the CNN has the advantage of combining the feature extraction and the classification activities in to a combined architecture. IN case of traditional machine learning algorithms the input are pre-process before extracting the relevant features and finally the classification is performed. This process makes this approach slightly inefficient and computationally expensive. The CNN based classification approach can extract the relevant and optimal feature set from the given input as part of the training process thus making the classification task more accurate. Traditional ML approaches were employed usually in case of 1D signal as the CNN or deep learning architectures were designed exclusively for processing 2D signals.

The common strategy for heart disease detection and arrhythmia identification is to generate a power- or log-spectrogram to transform each wave form to a 2D graphic [18, 19]. Using such deep convolution based networks, however, has some limitations and disadvantages. It is obvious that they require more computational efforts, demanding the use of dedicated equipment, particularly for training. As a result, designing applications that are running in real time environment inside a low power mobile device are not possible with 2D CNNs. Furthermore, proper deep CNN training necessitates a large dataset for training in order to acquire reasonable generalization ability. For numerous 1D signal processing where availability of labelled data is scarce, this may not be a viable solution.

1D-CNN architecture can be considered as a variant of the 2D-CNN and few literatures have proved the efficiency of the 1D-CNN architecture in text and speech analytics stating the following reasons:

Computational complexity – There is a great difference between 1D and 2D architecture of CNN in terms of the computational complexity; for a given input image of size NxN and convolution kernel size KxK the computational complexity is $O(N^2K^2)$ for 2D-CNN based processing whereas the 1D CNN's computational complexity for the same dimension of input NxK is $O(N*K)$. This shows that for the same topology and model hyper parameters the computational cost required for 1D-CNN based processing is lesser than that of 2D-CNN. Due to the minimal effort required in processing or analysis the 1D CNNs becomes the most promising architecture for real time and mobile phone/ hand held devices application.

On applications where labelled data is available in less numbers and signal fluctuations are more or if they are gathered from various sources, 1D CNNs have exhibited improved performance. In 1D CNNs, two separate layer types are suggested; "CNN-layers," which contain 1D convolutional layer, activation/transfer function, and subsampling (pooling), and "MLP-layers," which are similar to the layers of a normal Multi layered feed forward architecture and so named "MLP-layers." The following model parameters are responsible for the optimal functioning of a 1D-CNN; the count of convolution and dense layers/neurons that are concealed, Filter (kernel) size of every convolution layer, sub-sampling factor within every convolution layer, and pooling and transfer functions.

In 1D-CNN the layer which receives input is a receptive layer that accepts the un-processed audio signals, and the layer which yields the output is an MLP layer that contains neurons depending upon the number of classes in the training dataset. Usually the 1D filter kernel has a size of 3 or 5 or 7 and a sub-sampling factor of 2 where the k^{th} neuron present l^{th} hidden layer executes a series of convolution, and its total is passed via the transfer function, *f*, and then the sub-sampling process. This is the primary distinction among 1D and 2D CNNs, with 1D array replacing 2D matrix in both convolution filter and output.

The convolution layers process the raw input data and gives a compact representation of the same that are in turn fed to the dense layers for classification. Thus the extraction of feature and the classification are grouped in to a combined process which is repeated until the performance of the classifier is found to be satisfying. This makes the 1D-CNN computationally less expensive as the 1D convolutions can be considered as the linear weighted sums of 1D raw input and the 1D convolution kernel. The linear weighted sum operation can be parallelized during both the forward and backward propagation [20].

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} conv1D\left(w_{ik}^{l-1}, s_i^{l-1}\right) \qquad \text{Eq. 5.1}$$

where x_k^l is the raw input, and b_k^l indicates the k^{th} neuron's bias which at the l^{th} layer, s_i^{l-1} represents the i^{th} neural output at layer *l-1*, w_{ik}^{l-1}denotes the kernel used at layer $l-1$ of the i^{th} neuron. During implementation padding was not done on top of the output from the 1D- convolution operation. The output of a l^{th} layer neuron can be expressed as $y_k^l = f\left(x_k^l\right)$; $s_k^l = y_k^l \downarrow ss$ (s_k^l denotes the output of the neuron present at the l^{th} layer at the k^{th} position, and $\downarrow ss$ denote the down sampling process at a scaling rate of *ss*.

