IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 April 2024, accepted 29 April 2024, date of publication 6 May 2024, date of current version 15 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3397574

RESEARCH ARTICLE

An RNN-Bi LSTM Based Multi Decision GAN Approach for the Recognition of Cardiovascular Disease (CVD) From Heart Beat Sound: A Feature Optimization Process

N. A. VINAY[1], (Member, IEEE), K. N. VIDYASAGAR[2], (Member, IEEE), S. ROHITH[3], (Member, IEEE), DAYANANDA PRUTHVIRAJA[4], (Senior Member, IEEE), S. SUPREETH[5], (Member, IEEE), AND S. H. BHARATHI[2], (Member, IEEE)

[1]Department of Electronics and Communication Engineering, Dayanand Sagar College of Engineering, Bengaluru 560078, India
[2]School of Electronics and Communication Engineering, REVA University, Bengaluru 560064, India
[3]Department of Electronics and Communication Engineering, Nagarjuna College of Engineering and Technology, Bengaluru 562110, India
[4]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India
[5]School of Computer Science and Engineering, REVA University, Bengaluru 560064, India

Corresponding author: Dayananda Pruthviraja (dayananda.p@manipal.edu)

ABSTRACT The cardiovascular system is responsible for carrying the blood along with nutrition and oxygen throughout the body. This system consists of heart, blood, and blood vessels. The experts, or doctors called as cardiologists, analyze the sounds of heart's (lub-dub) beat and flow of blood to diagnose Cardio Vascular Disease (CVD) using a traditional stethoscope and phonological cardiogram technique. Through the stethoscope, the cardiologist will listen to vibration produced by heart beat and heart beat sound and murmur sound are popularly known as phonocardiogram (PCG) signals, which are being recorded for medical diagnosis purposes using a stethoscope. The development of a technique for the automatic recognition of HVD's assists the experts in recognizing the CVD effectively in the initial stage itself from PCG signals. There are many tools available to help doctors in a clinical setting for the accurate diagnose the CVD in a less time. The main aim of this proposed work is to provide an Artificial Intelligence (AI) based PCG signal analysis for the automatic and early detection of various cardiac conditions using supervised and unsupervised Recurrent Neural Network (RNN) based Bidirectional Long Short-Term Memory (Bi-LSTM) Machine Learning (ML) algorithm. Along with this algorithm, Generative Adversarial Networks (GAN's) is considered because they can create fresh, high-quality, pseudo-real data that resembles their training set which has been demonstrated by using their two unique networks: Discriminator Network (DN) and the Generator Network (GN). The proposed method is tested using heart sound signals from the well-known, freely accessible PhysioNet and Kaggle datasets. The Experimental results are validated based accuracy, precision, F1-score, sensitivity, and specificity.

INDEX TERMS RNN-Bi LSTM, multi decision GAN approach, recognition of cardiovascular disease, feature optimization process.

I. INTRODUCTION

Heart arrhythmia, also known as heart valve disorder (HVD), refers to the irregular heartbeat or sound characterized by an abnormal heart rate, either too slow or too fast. Detecting HVD at an early stage can significantly reduce the mortality rate associated with various cardiac failures, potentially by up to one-third [1]. The prevalence of cardiovascular diseases is alarmingly increasing worldwide, with projected costs of treating such conditions in the USA alone expected to reach

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. García-Fernández.

$1 trillion per year by 2030 [2]. In addition to irregular heartbeat, morphological parameters such as shortness of breath, dizziness, palpitations, fainting, and other symptoms may indicate the presence of arrhythmias or HVD. While not all arrhythmias or HVDs lead to mortality, certain types, including atrial fibrillation, premature ventricular contractions, and excessive supraventricular ectopic beats, are associated with various cardiovascular diseases (CVDs) such as stroke.

Cardiovascular disease (CVD) is a major global health concern, causing approximately 17.9 million deaths annually and accounting for 44% of all non-communicable disease-related deaths, as highlighted by the World Health Organization (WHO) [1]. Among the complications associated with CVD, cardiovascular arrhythmias stand out, characterized by abnormal electrical impulses in the myocardium, leading to irregular heart rhythms. These arrhythmias contribute significantly to the occurrence of cardiac arrests [2]. Hence, the prompt and accurate identification of cardiac arrhythmias becomes paramount in order to save lives and mitigate the impact of these life-threatening conditions.

The heart's rhythm is governed by an electrical impulse generated by the sinoatrial node (SA node) located in the right atrium. This electrical activity of the heart can be detected by placing electrodes on the patient's body surface and recording the resulting electrical potential differences. These measurements are visually represented and captured in an electrocardiogram (ECG), providing valuable insights into the heart's electrical activity.

An electrocardiogram (ECG) is a widely used non-invasive diagnostic tool that records the electrical activity of the heart using a group of electrodes typically placed on the patient's skin. It is an effective method for detecting heart conditions, including arrhythmias.

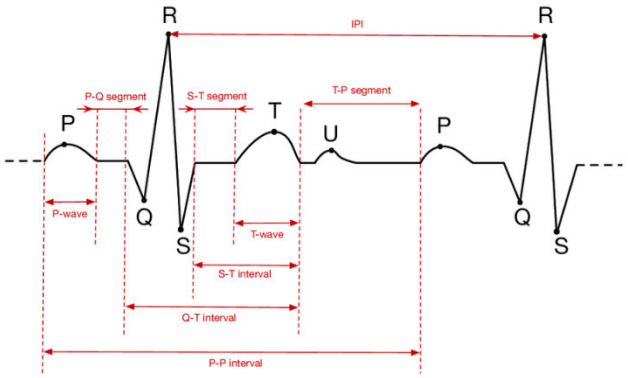


FIGURE 1. A typical waveform of an ECG signal.

The ECG waveform consists of various components such as the P wave, Q wave, R wave, S wave, T wave, and U wave, which collectively form a normal ECG cycle. These components are characterized by time-domain features like amplitude, duration, interval, and segment [3], as illustrated in Figure 1. The sequential depolarization of the atria and ventricles generates distinct waveforms [4]. Abnormalities in the ventricles or atria can cause irregular ECG waveforms, enabling the diagnosis of various cardiac arrhythmias such as premature ventricular contraction (V), atrial premature beat (A), and right bundle branch blocks (R). These deviations in the ECG waveform are indicative of alterations in the origin and propagation of electrical impulses in the heart, reflecting the presence of cardiac arrhythmias.

The recognition of heart valve disorders (HVDs) from phonocardiogram (PCG) signals, which capture the heart sound, has gained popularity as a cost-effective and non-invasive diagnostic approach. Digital stethoscopes equipped with digital signal processing kits enable the acquisition and interpretation of PCG recordings, aiding in the diagnosis and treatment of valve-related disorders. In this study, our focus is on the automatic recognition of various HVDs based on the distinct lub-dub heart sound in the PCG signal.

A method for classifying HVDs based upon the formant characteristics of PCG signals, whose acoustic properties represent heart sounds. The conventional time frequency domain statistical features can be extracted from PCG signals, clearly showing big fluctuations specific to various HVDs. The direct classification of HVDs is also done with PCG signals using sequential networks. The spectrum of the PCG signal shows distinct peaks comparable to the formant peaks in voiced speech, especially in the voiced component of the heart sound (lub-dub).

In a normal person, we can hear two prominent heart sounds, S1 (lub) and S2 (dub), using a stethoscope. The first heart sound (S1 or lub) occurs due to the closure of the mitral and tricuspid valves, which follows ventricles contracting in systole [3]. S2, or dub, is generated in diastole as the aortic and pulmonic valves are closing. These include another softer sound known as the S3, S4, murmur due to arterial turbulence of blood flow, ejection click (EC) during systole, opening snap/snaps (OS) in diastole and mid systolic snicks [4]. One of the early means of diagnosis for diseases of the valvular heart includes cardiac auscultation, which entails listening to the heartbeat using a stethoscope. Therefore, the early recognition of HVDs and cardiovascular diseases relies on analyzing the first two formants of PCG signals, as these capture the acoustic characteristics of the heart sound. The main contributions of this work are:

- Our proposed work focuses on efficient feature extraction techniques to extract the magnitude, frequency, and absolute value of PCG signals.
- We also employ a deep learning model to automatically extract these parameters from the lub-dub sound.
- Furthermore, a hybrid neural network incorporating Generative Adversarial Networks (GANs) is utilized for the classification of normal and abnormal cardiac sounds.
- To address false-positive classifications and improve accuracy, a hybrid RNN-based Bi-LSTM model is proposed, which classifies abnormalities based on true-positive signals.

➢ The proposed algorithm is evaluated on PhysioNet and kaggle datasets, and the results are validated using an object-oriented approach where the morphological parameters of heart sound are segmented into delta-groups.

II. RELATED WORK

Irregularity in heart beat is one of the chronic heart conditions that affects older people the most persistently and has a high morbidity and mortality rate, including stroke, cardiac failure, and coronary artery disease. Automatic detection and classification of arrhythmia heartbeats using electrocardiogram (ECG) signals is important for cardiac patients. So, to extract the desired or required features the Hao Dang, et.al has developed a three-layer Convolutional Neural Network, which consists of one-layer simple CNN and two- layers are of multi-scale fusion CNN. The simple CNN is a baseline network with manifold convolutional layers and a straightforward CNN architecture that is used to test the single dimension CNN's capacity for processing ECG signals. It is suggested that the MSF-CNN A will enhance the plain-capacity CNNs for learning. In this proposed work, the experimentation is conducted using normal beat (N), supraventricular ectopic (S), ventricular ectopic (V), fusion beat (F), and unknown beat (Q) types of signals which are available publicly in MIT-BH database and in order to evaluate the effectiveness of these models, six groups of ablation experiments are also carried out [1].

ECGs form an essential diagnostic instrument that is widely used by cardiologists when dealing with cardiac disease and pathology. The cardiac electrical activity is accumulated as a function of time, and subsequently analyzed via ECGs, whereby relevant electronic signals are gathered from various external electrodes attached on the skin. Cardiac dysfunction and arrhythmias that often accompany such disorders also increase the risk of sudden death. Hence, ECG signal studies have raised widespread interest among both the computer and biomedical circles. Due to variations in time and amplitude, of ECG signal, BLSTM model which is one form of common RNN architecture is employed for analysis of time series. However, an ELM based on local receptive fields (ELM-LRFs) is among the fastest techniques employed in the segmentation and classification of time series signals. DELM-LRF-BLSTM [2] is a faster and accurate hybrid deep learning model for ECG signal recognition proposed by FENGJUAN QIAO & Co. It yielded very high levels of accuracy and sensitivity of about 99.32% and 97.15%, respectively when applied to the MIT-BIH Arrhythmia dataset. This ensures that the model is viable and effective. Additionally, a one-time activation of heartbeat detection takes about six and a half milliseconds. This makes the provided algorithm a good practice since it attains desirable high performance with minimal computation.

For instance, according to World Health Organization (WHO) in their latest report, CADs have become the main cause of unexpected deaths with nearly seventeen millions of it recorded across the world and 44 per cent of the global non-communicable diseases related mortalities are Many cardiac arrests stem from cardiac arrhythmias that are associated with abnormal electrical pulse formation and propagates in the myocardial tissues [1]. So, the accurate analysis cardiac arrythmias reduces the mortality towards this Huang, Y., Zhang, F., Wang, D., Guan, Y., Zhou, F., Pan, Y., & Liu, W. Reference [5] proposed a novel ensemble classification algorithm KSMAX for accurate detection of cardiac ventricular and atrial the present study considers only the wavelet decomposed ECG signal in four level with respect to the morphological parameters are the morphological visual representation associated with the QRS complexity and the basic components of the third and the fourth stage. The AAMI standard showed that this proposed method achieved an overall accuracy of about 98.68% from which fifteen different heartbeat waveforms were extracted from a public MITDB. The classification accuracy for each of the six major types is as follows: We obtained 98.75% for N, 99.77% for R, 99.70% for L, 94.81% for A, 98.57% for V and 99.94%. Table1 shows the summary of related work. The summary of the related literature is discussed in table 1. After referring to the literature from [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], and [44], the recognition of various irregularities of cardiac is done based on ECG signals and few works related to PCG signals [14], [16], [18] is done. Because of lack of PCG signal analysis there is large research gap is identified in the field of recognition of irregularities.

III. DESIGN AND IMPLEMENTATION OF PROPOSED METHODOLOGY

A. AUSCULTATION ANALYSIS

The heart Auscultation contains two segments: sound acquisition and analysis. For the acquisition of Heart sound stethoscope is placed at the proper location on the chest of the patient with a small amount force to listen the heart sound as shown in the figure 2.

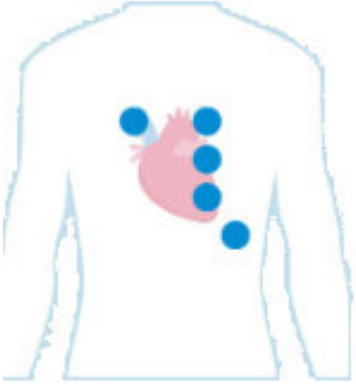


FIGURE 2. Stethoscope position on the chest of the patient.

TABLE 1. Summary of related work.

Ref No.	Types of Classifiers	Types of features	Pathology	Recognition rate	Datasets used	Methods
[6]	CNN-LSTM	Local characteristics of Atrial Fibrillation (AF) from ECG signals	ECG SIGNAL	Accuracy: 99.59% Sensitivity:99.93% Specificity: 97.03%	MIT-BIH: AF databases	Accurate analysis of RR intervals and QRS interval from ECG sample
[7]	Novel Integration of Open loop And Closed-Loop Switch Modes Is Proposed	Morphological parameters of heart beat from ECG signal using a wearable device.	ECG SIGNAL	-	Database is created using a wearable device	• R-peak detection algorithm for initial diagnosis • Abnormality detection • Classification ECG signals in cloud
[8]	Optimized Convolutional Neural Network (CNN)	Cardiac arrhythmias- ECG	ECG SIGNAL	Accuracy: 98.45%	MIT-BIH arrhythmia	Genetic algorithm-based CNN
[9]	A Deep Learning-Based Multi-Model System	Cardiac arrhythmias- ECG	ECG SIGNAL	Accuracy: 95.81%	MIT-BIH arrhythmia	CNN-LSTM integrated with RRHOS-LSTM
[10]	Multi-Layer Perceptron (MLP), k-Nearest Neighbor, Support Vector Machine, Random Forest, and Extra Decision tree	ECG time-series	ECG SIGNAL	Accuracy: 95.21%	MIT−BIH arrhythmia database & MIT−BIH Supraventricular arrhythmia	Multi-objective particle swarm optimization (MOPSO)
[11]	Neural Network, SVM, and KNN	Amplitude, interval, and duration- ECG	ECG SIGNAL	Accuracies for: • N:97.79%, • L:99.50%, • R: 99.59% • V: 97.69%, • A: 89.70%, and • P: 99.92%	MIT-BIH arrhythmia database	
[12]	Convolutional Neural Network (CNN)	Premature ventricular contraction- ECG	ECG SIGNAL	Accuracy: 95.59% Sensitivity:97.93% Specificity: 98.03%	MIT-BIH database	OTSU Method
[13]	Dual Heartbeat Coupling Based on Convolutional Neural Network	SVEB or S beats and VEB- ECG	ECG SIGNAL	sensitivity and positive predictive rate is increased by 12.2% and 11.9%	MIT-BIH database	
[14]	Ensembled Bagged Trees (EBT)	Heart valve disorder (HVD) analysis based on magnitude, frequency and phase each PCG formants	PCG SIGNAL	PhysioNet: 93.46% and CinC : 99.28%	PhysioNet CinC	
[15]	Bidirectional LSTM (Bi-LSTM)	RR interval, P wave, QRS wave, ECG-intervals, QRS interval, PR interval, ST interval, ST level- ECG	ECG SIGNAL	-	MIT-BIH Arrhythmias Database, SVDB and NSTDB	Stacked Denoising Autoencoders (SDAE), as encoder, automatically learns semantic encoding of heartbeats without any complex feature extraction in unsupervised way
[16]	End-To-End Deep Multi-Scale Fusion convolutional neural network	Abnormal rhythms and noise distribution- PCG	PCG SIGNAL	Precision:0.838 Recall: 0.822 F1 Score: 0.828	CPSC_2018 and PhysioNet/CinC_2017	multi-class arrhythmia detection
[17]	Discrete-Time State-Space	Extract features of heartbeats via ECG records	ECG SIGNAL	Accuracy: 99.3% Sensitivity:99.6% Specificity: 99.8%	MIT-BIH Arrhythmia	q-lag unbiased finite impulse response (UFIR) smoother

TABLE 1. *(Continued.)* **Summary of related work.**

[18]	Convolutional Neural Network (CNN)	Early detection of cardiac pathologies- PCG	PCG SIGNAL	Accuracy: 99.66%	Three data set is used for experimentation: first recorded cardiac signal using digital stethoscope and other two taken with the help of mobile smart devices	CNN in the form of audio samples as well as in spectrogram format
[19]	Feature Enrichment - convolutional neural network (FE-CNN)	Heartbeat classification from electrocardiogram (ECG) using Feature Enrichment (FE)	ECG SIGNAL	Beat Type: S Beat Sen:75.6% Ppr:90.6% F1 Score:0.82 Beat Type: V Beat Sen:92.8% Ppr:94.5% F1 Score:0.94	MIT-BIH Arrhythmia	-
[20]	End-To-End Approach and A Two-Stage Hierarchical Approach—Based On Deep Convolutional Neural Networks (CNN's)	Augmenting the original imbalanced dataset with generated heartbeats	ECG SIGNAL	Accuracy: 98.0% Sensitivity:97.7% Specificity: 97.4% Precision: 90.0%	MIT-BIH arrhythmia	novel data-augmentation technique using generative adversarial networks (GANs)
[21]	Heartbeat classification using Convolutional Neural Network	Segment label	ECG SIGNAL	Beat Type: S Beat Sen:0.92 Ppr:0.76 F1 Score:0.74 Beat Type: V Beat Sen:0.98 Ppr:0.76	MIT-BIH arrhythmia	
[22]	New generative adversarial network-based deep learning method called HeartNet	--	ECG SIGNAL	Accuracy: 99.67 ± 0.11	MIT-BIH dataset	multi-head attention mechanism on CNN architecture.
[43]	CNN	-	PCG Signal	96.6%	Raw Dataset	CNN based classifier.
[45]	Progressive Dense Fusion Network	-	ECG and PCG signals	Sensitivity: 94.85 ± 5.14 Specificity: 93.97 ± 4.97 accuracy: 94.41 ± 3.04 AUC: 0.973 ± 0.026	PhysioNet/CinC 2016 dataset	Co-learning-assisted progressive dense fusion network
[46]	MLP, RF, Conv1D, DNN, XGB	Spectrograms, MFCCs	PCG Signals	Accuracy: 95.65%	PASCAL CHALLENGE	Features are augmented with synthetic noise, ensembled, and various models employed including deep learning and machine learning.
[47]	Support Vector Machine (SVM), deep neural network (DNN) and centroid displacement based k nearest neighbor	Audio PCG signals	PCG Signals	Accuracy upto: 97%	Classification-of-Heart-Sound-Signal-Using-Multiple-Features	Artificial intelligence framework for heart disease classification

To determine whether a recorded sound is associated with a healthy or sick heart, heart sound analysis is used. This research is a component of a larger effort to develop a robotic system that can acquire and analyze heart sounds called a remote auscultation system. This work only addresses the analysis because acquisition is primarily a mechanical process. Further, the work does not attempt to identify any disorders; rather, it only seeks to identify their presence.