The process of back-propagation starts at the output layer and while considering $l = 1$ for input layer, $l = L$ for the output layer, N_L denotes the available classes in the dataset. When the input represented as p, and the respective actual expected $t^p$with output vectors $[y_1^L, \ldots\ldots, y_{N_L}^L]$ the cross entropy error can be represented as

$$-\sum_{c=1}^{N} t_c^p \log(y_c) \qquad \text{Eq. 5.2}$$

Using the estimated delta error the derivative of the error for each weight/ bias can be computed. To update the weights/bias of neurons in a layer the chain rule can be applied,

$$\frac{\partial E}{\partial w_{ik}^{l-1}} = \Delta_k^l y_i^{l-1} \qquad \text{Eq. 5.3}$$

$$\frac{\partial E}{\partial b_k^l} = \Delta_k^l \qquad \text{Eq. 5.4}$$

The backpropagation of error between first layer of the dense network to the final layer of the convolution network can be expressed mathematically as follows;

$$\frac{\partial E}{\partial s_k^l} = \Delta s_k^l = \sum_{i=1}^{N_{l+1}} \frac{\partial E}{\partial x_i^{l+1}} \frac{\partial x_i^{l+1}}{\partial s_k^l} = \sum_{i=1}^{N_{l+1}} \Delta_i^{l+1} w_{ki}^l \qquad \text{Eq. 5.5}$$

Then the sensitivities of the parameters can be expressed as follows;

$$\frac{\partial E}{\partial w_{ik}^l} = conv1D(s_k^l, \Delta_i^{l+1}) \qquad \text{Eq. 5.6}$$

$$\frac{\partial E}{\partial b_k^l} = \sum_n \Delta_k^l\,(n) \qquad \text{Eq. 5.7}$$

Using the computed bias and weight sensitivities the bias and weights are estimated with respect to the learning rate α

$$w_{ik}^{l-1}(t+1) = w_{ik}^{l-1}(t) - \varepsilon \frac{\partial E}{\partial w_{ik}^{l-1}} \quad \text{Eq. 5.8}$$

$$b_k^l(t+1) = b_k^l(t) - \varepsilon \frac{\partial E}{\partial b_k^l} \quad \text{Eq. 5.9}$$

The sensitivity of the output of a neuron k at lth layer of the convolution network is derived by back-propagation of gradient of errors (Δ_i^{l+1}) estimated at the $l+1^{th}$ layer. The overall iterative procedure of the back-propagation based training can be enumerated as follows;

1. Initial values of network weights and bias are chosen arbitrarily (a random value in the order of $1E^{-4}$ to $1E^{-3}$)
2. For every iteration do the following steps;
   a. For every sample of the traiing data;
      i. Forward Propagation: At each layer, estimate the output of each neuron denoted as $s_i^l, \forall i \in [1, N^l], and\ \forall l \in [1, L]$.
      ii. Back Propagation: Estimate the error gradient at the neurons of the output layer and propagate the error back to the first layer in the hidden network to estimate the respective delta error.

$$\Delta_k^l, \forall k \in [1, N_l], and\ \forall l \in [1, L].$$

      iii. Compute the change in the bias and weights and then update the weights and bias using the same scale with the learning parameter α.

6. RESULTS AND ANALYSIS

The proposed neural network for the binary classification of abnormality of the heart was constructed based on the concepts of deep learning based on a train and test dataset. The dataset was divided for training and validation based on k-fold cross validation technique. The model performance was evaluated primarily based on the accuracy and F1-score. The table below shows the summary of accuracy and loss values during the training process. The other evaluation metrics used to evaluate the performance of the proposed classifier are False Positive Rate, False Negative rate, sensitivity/ recall, precision, accuracy, F1 score, cohen kappa metric, RoC curve (visual representation of the relationship among the true positive rate and false positive rate). The Table 1 summarizes the mathematical expression for estimating the evaluation metrices and their estimated values.