In other words, there are two categories for heart sounds: ''healthy'' and ''diseased.'' This problem is simpler compared to multi-class disease diagnosis, but it allows one method which does not include the need to segment heart sounds. This also makes the speech recognition system insensitive to difficult-to-segment sounds where the signal has been corrupted.

The first heart sound (S1) and the second heart sound are the two events that make up heart sounds in healthy adults (S2). They are collectively known as the ''basic heart sound'' (FHS). The period from the start of S1 to the start of S1 after that is known as a cardiac cycle or a single heartbeat. Systole refers to the period from the end of S1 to the start of S2 of the same cycle, and diastole refers to the period from the end of S2 to the start of S1 of the following cycle. The components of abnormal heart sounds' cardiac cycles do not belong to the FHS. Extra heart sounds and murmur sounds are two categories into which these elements can be divided. Murmurs are the second category of abnormal heart sounds. They result from turbulent blood flow through a stenosed (blocked) valve or from retrograde flow through a regurgitated (leaking) valve. Depending on the underlying condition, these sounds can be heard during either systole or diastole. A murmur is a reliable sign of valvular (valve-related) disorders. The appearance of S3 resembles the extra FHS and it needs to be segment properly after second cycle [36].

B. PROPOSED METHODOLOGY

The general structure of the PCG signal processing is given in figure 3. It includes the following three major operation: Data Acquisition, Pre-processing and Segmentation, Processing.

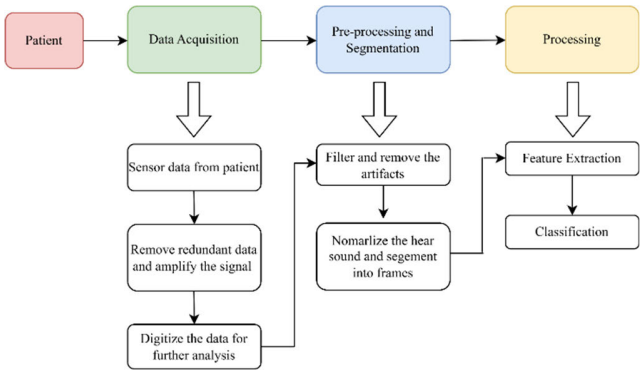


FIGURE 3. General structure of PCG signal processing.

The data acquisition system, records the PCG data in real time and it is recoded using stethoscope at various auscultation points as shown in figure 3. In general, the stethoscope comes with microphone, piezoelectric sensor, capacitive sensor, T-type MEMS heart sound sensor, MEMS piezoresistive electronic heart sound sensor [35]. To detect the cardiac related disorders the PCG signals needs to processed properly, by removing the unwanted signals at the pre-processing stage and always processing longer is difficult compare smaller length sequences, so the recorded sample is segmented into shorter frames. Later, in processing the desired features of PCG signal is extracted and classified for healthy and disorder heart beats.

The proposed work focuses majorly on processing stage, where the feature extraction and classification are done in a lesser time. The block diagram of the proposed methodology is shown in figure 4.

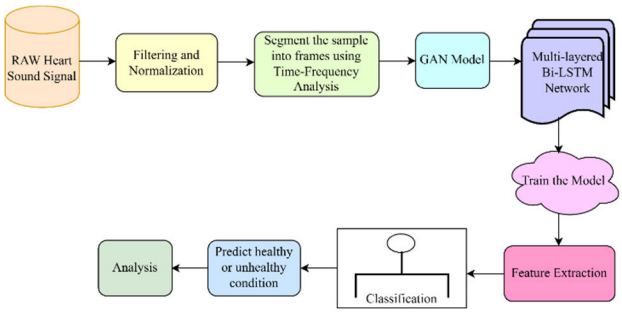


FIGURE 4. Proposed methodology.

C. SEGMENTATION USING HAUG-HILBERT TRANSFORM

PCG, which is the heart sound recorded via acoustic sensors, is the raw signal in the analysis of the heart sound. The two main segments of normal sound conduction include S1 and S2 which relate to the end of the mitral-tricuspid and the aortic/pulmonary valves, accordingly. Nevertheless, the waveform can have murmurs and according to their structure and position, they can indicate different cardiac problems. Such for example, one may experience aortic stenosis murmur between S1 and S2 during systole. Unlike most adults' PCG whose S3 is not included in their children may still have the innocence S3. PCG signals are not high in frequency (less than 70 Hz) and very short in duration (S1 is around 120 ms and S2 around 100 ms). Therefore, just by using time-domain representation will not reveal all the information that it has Hence, most PCG signals are transformed into the time-frequency spectrum by employing Time-Frequency analysis that is used on both the time and frequency scales. Time-Frequency analyses are applied in assessing EEG, EMG, ECG, and other bio-signals. Conversion of 1-D time domain signal to 2-D space with time and frequency information would aid in the simpler process of visualization of the frequency data. They can also make frequency information removal less challenging by acting as inputs for RNNs.

There are many investigations on Segmentation as the most difficult initial step for heart sound analysis. In fact, the best-known approach to segmentation is referred to as the ''envelope analysis''. The method envelopes a heart sound, discovers peak points in envelope signal, selects peaks as corresponding to s1 or s2 and constructs cardiac cycles with the aid of s1 – s1 interval.

1) HILBERT TRANSFORM

In the proposed methodology, the Hilbert-Huang Transform (HHT) is used to study behaviors of the unseen way

of variations in PCG signals and to observe their time-based changes. The use of HHT in EEG and Heart beat classification is discussed in [23] and [24] and the steps followed to obtain the HHT of PCG in the proposed work is shown in figure 5.

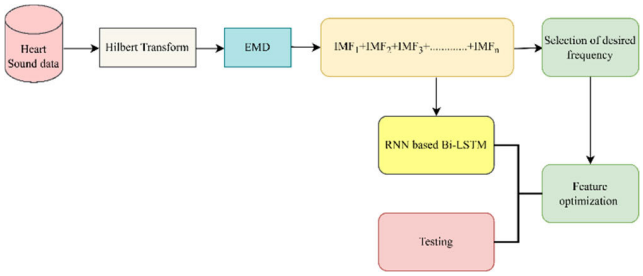


FIGURE 5. Steps in obtaining the HHT of PCG signal.

HHT is a deformation procedure giving rise to a comprehensive and adaptive basis in order that we may describe those oscillations' modes. This differs from most of the other conventional approaches that use a specific set of priori basis. HHT consists of two procedures namely, Hilbert spectrum analysis and the Empirical-Mode-Decomposition (EMD). Discrete signals are decomposed into IMF's using EMD. IMF is defined as a signal that satisfies the following two conditions: The number of maxima, maxima zero crossings should either differ by one or the same as the minima, maxima zero crossings. These two conditions of IMF make Hilbert Transform to maintain significant instantaneous frequency. The EMD of the signal is derived as:

i. The first step of the proposed approach involves utilizing a spline function to generate upper and lower envelopes, denoted as u[n] and l[n] respectively, for the original signal x[n] or temporal result h[n]. These envelopes serve as smoothed representations that encapsulate the variations and trends present in the signal, providing valuable insights into its characteristics.

ii. To obtain the temporal result h[n], the median of the envelope, denoted as m(n), is derived by averaging the upper and lower envelopes. This process involves combining the smoothed representations of the signal, u[n] and l[n], and computing their average. Subsequently, h[n] is obtained by subtracting this envelope average from the original signal, resulting in a modified representation that highlights the deviations and fluctuations within the signal.

$$m(n) = \frac{u(n) + l(n)}{2} \tag{1}$$

$$h_1(n) = x(n) - m(n) \tag{2}$$

Extract the temporary local oscillation

$$h_{11}(n) = h_1(n) - m_{11}(n) \tag{3}$$

iii. Ideally, the temporal result $h_{11}(n)$ corresponds to the first Intrinsic Mode Function (IMF). However, if it does not meet the criteria of an IMF, it is treated as a new signal, and the above steps are repeated on this new signal. This iterative process ensures that each subsequent result aligns with the characteristics of an IMF, enabling the extraction of meaningful and distinct components from the original signal.

$$\begin{aligned} h_{12}(n) &= h_{11}(n) - m_{12}(n) \\ h_{13}n &= h_{12}(n) - m_{13}(n) \\ &\vdots \\ &\vdots \\ h_{1k}(n) &= h_{1k-1}(n) - m_{1k}(n) \end{aligned} \tag{4}$$

Finally, the coefficient of the first Intrinsic Mode Function (IMF) is obtained. This coefficient represents an essential feature derived from the decomposition process and carries valuable information about the underlying oscillatory patterns and variations within the signal.

$$C_{11}(n) = h_{1k}(n) \tag{5}$$

iv. Next, it is important to verify certain conditions for the subtracted data. Confirm that the local maximum value is greater than 0, indicating an upward trend, and that the local minimum value is less than 0, indicating a downward trend. If these conditions are not met, then a standard deviation (SD) threshold is set. This threshold serves as a criterion for determining the presence of significant variations in the data, allowing for further analysis and processing.

$$SD = \sum_{n=0}^{N} \frac{|h_{1k-1}(n) - m_{1k}(n)|^2}{(h_{1k-1}(n))^2} \tag{6}$$

The value of SD must be $0.2 < SD \leq 0.3$

$$\left.\begin{aligned} r_1(n) &= x(n) - C_{11}(n) \\ r_2(n) &= x(n) - C_{12}(n) \\ &\vdots \end{aligned}\right\} \tag{7}$$

v. The subtracted data obtained in the previous step is applied to Step 1 iteratively to create a new Intrinsic Mode Function (IMF). This iterative process continues until two conditions are met:

➢ The difference between the subtracted data and the newly subtracted data falls below a predefined threshold value.

➢ The subtracted data becomes a monotonic function, either consistently increasing or decreasing.

By repeatedly applying Step 1 and evaluating these conditions, the decomposition process ensures that the resulting IMF accurately captures the intrinsic oscillatory behavior and effectively represents the underlying components of the original signal.

$$x(n) = \sum_{i=1}^{N} C_i(n) + r_i(n) \tag{8}$$

Once the Intrinsic Mode Function (IMF) is obtained for each frame, the Hilbert Transform is applied to each IMF-based frame to determine the instantaneous amplitude and frequency. This process yields valuable information about the varying amplitudes and frequencies present in each frame. The IMF coefficients are then stored in a matrix format, organized in accordance with their respective frames. This matrix serves as a comprehensive representation of the IMF coefficients, facilitating further analysis and interpretation of the signal characteristics [25].

$$C_{11} = \begin{bmatrix} C_{11}^1 & C_{12}^1 & C_{13}^1 \ldots\ldots C_{1K}^1 \\ C_{21}^1 & C_{22}^1 & C_{23}^1 \ldots\ldots C_{2K}^1 \\ C_{31}^1 & C_{32}^1 & C_{33}^1 \ldots\ldots C_{3K}^1 \\ & & \vdots \\ & & \vdots \\ & & \vdots \\ C_{n1}^1 & C_{n2}^1 & C_{n3}^1 \ldots\ldots C_{nK}^1 \end{bmatrix}$$

$$C_{12} = \begin{bmatrix} C_{11}^2 & C_{12}^2 & C_{13}^2 \ldots\ldots C_{1K}^2 \\ C_{21}^2 & C_{22}^2 & C_{23}^2 \ldots\ldots C_{2K}^2 \\ C_{31}^2 & C_{32}^2 & C_{33}^2 \ldots\ldots C_{3K}^2 \\ & & \vdots \\ & & \vdots \\ & & \vdots \\ C_{n1}^2 & C_{n2}^2 & C_{n3}^2 \ldots\ldots C_{nK}^2 \end{bmatrix}$$

$$C_{32} = \begin{bmatrix} C_{11}^3 & C_{12}^3 & C_{13}^3 \ldots\ldots C_{1k}^3 \\ C_{21}^3 & C_{22}^3 & C_{23}^3 \ldots\ldots C_{2K}^3 \\ C_{31}^3 & C_{32}^3 & C_{33}^3 \ldots\ldots C_{3K}^3 \\ & & \vdots \\ & & \vdots \\ & & \vdots \\ C_{n1}^3 & C_{n2}^3 & C_{n3}^3 \ldots\ldots C_{nk}^3 \end{bmatrix} \quad (9)$$

In the given matrix, K represents the number of frames, while n represents the number of speech samples within each frame. Each column in the matrix corresponds to a time duration of 5 milliseconds. This organization allows for a clear representation of the temporal segmentation of the speech signal, providing insights into the variations and characteristics of the signal over time.

Further, decomposition of stage 1 IMF component is given by:

$$y(n) = C_{11} + C_{12} \quad (10)$$

$C_{11}-$ Frequency component of IMF stage 1 and it is low compared to C_{12} and C_{12} has high energy than C_{11} so C_{12} is considered for further decomposition.

The high energy components of IMF stage 1 are represented by:

$$C_{21} = C_{31} + C_{32} \quad (11)$$

Finally, the Hilbert Transform is defined as:

$$\begin{aligned} &\mathrm{HT}(\mathrm{x(n)}) \\ &= \mathrm{X(n)} = \hat{x}(\mathrm{n}) \\ &= \begin{cases} -\mathrm{jx(n)}, \mathrm{n} = 1 \ldots\ldots\ldots \dfrac{\mathrm{N}}{2} - 1, & \text{when n is even} \\ \mathrm{jx(n)}, \mathrm{n} = \dfrac{\mathrm{N}}{2} + 1 \ldots\ldots \mathrm{N} - 1, & \text{when n is odd} \end{cases} \end{aligned} \quad (12)$$

D. DESIGN OF GAN TRANSFORM

In the analysis of sound, the most important step is to synthesis data of generative models [44] for the analysis of heart sound (lub-dub). The algorithm needs to give better results after the distributing the samples of original signals from newly generated models. For this process, the system requires a precise generative model like Generative Adversarial Network (GAN). The architecture of GAN consisting of two major elements or components that is: Discriminator and Generator, the working of these two networks is exactly opposite to each other. The major role of generator network, is to generate the data samples which resembles the original sample, whereas, the discriminator role is differentiate the original sample and synthesized sample [20]. Because of this reason, GAN is applicable to synthesis of: image, video, audio. The arrangement of generator and discriminator is arranged as shown in figure 6.

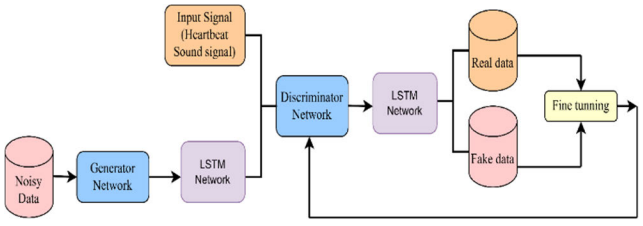


FIGURE 6. Basic structure of GAN to train the two networks: Generator and discriminator network.

In [20], the two major issues related to GAN is discussed. First issue is, when the heartbeat from generator network is not realistic then discriminator is got clowned and under this

scenario it will concentrates on generating the desired group of features to enhances the network loss and to avoid the mode collapse problem. Second issue is, generator network is not trained to differentiate the heartbeat samples if samples of different data set is applied simultaneously and it only generate the similar kind or forged. So, this can be avoided by training the training the generator network to generate the synthetic data for tested samples of different datasets. This procedure is continued till the value of loss function is get saturated. In this proposed work, as mentioned in [20] the loss function computed for both discriminator network and generator network and it is given by:

$$\begin{aligned}&\text{Discrimintor Network loss}\\&= \mathrm{DN}_{\mathrm{loss}}\\&= -\frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}}\log\left(\mathrm{DN}\left(\mathrm{r}^{\mathrm{k}}\right)\right)+\log\left(1-\mathrm{DN}\left(\mathrm{GN}\left(\mathrm{x}^{\mathrm{k}}\right)\right)\right)\end{aligned} \tag{13}$$

$$\begin{aligned}&\text{Generator Network loss}\\&= \mathrm{GN}_{\mathrm{loss}}\\&= -\frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}}\log\left(1-\mathrm{DN}\left(\mathrm{GN}\left(\mathrm{x}^{\mathrm{k}}\right)\right)\right)\end{aligned} \tag{14}$$

where x the noisy vector of fake sample is, r is the real vector original sample, N is the number of samples.

The proposed architecture of generator network is as shown in the figure 7. It **consists of fully connected 4 layers to generate the fake samp**le of 150 × 1 size the generator network is fed with 50 × 1 sample input. The discriminator network architecture is shown in figure 8.

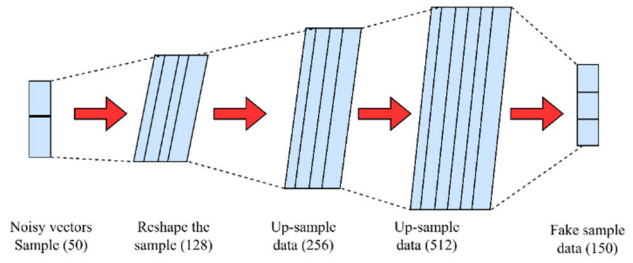


FIGURE 7. Generator network architecture.

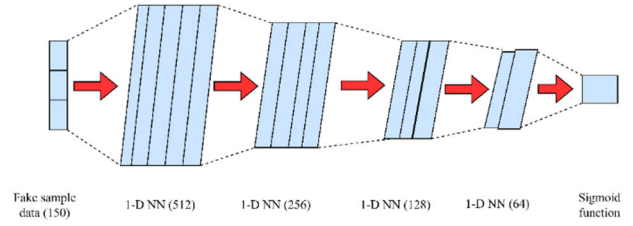


FIGURE 8. Discriminator network architecture.

In the initial stages of processing time series signals, the peak of the signal is very sharp and it is required to smoothen the sharp edges but the information at peak should not get loss also, signal-to-noise ratio should be reduced. But normally when the signal is subjected for smoothing there is a chance of loss of information so, the sharp peak values or edges are replaced by new values obtained from 2m + 1 number of adjacent polynomials.