Table 1. Summary of Evaluation metrics

Note: tp- true positive, fp – false positive, tn- true negative, tp – true positive

SNo	Metric	Mathematical Expression	Estimated Value for proposed 1D-CNN
1	False Positive Rate	$FPR = \frac{fp}{fp+tn}$	0.0548
2	False Negative rate	$FNR = \frac{fn}{tp+fn}$	0.0646
3	Sensitivity/ Recall	$TPR\ (or) Sensitivity = \frac{tp}{tp+fn}$	0.9353
4	Precision	$Positive\ Predict\ Value$	0.9487

		$(or)\ Precision = \frac{tp}{tp + fp}$	
5	Accuracy	$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$	94.92%
6	F1 score	$F1 = 2\frac{precision * recall}{precision + recall}$	0.9428
7	Cohen Kappa Metric	$k = \frac{p_t - p_o}{1 - p_o}$	0.9145
8	AUC RoC		0.9068

The accuracy of the CNN model after 50 epochs was observed to be 96.88% over the training period and while validation the model yields and accuracy of 94.92%. For analysing the performance of the 1D CNN model the results are compared with the results specified in [15, 16] which utilizes a deep neural classifier and a conventional machine learning classifier for the binary classification task. The loss and accuracy during the training process are tabulated in Table below. The dropout regularization used during the training process helps to avoid overfitting and aids the model to learn the patterns available in the heart sound to discriminate between the normal or abnormal sounds. The dropout value was fixed as 0.15 during the training process and the value was fixed based on trial and error approach. Similarly the initial learning rate was fixed as 0.01 and reduced gradually over the epochs linearly to make the model to learn faster. The high learning rate in the early epochs provides rapid learning, but when the learning rate decreases as it approaches minima, it oscillates in a closer region around local minimum rather than straying far away. This was clearly presented in the training loss and accuracy curve in Fig.

To validate the efficiency of the proposed model architecture, a set of additional convolutional layers were used to classify the heart sounds. Monte carlo method was used to find the optimal values for the model hyper-parameters summarized in Table 2. The experiments were repeated for different number of epochs and average performance of the model reached to a satisfactory level in the 50th epoch. The table below presents the list and value of the hyper-parameters used in the study.

The model yields a good AUC thus representing the performance of the model for different threshold values. AUC is the probability that the trained model gives higher rank for a positive test case when compared to a negative test case. Irrespective of the chosen threshold values the AUC helps to quantify the quality of the model's prediction.

Table 2. Hyperparameter values (Estimated by Monte carlo method)

Parameter	Value
Initial learning rate	0.001
Optimizer	ADAM
Dropout rate	0.15
Number of convolutional layers	12
Kernel size	3
Batch size	32
Pooling size	4

Further increasing the training epochs the model tends to overfit and hence the training was stopped at 50th epoch. The comparison with the other classification approaches (classification of spectrogram using 2D-CNN, and machine learning approach) proved that 1D-CNN converges faster with higher accuracy.

Table. 3 Model performance

Epoch No.	Loss	Accuracy	Validation Loss	Validation Accuracy
5	1.5893	0.7854	2.1584	0.6785
10	1.2071	0.8491	1.9587	0.7642
15	1.0042	0.8654	1.4516	0.8115
20	0.9856	0.8915	1.2564	0.8459
25	0.8420	0.9145	1.1520	0.8796
30	0.7645	0.9305	0.9812	0.9125
35	0.7241	0.9452	0.9145	0.9288
40	0.6654	0.9488	0.8455	0.9355
45	0.5948	0.9586	0.7856	0.9412
50	0.5142	0.9688	0.7154	0.9492

The experimental results and comparison with the conventional approach presented in Fig. 7 indicates that the deep features used in the classification include additional details thus improving the classification accuracy. The training and validation statistics were plotted as graph in Fig. 6 which shows that model converge faster with any much deviations as the optimal value for the hyper-parameters were fixed. Even though the 2D-CNN and the transfer learning techniques yields higher accuracy in image classification; their performance in the audio or