The neighboring polynomials are considered as an array of smaller segments, these segmented arrays are merged to define the polynomial filter of a signal [34]. The least square value is estimated between the filtering window coefficients (ω) and column coefficients of polynomial functions (P). It is given by:

$$\omega = \mathrm{Xp} + \epsilon \tag{15}$$

In equation, the error polynomial is defined by ϵ. In terms of matrix is given by:

$$\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\\ \vdots\\ \vdots\\ \vdots\\ \omega_n\end{bmatrix} = \mathrm{X}^{\mathrm{T}}\begin{bmatrix}\mathrm{p}_1\\ \mathrm{p}_2\\ \mathrm{p}_3\\ \vdots\\ \vdots\\ \vdots\\ \mathrm{p}_n\end{bmatrix} \tag{16}$$

For smoothing the coefficients, the equation (3) can be written using least square (LS) method as:

$$\mathrm{p} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\omega \tag{17}$$

But, the projection matrix $(\hat{\mathrm{p}})$ for the prediction of values is given by:

$$\hat{\mathrm{p}} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}} \tag{18}$$

To fit the predicted values, the approximation is taken for LS values of X in equation (3). It is given by:

$$\mathrm{X} = \begin{bmatrix}1 & x_0 \cdots\cdots\cdots & x_0^n\\ 1 & x_1 \cdots\cdots\cdots & x_1^n\\ & \vdots & \\ & \vdots & \\ & \vdots & \\ 1 & x_m \cdots\cdots\cdots & x_m^n\end{bmatrix} \tag{19}$$

The matrix defined in equation (7) is known as Vandermond matrix. The polynomial representation of equation (3) is given by:

$$\begin{aligned}\omega_{\mathrm{j}} = \mathrm{p_i X_j^i} + \mathrm{p_{i-1} X_j^{i-1}} + \cdots\cdots\cdots + \mathrm{p_1 X_j^1} + \mathrm{p_0 X_j^0},\\ \text{for j} = 1\ldots\ldots\ldots 2\mathrm{q}+1\end{aligned} \tag{20}$$

In the proposed work, the sound of heart is recorded known as PCG and the behavior of PCG signal is non-stationary. So, as the signal varies the window coefficient size will also get varies hence and this can be avoided in moving window

technique which is of fixed length to generate the coefficients which the improves in controlling of variation of standard deviation, maximum and minimum values. In this proposed work, the moving window is implemented by smoothing the n^{th} coefficient which helps for the analysis of spatial dependency. Then the n^{th} location derivative is given by:

$$\hat{p}_i^n = \left(\left(X^T X\right)^{-1} X^T\right)_i^n \cdot p \tag{21}$$

$$\hat{p}_i^n = (X)_i^n \cdot p \tag{21a}$$

$$\hat{p}_i^n = (W)_i^n \cdot p \tag{21b}$$

$(W)_i^n$ – Moving Window function of nth order.

E. MULTI-LAYERED RNN BASED BI-LSTM

The research on Recurrent Neural Networks (RNNs) have proven that it is an effective modelling tools for sequential data. For several applications, most notably time-series dependency application, RNN have contributed to significant advancement. A RNN cell combines two inputs—one is from current sequence and the other is the output from the previous step and this forms the basic building block of an RNN. But, in recent years the research on of Deep learning is increased significantly in many fields. Because, the deep learning algorithm has many benefits over the traditional machine learning algorithm, such as the feature optimization for complex data preprocessing. In this deep learning architecture, the model is fed with raw data as its input, the unique structure and deeply interconnected network trains the model effectively to process the complex data. So, likewise the RNN also requires such unique structure to process the complex long dependency data and to overcome gradient vanishing problem. One such architecture is, Long Short-Term Memory (LSTM) based RNN and this structure helps to overcome this major disadvantage of RNN. The basic structure of single layer LSTM based RNN is shown in the figure 9.

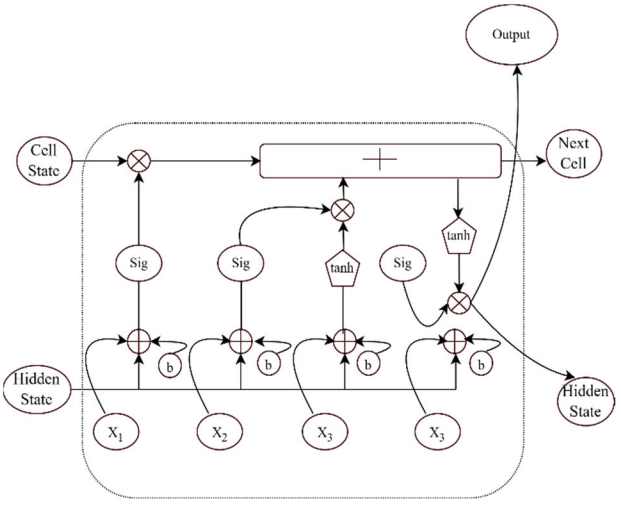


FIGURE 9. Basic structure of Bi-LSTM.

In this proposed work, to process the complex data of cardiac signals, a multi-layered Bi-LSTM structure is used. It has two hidden layers to process the data and the mathematical implementation of the LSTM is:

$$l^t = \tanh\left(w_z X^t + M_z d^{t-1} + n_z\right) \tag{22a}$$

$$g^t = \sigma\left(w_l X^t + M_l d^{t-1} + V_l \odot a^{t-1} + n_l\right) \tag{22b}$$

$$e^t = \sigma\left(w_e X^t + M_f d^{t-1} + V_f \odot a^{t-1} + n_f\right) \tag{22c}$$

$$a^t = g^t \odot l^t + e^t \odot a^{t-1} \tag{22d}$$

$$h^t = \sigma\left(w_h X^t + M_h d^{t-1} + V_h \odot a^t + n_h\right) \tag{22e}$$

$$d^t = h^t \odot \tanh\left(a^t\right) \tag{22f}$$

In the described model, the logistic sigmoid function, denoted as σ, is applied to process the shifted delta coefficient g^t as an input. The desired output vectors h^t are generated based on the input provided. The LSTM (Long Short-Term Memory) architecture, being a memory-based neural network, utilizes the forget gate e^t to determine which information should be retained and which should be disregarded. The internal state vectors are represented by a^t, while n represents the bias vectors. The weight vectors V and recurrent weight matrices M are used in the computation. The input vector at time t is denoted as X^t, and w represents the input weight vectors. The activation function tanh and the element-wise product $\odot$ are employed to address the gradient vanishing problem commonly encountered in RNNs.

Where the logistic sigmoid function is σ, the shifted delta coefficient g^t is fed as an input, the desired output vectors h^t is generated for the applied input, since LSTM is a memory based neural network the value of forget gate e^t helps to decide which information has to be memorized and which one ignore, a^t refers to internal state vectors, n to bias vectors, V is weight vectors and M refers to the recurrent weight matrices, X^t is input vector at time t, w are input weight vectors, tanh and $\odot$ is the element-wise product of the vectors which helps in overcome the gradient vanishing problem in RNN.

The multi-layered RNN-based Bi-LSTM model, as depicted in Figure 10, is structured in this manner. The dataset is initially trained and tested, allowing past data to be analyzed in order to predict future data. Subsequently, a two-layered Bi-LSTM model is trained to predict the occurrence of cardiovascular disease (CVD) based on the phonocardiogram (PCG) signals. This approach leverages the power of the Bi-LSTM architecture to capture temporal dependencies and make accurate predictions regarding the presence of CVD from the provided PCG data.

F. FEATURE EXTRACTION AND SELECTION

In the realm of deep learning, the extraction and selection of desired features pose significant challenges. Signals such as sound, including speech, heartbeat, and pulmonary sounds, require particular attention to maintain the long-term dependencies between each frame throughout the signal.

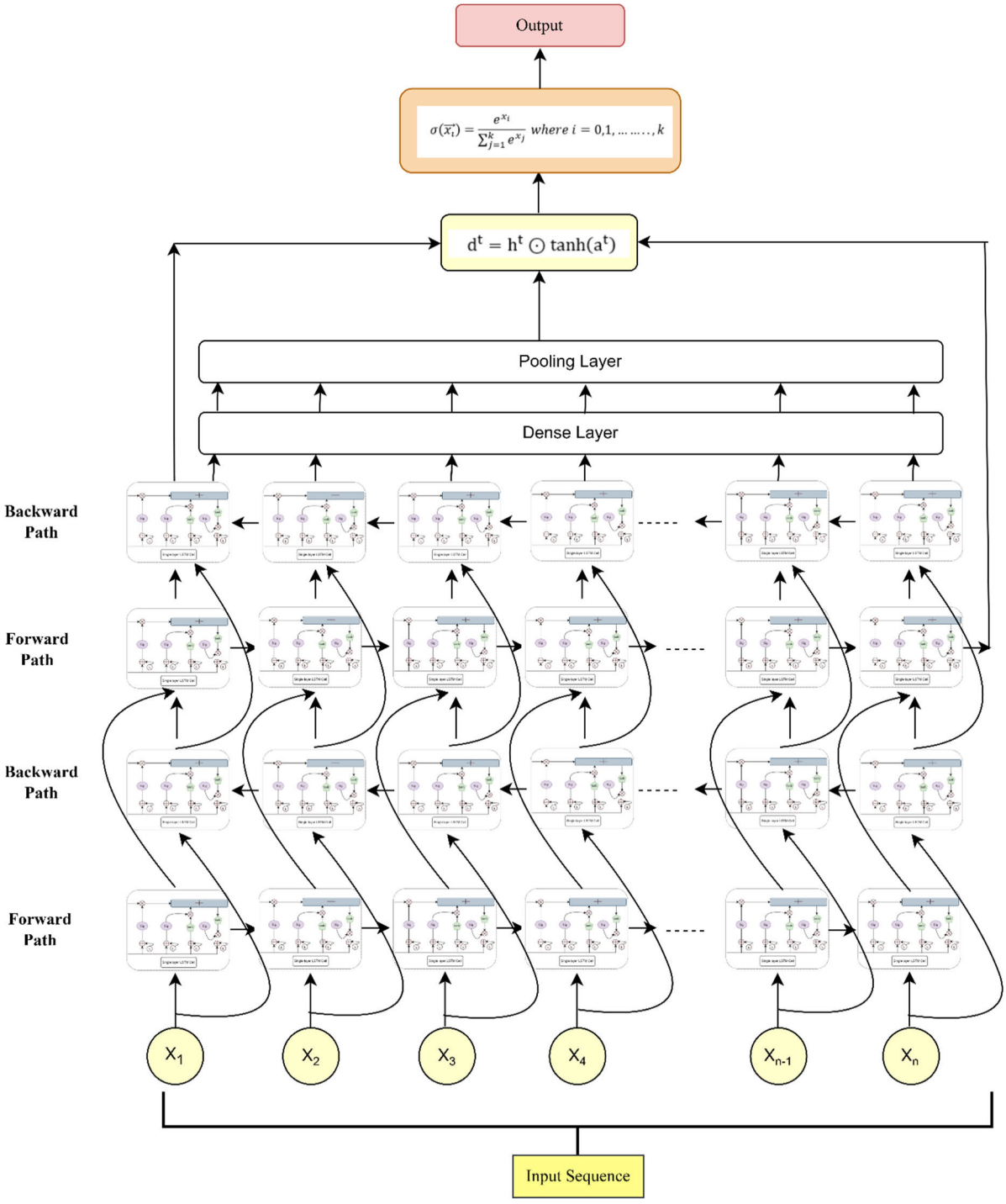


FIGURE 10. Multi-layered RNN based Bi-LSTM.

Time series neural networks, such as RNN-based Bi-LSTM, generate a large set of features from the input samples. Therefore, a pooling process becomes crucial in selecting the desired features from the neural network.

In this proposed work, a comparative study of five pooling functions by [41] and [42] was considered. Based on these reviews, the authors of this paper decided to adopt pooling in the prediction technique, specifically using the exponential SoftMax function.

The k-dimensional vector serves as the input to the Bidirectional Long Short-Term Memory (Bi-LSTM) Recurrent Neural Network (RNN). These k-dimensional Shifted Delta Coefficient (SDC) vectors are transformed into k-dimensional real vectors, with their sum equating to "1".

This transformation is achieved through the SoftMax function. In the Bi-LSTM RNN, the input to the SoftMax function can take on positive, zero, or negative values. However, the SoftMax function normalizes these values to fall between 0 and 1, aligning the Neural Response Coefficients (NRCs) with the output of the neural network. Thus, the SoftMax function converts the values into a normalized probability distribution, where smaller values correspond to lower probabilities and larger values correspond to higher probabilities. Importantly, the sum of these probabilities always equals 1.

Due to its ability to convert values into a normalized probability distribution, the SoftMax function is widely used for classifying complex single-dimensional signals, such as speech.

The SoftMax is defined mathematically as:

$$x = \frac{\sum_j x_j \cdot e^{x_j}}{\sum_j e^{x_j}} \quad where\ j = 0, 1, \ldots\ldots\ldots, k \tag{23a}$$

The gradient of SoftMax function is:

$$\sigma\left(\vec{x_i}\right) = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}} \quad where\ j = 0, 1, \ldots\ldots\ldots, k \tag{23b}$$

where $\vec{x_i}-$ the input vectors to SoftMax function, expanded up to kth term $(x_0, x_1 \ldots\ldots\ldots x_k)$

x_i- are the input vectors.

$\sum_{j=1}^{k} e^{x_j}-$ is the normalization term which makes sure that the output of the function varies between 0 and 1.

Another, challenging task in this work classifying of healthy and un-healthy heart beat based the features selected. So, based on the values SoftMax layer the healthy and un-healthy event is classified. Finally, at the end we need to predict the following classes of heart beats: N (Normal).

IV. RESULTS AND DISCUSSION

A. DATASET USED FOR EXPERIMENTATION

The analysis two datasets taken from the Kaggle (previously PASCAL) Heart Sound Classification Contest. The first one is called A which contains actual field trials data gathered by means of using a digital stethoscope, usually containing accompanying noise like speech, traffic, or even accidental contacts with microphones when on clothes or skin. Audio file durations are about 1-30s as well. The initial step involved splitting the Pascal competitions datasets into training and test set. Dataset A had four classes (normal, murmur, Extrasystole, and echocardiography artifacts), while dataset B had three classes (normal, murmur and Extrasystole) as shown in Table 2.

B. HEART SOUND ANALYSIS

The 4 types of heart sound and related parameters are the original heart sound is shown in figure 11: (a) Normal (b) Murmur (c) Extrasystole (d) Artifact

The proposed work we followed steps for the recognition of CVD as mentioned in figure 4: (i) Pre-processing (ii) Time-Frequency analysis (iii) Feature extraction (iv) Classification (iv) decision making. In general, the processing of time-series using Machine learning algorithms includes following features: (i) Energy Estimation (ii) Frequency estimation (iii) Bandwidth Estimation (iv) Spectral centroid (v) Spectral roll-off (vi) Features representation using MFCC [43]. The results of each steps is discussed in the following figure 12 and 13.

The analyses of Harmonic Mel Spectrum, Percussive Mel Spectrum, Spectral Centroid, and Spectral Roll-off of a normal heartbeat shown in figure 12 (a), (d), (e), (g), provide insights into the harmonic structure, transient components, tonality, and frequency distribution of the heart sounds, which can be useful in various applications such as heart rate monitoring, heart sound analysis, and cardiac health assessment.

The analysis of Harmonic Mel Spectrum, Percussive Mel Spectrum, Spectral Centroid, and Spectral Roll-off of murmur heart beat shown in figure 12 (b), (c), (f), (h) can provide complementary information about the harmonic structure, transient components, tonality, and frequency distribution of a heart murmur. The Harmonic Mel Spectrum and Percussive Mel Spectrum can reveal insights into the harmonic and transient characteristics of the murmur, respectively, while the Spectral Centroid and Spectral Roll-off can provide information about the tonality and frequency distribution of the murmur, respectively. Analyzing these features collectively can aid in identifying the underlying characteristics of a heart murmur and can be helpful in the diagnosis and management of cardiac conditions, under the guidance of a medical professional.

The frequency analysis parameters like: Harmonic Mel Spectrum, Percussive Mel Spectrum, Spectral Centroid, and Spectral Roll-Off of artifact heart beat shown in figure 13 (a), (d), (e), (g), are different techniques used for analyzing the artifact heartbeat signal from different perspectives, including harmonic content, percussive or transient characteristics, spectral balance, and high-frequency content. These techniques can provide valuable insights into the acoustic properties and characteristics of the artifact heartbeat signal, and can be used in various applications such as heartbeat analysis, heart rate monitoring, and physiological signal processing.

In figure 13 (b),(c),(f),(h) the Harmonic Mel Spectrum and Percussive Mel Spectrum focus on the harmonic and percussive content of the heart beat signal, respectively, while the Spectral Centroid and Spectral Roll-off provide information about the frequency characteristics of the signal, such as the perceived pitch and energy distribution. These features can be useful in analyzing the spectral properties of an Extrasystole heart beat and may provide insights into the underlying physiological or pathological conditions associated with the abnormal heartbeat.

The Energy and power spectrum are two different features that can be computed from heart beat signals, including

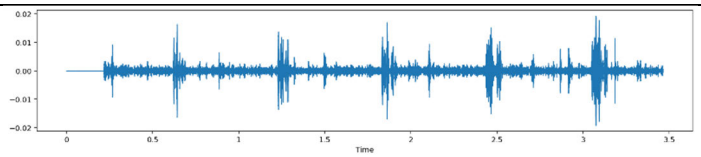
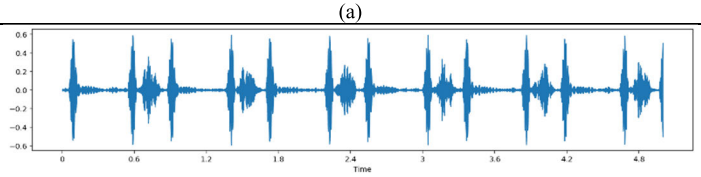
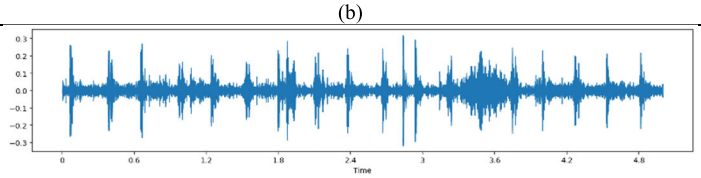
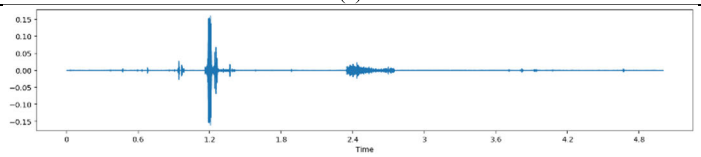
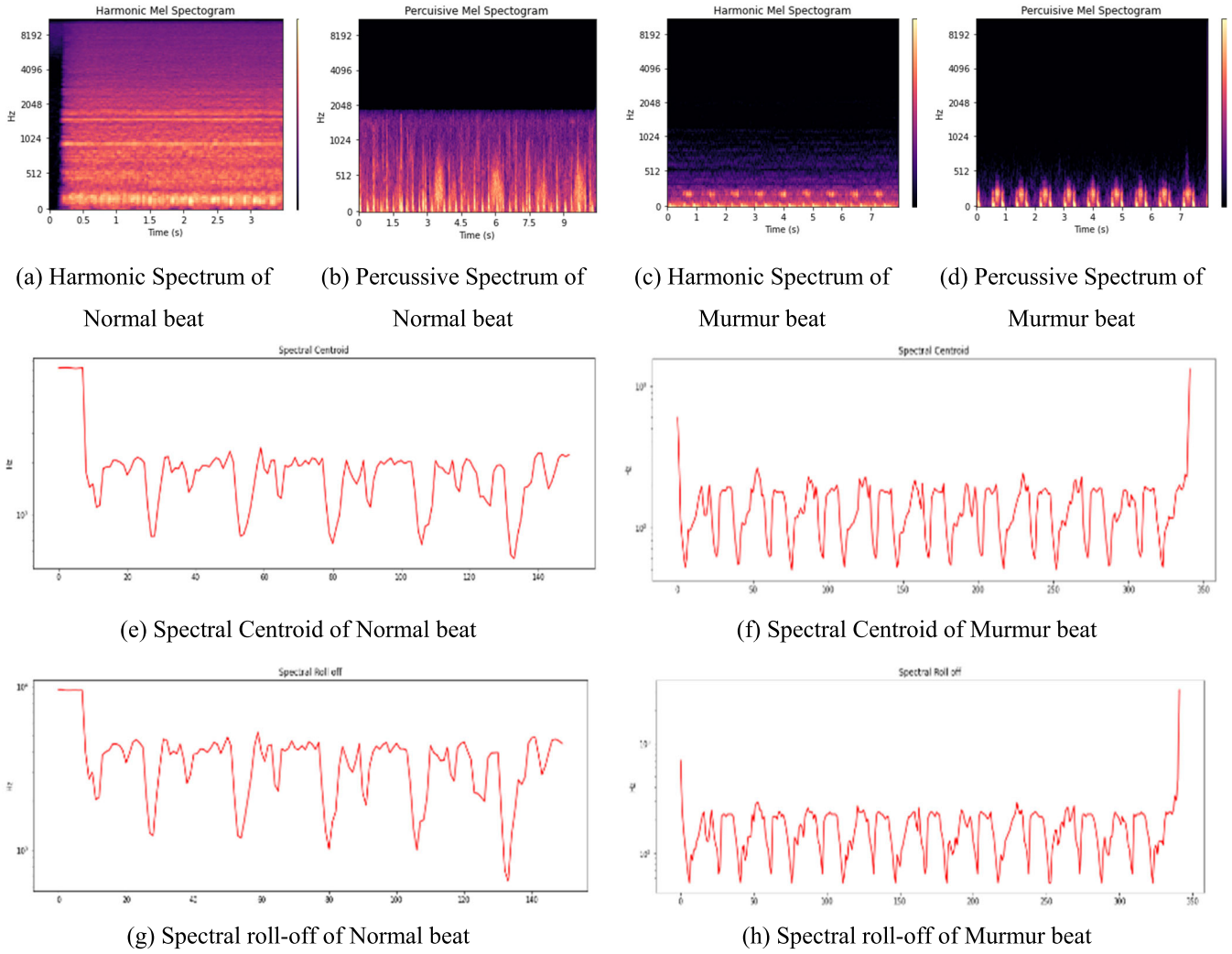
Heart beat wave form	Duration	Sample Rate	No. of Samples
 (a)	3.46	44100	152737
 (b)	7.935	44100	349958
 (c)	10.408	41635	349958
 (d)	9.0	44100	396900