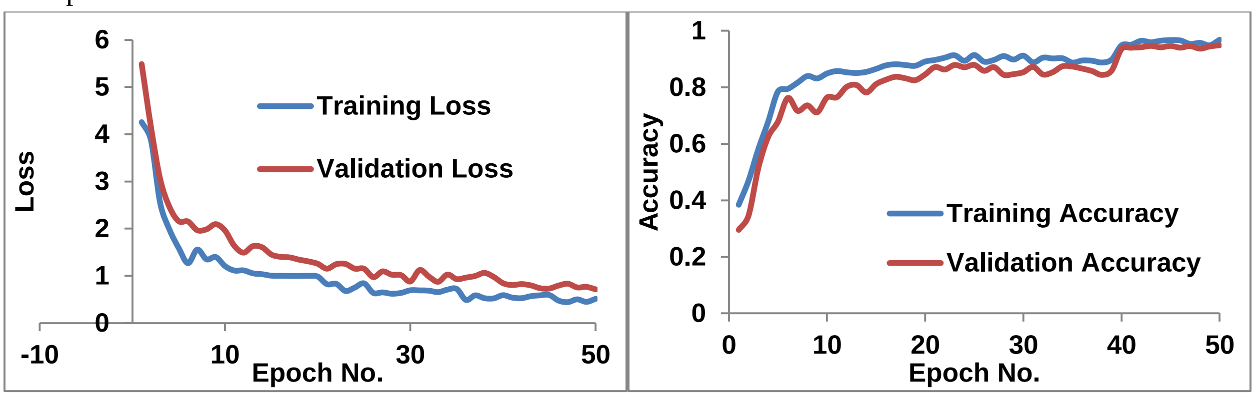

Fig. 6 a) Comparison of Training and Validation loss b) Comparison of model accuracy

sound classification using spectrograms are observed to be lesser as the sound signals are transparent in nature [2]. On a spectrogram, occurrence of discrete sound does not spread across multiple layers; but in turn, they all add and formed as a unique holistic representation. That is, a specific measured frequency that is present within spectrogram cannot be believed to correspond to a class of sound since its amplitude might have been formed by a set of accumulated sounds or even by complicated mixture of sound waves such as phase cancellation. This causes classification based on spectrogram representations of audio/ voice signals more challenging.

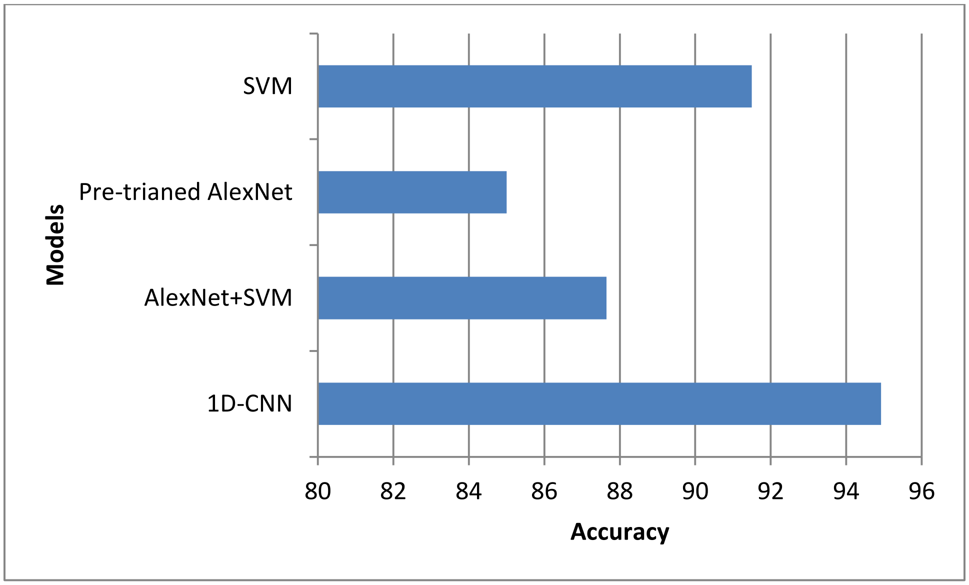

Fig. 7 Comparison of proposed model performance with other approaches

The two dimensions of spectrograms indicate fundamentally distinct units, one being frequency intensity and another is time. A sound event's place in time is offset when it is moved horizontally, and a sound event implies the same information despite of when it occurs. Increasing a sound upward, on the other hand, might change its definition: moving the frequencies of a heart sound higher, for example, can change its nature from normal to pathological or noisy. A sound event's spatial breadth can be altered by applying frequency shifts [2]. As a result, 2D CNNs which are considered to be spatially invariant may not be as effective for this category of data. Similar surrounding pixels in pictures are frequently thought to correspond to a certain visual object, while frequencies in sound are frequently scattered (not confined to a particular area) on the spectrogram [4]. Sounds that occur frequently in between periodic intervals are made up of a basic frequency and a series of harmonics that are separated by characteristics determined by the sound source. The nature and character of the sound is determined by the combination of these harmonics. Acoustic frequencies are not clustered locally, but they do migrate together in a predictable pattern. This complicates the challenge of utilising 2D convolutions to detect local patterns in spectrograms since they are typically unevenly spaced apart despite moving according to the same variables. Thus the 1D-CNN based classification of the sound signals yields better results when compared to the spectrogram based 2D-CNN classification.

CONCLUSION

This study focussed on methods for segmenting and classifying the heart sounds using deep neural classifier which eliminates the need for manual extraction of features from the sound signals. In the conventional machine learning approach the manual feature extraction process was complex and makes the classifier to perform less efficiently. The proposed method involves a deep neural classifier to extract the required features for differentiating the normal and abnormal heart sounds. The pre-processing techniques employed were focused on removing the noise in the heart sounds especially the spectral gating methods eliminates the noise present in the signals based on its frequency and duration. The 1D-CNN was less complex to implement and from the graphical representation of the result analysis and comparison, it is very clear that it performs better than the 2D-CNN and conventional machine learning approaches.

Acknowledgements

Not applicable

Author's Contribution

V. Anantha Natarajan has contributed in problem framing and developed, optimized and evaluated the deep neural model used for performing the classification task. Sunil Kumar has contributed the detailed literature study of various related works. Naresh T helped in data pre-processing and prepared the evaluation reports including the graphs and table. Suneetha K drafted the manuscript. The authors read and approved the final manuscript.

Funding

Not applicable

Availability of data and materials

The present manuscript reports an open-source licensed software packages and the data used in this manuscript were from the PhysioNet Challenge hosted in 2016.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Computer Science and Engineering,
Sree Vidyanikethan Engineering College (Autonomous), Tirupati, Andhra Pradesh, India.
vananthanatarajan@vidyanikethan.edu

[2]Department of Computer Science and Engineering,
Sree Vidyanikethan Engineering College (Autonomous), Tirupati, Andhra Pradesh. India.
sunilmalchi1@gmail.com

[3]Department of Information Technology,
Aditya Institute of Technology and Management, Tekkali, K Kotturu, Andhra Pradesh. India.
itsajs@gmail.com

[4]School of Computer Science & IT, Jain (Deemed-to-be University),
Bangalore, India
keerthisuni.k@gmail.com

REFERENCES

1. Ghosh, Samit Kumar, et al. "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features." *IEEE Sensors Letters* 3.12 (2019): 1-4.
2. L. Wyse. 2017. Audio Spectrogram Representations for Processing with Convolutional Neural Networks. Proceedings of the First International Workshop on Deep Learning and Music joint with IJCNN. Anchorage, US. May, 2017. 1(1). pp 37–41.
3. Langley, Philip, and Alan Murray. "Heart sound classification from unsegmented phonocardiograms." *Physiological measurement* 38.8 (2017): 1658.
4. Ren, Zhao, et al. "Learning image-based representations for heart sound classification." *Proceedings of the 2018 International Conference on Digital Health*. 2018.