FIGURE 11. Four types of heart sound waveform and its related parameters.

TABLE 2. Number file is A & B dataset-PASCAL database.

Category	Normal	Murmur	Artifact	Extrasystole	Arrhythmia	Frequency (Hz)	Bit size
Dataset A(Testing)	31	34	40	19	66	44100	16
Dataset B(Training)	200	95	-	45	-	44100	16

normal, murmur, artifact, and Extrasystole heartbeats and the results of energy and power spectrum of 4 types of heart beat is discussed in figure 14. In figure 14 (a) the energy and power spectrum of normal heart beat is estimated, where the energy of a normal heartbeat signal would typically be moderate, reflecting the regular and consistent nature of a healthy heart's rhythmic contractions. The power spectrum of a normal heartbeat signal would typically exhibit a dominant frequency at the heart rate or fundamental frequency, with smaller energy contributions at harmonics or other frequency components. A murmur is an abnormal sound heard during a heartbeat, which can be caused by various physiological or pathological conditions. The energy of a murmur heartbeat signal may vary depending on the severity and characteristics of the murmur. It can range from low to high, depending on the intensity and duration of the murmur sound. The power spectrum of a murmur heartbeat signal may exhibit additional frequency components or peaks, corresponding to the abnormal sounds generated by the murmur. These additional peaks may vary in frequency, amplitude, and duration depending on the characteristics of the murmur as shown in the figure 14 (b). Artifact refers to any unwanted or extraneous signal that may be introduced during the recording or processing of a heart beat signal, such as noise or interference. The energy of an artifact heartbeat signal would depend on the nature and level of the artifact, which can vary widely. The power spectrum of an artifact heartbeat signal may exhibit irregular or spurious frequency components, corresponding to the unwanted or extraneous signals introduced during the recording or processing of the signal as shown in the figure 14 (c). These frequency components may not be related to the physiological characteristics of the heart. Extrasystole is an abnormal heart rhythm characterized by premature or irregular heartbeats. The energy of an Extrasystole heartbeat signal may vary depending on the timing, frequency, and intensity of the Extrasystole events. It can be low to moderate, depending on the severity and frequency of the Extrasystole occurrences. The power spectrum of an Extrasystole heartbeat signal may exhibit irregular or abnormal frequency components, corresponding to the premature or irregular heartbeats as shown in figure 14 (d). These frequency components may deviate from the normal heart rate or fundamental frequency, indicating the abnormal rhythm of the Extrasystole events.



(a) Harmonic Spectrum of Normal beat

(b) Percussive Spectrum of Normal beat

(c) Harmonic Spectrum of Murmur beat

(d) Percussive Spectrum of Murmur beat

(e) Spectral Centroid of Normal beat

(f) Spectral Centroid of Murmur beat

(g) Spectral roll-off of Normal beat

(h) Spectral roll-off of Murmur beat

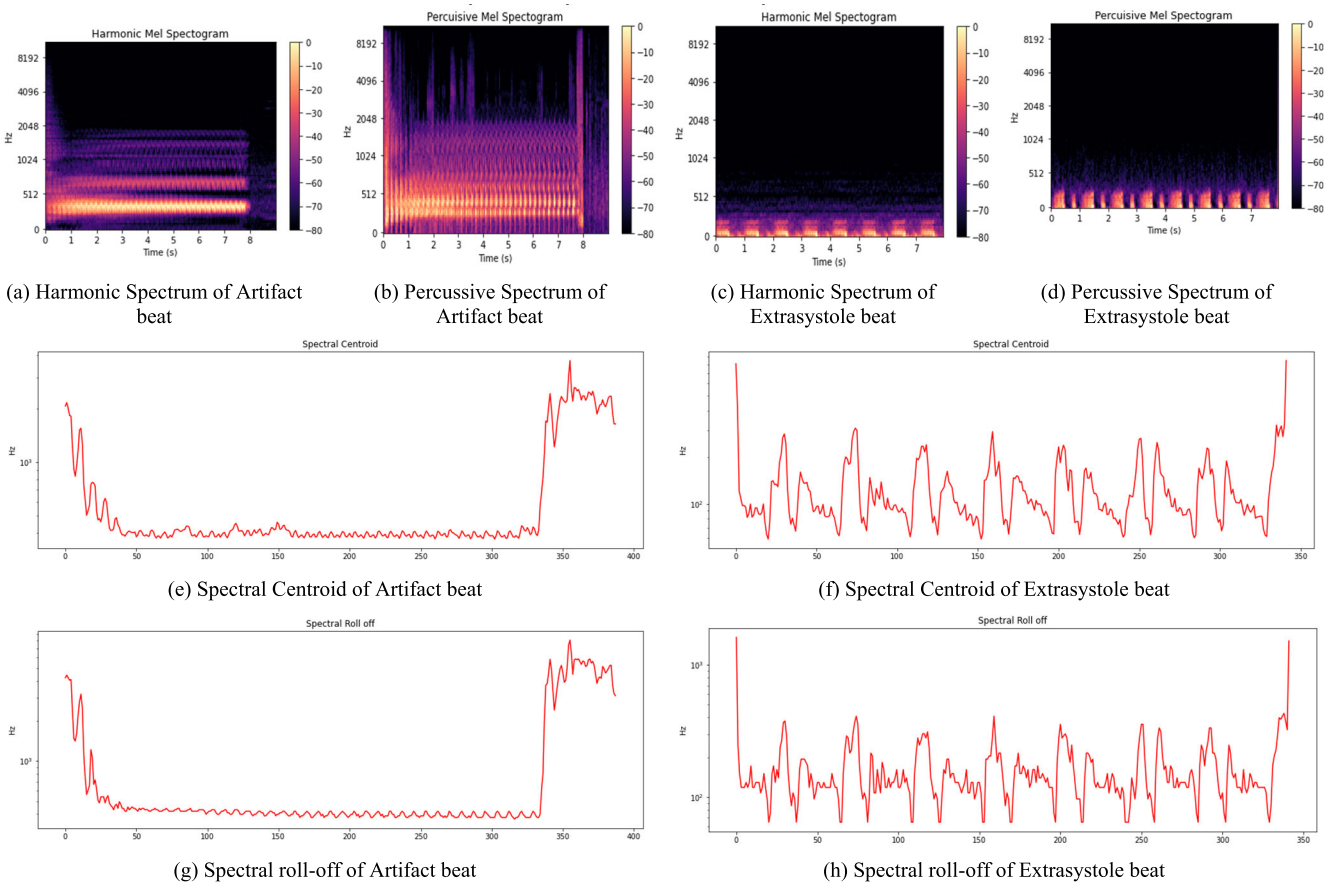
FIGURE 12. **Analysis of harmonic mel-spectrum, percussive mel-spectrum, spectral centroid and spectral roll-off of normal and murmur heart beat.**

Onset detection of a normal heartbeat refers to the process of identifying the beginning or onset of a heartbeat from a physiological signal, such as an electrocardiogram (ECG) or a phonocardiogram (PCG). The onset of a normal heartbeat typically corresponds to the point where the electrical or acoustic activity generated by the heart starts to deviate from the baseline, indicating the contraction of the heart muscle.

In figure 15, the ONSET calculation is done for heart sound-based power spectrum. Power spectrum analysis is a technique used to analyze the frequency components present in a signal. In figure 12(a) the context of a normal heartbeat, it can help identify the dominant frequency associated with the cardiac cycle. Onset Detection: Monitor the changes in the dominant frequency across successive windows. The onset of a normal heartbeat can be detected by identifying a significant increase or change in the dominant frequency.