5. Pratima Upretee, Mehmet Emin YU (2019) ¨ KSEL "Accurate Classification of Heart Sounds for Disease Diagnosis by A Single Time-Varying Spectral Feature: Preliminary Results" 978-1-7281- 1013-4/19/$31.00, 2019 IEEE
6. Sachin Vernekar, Saurabh Nair, 2016 "A Novel Approach for Classification of Normal/Abnormal Phonocardiogram Recordings using Temporal Signal Analysis and Machine Learning" Physionet, challenge2016
7. Pedrosa J, Castro A, Vinhoza TT. 2014 "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, IL: IEEE; 2014. p. 2294–2297.
8. Jiang, Zhongwei, and Samjin Choi. (2006) "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope." Expert Systems with Applications 31, no. 2: 286-298.
9. Subhamoy Mandal,Roshan J Martis, K M Mandana, 2014 "Practice of Cardiac Auscultation: Clinical perspectives and its implications on computer aided diagnosis" Dec. 30, 2014.
10. Anthony Bouril, Darya Aleinikava, 2016 "Automated Classification of Normal and Abnormal Heart Sounds using Support Vector Machines" Computing in Cardiology 2016; 43:549-555.
11. Simarjot kaur, Mandeep singh, 2015 "Classification of heart sound signals using multi-model features" procedia computer science 58 (2015) 165-171.
12. Ilias Maglogiannisa, Euripidis Loukis, 2009 "Support Vectors Machine-based identification of heart valve diseases using heart sounds" computer methods and programs in biomedicine 9 5 (2 0 0 9) 47–61.
13. Omer Deperlioglu (2018) "Classification of segmented heart sounds with Artificial Neural Networks" International Journal of Applied Mathematics Electronics and Computers (IJAMEC) IJAMEC, 2018, 6(4), 39–44.
14. Christopher Sch¨olzel, Andreas Dominik (2016) "Can Electrocardiogram Classification be Applied to Phonocardiogram Data? –An Analysis Using Recurrent Neural Networks", Computing in Cardiology 2016; VOL 43 ISSN: 2325-887X DOI:10.22489/CinC.2016.167-215.
15. Ganesh, D., et al. "Improving Security in Edge Computing by using Cognitive Trust Management Model." 2022 International Conference on Edge Computing and Applications (ICECAA). IEEE, 2022.
16. B. E. Manjunath Swamy. "Personalized Ranking Mechanism Using Yandex Dataset on Machine Learning Approaches." Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1. Singapore: Springer Nature Singapore, 2022.
17. Burada, Sreedhar,"Computer-Aided Diagnosis Mechanism for Melanoma Skin Cancer Detection Using Radial Basis Function Network." Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1. Singapore: Springer Nature Singapore, 2022.
18. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Efficient energy management for reducing cross layer attacks in cognitive radio networks." Journal of Green Engineering 11 (2021): 1412-1426.
19. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Novel Defense Framework for Cross-layer Attacks in Cognitive Radio Networks." International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020. Springer Singapore, 2021.
20. Kumar, M. S, et al. "Deep Convolution Neural Network Based solution for Detecting Plant Diseases." Journal of Pharmaceutical Negative Results (2022): 464-471.
21. Prasad, Tvs Gowtham, et al. "Cnn Based Pathway Control To Prevent Covid Spread Using Face Mask And Body Temperature Detection." Journal of Pharmaceutical Negative Results (2022): 1374-1381.
22. Venkateswara Reddy, L., et al. "Bio-Inspired Firefly Algorithm for Polygonal Approximation on Various Shapes." Intelligent Computing and Applications: Proceedings of ICDIC 2020. Singapore: Springer Nature Singapore, 2022. 95-107.
23. Kumar, A., Kumar, M. S., Vikhar, P. A., Ghodke, V., Waghulde, R. R., & Rathod, M. (2022). Improving the Software Privacy in the OFDM 5G Communication Integrated with License Key in the Hardware

Communication Parameters. International Journal of Intelligent Systems and Applications in Engineering, 10(2s), 236-240.

24. Natarajan, V. A., Kumar, M. S., Tamizhazhagan, V., & Chevdumoi, R. M. (2022). Prediction Of Soil Ph From Remote Sensing Data Using Gradient Boosted Regression Analysis. Journal of Pharmaceutical Negative Results, 29-36.
25. Kumar, M. Sunil, et al. "Deep Convolution Neural Network Based solution for Detecting Plant Diseases." Journal of Pharmaceutical Negative Results (2022): 464-471.
26. P. Sai Kiran, and M. S Kumar. "Resource aware virtual machine placement in IaaS cloud using bio-inspired firefly algorithm." Journal of Green Engineering 10 (2020): 9315-9327.
27. Balaji, K., P. Sai Kiran, and M. Sunil Kumar. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." Applied Nanoscience (2022): 1-9.
28. Kumar, M. Sunil, et al. "APPLYING THE MODULAR ENCRYPTION STANDARD TO MOBILE CLOUD COMPUTING TO IMPROVE THE SAFETY OF HEALTH DATA." Journal of Pharmaceutical Negative Results (2022): 1911-1917.
29. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Efficient energy management for reducing cross layer attacks in cognitive radio networks." Journal of Green Engineering 11 (2021): 1412-1426.
30. Haya Alaskar, Nada Alzhrani, Abir Hussain, and Fatma Almarshed. The implementation of pretrained AlexNet on Pcg classification. In International Conference on Intelligent Computing, pages 784-794. Springer, 2019.
31. Tamer Olmez and Zu¨mray Dokur. Classification of heart¨ sounds using an artificial neural network. Pattern Recognition Letters, 24(1-3):617–629, 2003.