Detecting the onset of a murmur heartbeat using power spectrum analysis can be a useful technique to identify abnormal heart sounds associated with heart murmurs. As shown in the figure 12 (b), Heart murmurs typically occur in specific frequency bands, such as low-frequency (below 200 Hz) or high-frequency (above 200 Hz) ranges. Peaks in the power spectrum correspond to frequency components that are prominent in the audio signal.

Monitor the power spectrum across successive frames or time segments. Look for changes or increases in the power or amplitude of the frequency components associated with heart murmurs. Significant changes or sudden increases in power can indicate the onset of a murmur. The onset detection of artifact heart sound is shown in figure 12(c), detecting the onset of an artifact heartbeat using power spectrum analysis can be challenging since artifacts are typically unwanted interference or noise in the recorded cardiac signal. However, if the artifact exhibits distinct frequency characteristics (50 or 60 Hz), it might be possible to detect its onset using power spectrum analysis. Analyze the power spectrum to identify any frequency peaks or components associated with the artifact. If the artifact has distinct and



(a) Harmonic Spectrum of Artifact beat
(b) Percussive Spectrum of Artifact beat
(c) Harmonic Spectrum of Extrasystole beat
(d) Percussive Spectrum of Extrasystole beat
(e) Spectral Centroid of Artifact beat
(f) Spectral Centroid of Extrasystole beat
(g) Spectral roll-off of Artifact beat
(h) Spectral roll-off of Extrasystole beat

FIGURE 13. Analysis of harmonic mel-spectrum, percussive mel-spectrum, spectral centroid and spectral roll-off of artifact and extrasystole heat beat.

consistent frequency content, it might manifest as a peak or energy concentration in the power spectrum at that particular frequency. The onset detection of extrastole heart sound is shown in figure 12 (d), detecting the onset of an ectopic or premature heartbeat, specifically an extrasystole or premature ventricular contraction (PVC), using power spectrum analysis can provide insights into the abnormal rhythm. The frequency components of the Extrasystoles beat often exhibit different frequency characteristics compared to the regular cardiac rhythm. Look for significant differences or prominent peaks in the power spectrum of the ectopic beat.

In figure 15 (e) to (h) different onset is calculated to identify the abnormalities in heart sound. In general, the term ''onset'' typically refers to the beginning or initial occurrence of an event or sound. In the context of heart sounds, there are different types of onsets that can be discussed: onset of heart sound, raw onset of heart sound, and backtracked onset of heart sound.

1) ONSET OF HEART SOUND

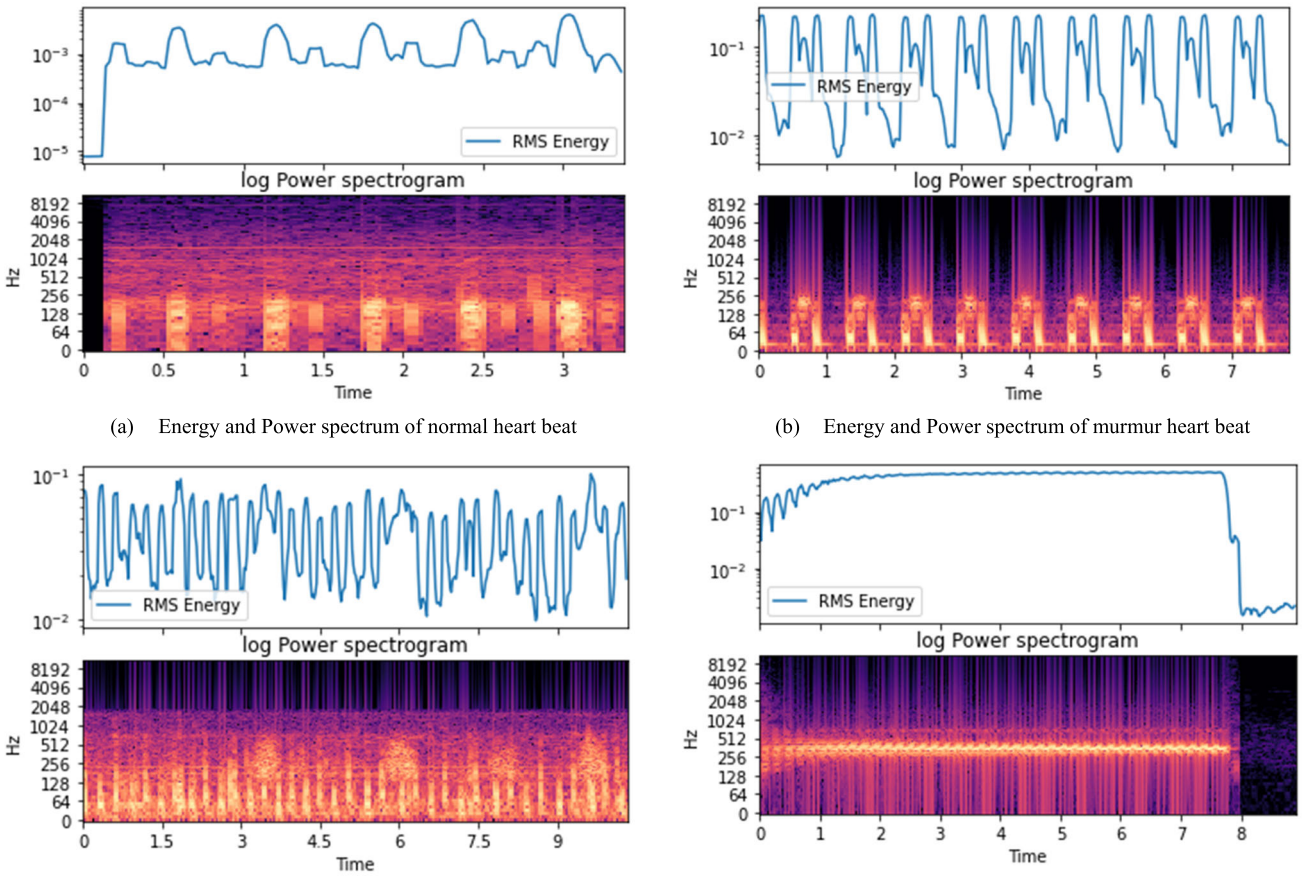
The onset of a heart sound refers to the exact moment when a specific heart sound component starts. In a typical cardiac cycle, there are four primary heart sounds: S1, S2, S3, and S4. The onset of S1 represents the beginning of the systole (contraction) phase, while the onset of S2 marks the beginning of the diastole (relaxation) phase. The onsets of S3 and S4, if present, can indicate abnormal heart sounds related to ventricular filling.

2) RAW ONSET OF HEART SOUND

The raw onset of a heart sound refers to the moment when the sound wave associated with a specific heart sound component reaches a detectable level above the baseline noise. This onset is typically determined by analyzing the amplitude or intensity of the sound waveform recorded during auscultation or with other measurement devices. The raw onset can be measured using various techniques, including visual inspection, signal processing algorithms, or amplitude threshold-based methods.

3) BACKTRACKED ONSET OF HEART SOUND

The backtracked onset of a heart sound refers to the estimated or reconstructed onset of a specific heart sound component based on post-processing analysis. It involves using advanced signal processing techniques to analyze the heart sound waveform, extract relevant features, and determine the likely point of onset. This estimation may



(a) Energy and Power spectrum of normal heart beat

(b) Energy and Power spectrum of murmur heart beat

(c) Energy and Power spectrum of artifact heart beat

(d) Energy and Power spectrum of Extrasystole heart beat

FIGURE 14. Energy and power spectrum of 4 type heart beat.

involve sophisticated algorithms that consider the temporal and spectral characteristics of the heart sound signal.

It's important to note that the precise determination of the onsets of heart sounds can be challenging due to various factors, such as the presence of background noise, overlapping sounds, and variations in individual physiology. Additionally, the interpretation and diagnosis of heart sounds require clinical expertise and should be performed by healthcare professionals, as they consider multiple factors, including the entire phonocardiogram (PCG) signal, patient history, and other clinical information.

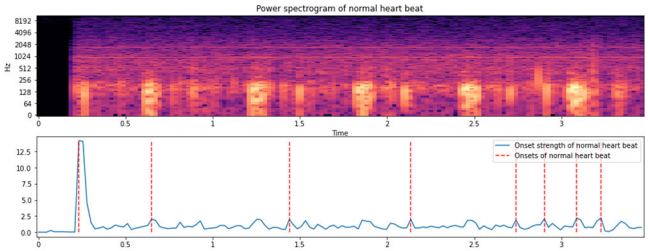
In clinical practice, specialized tools such as phonocardiography and computer-aided analysis can assist in the accurate determination of heart sound onsets. These tools help in detecting and analyzing the timing and characteristics of heart sounds to aid in the diagnosis and assessment of cardiac conditions.

The Constant Q Transform (CQT) is a specific type of time-frequency analysis method, which is often used for analyzing non-stationary signals, such as heartbeats. While the CQT itself does not directly provide mean and median values, it can be used as a tool to extract features from the signal that can subsequently be used to calculate these statistics for normal heartbeats. By applying CQT to each selected normal heartbeat signal, it provides a time-frequency representation of the signal, capturing the frequency content and its evolution over time.

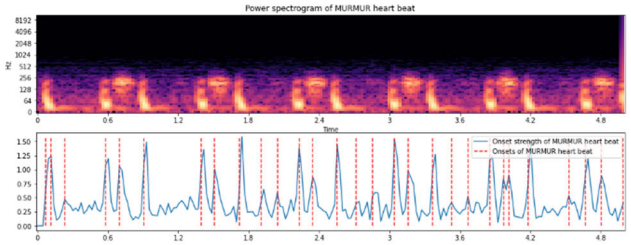
Then extract relevant features from the CQT representation of the normal heartbeats. These features can include statistical measures such as mean and median values. Calculate the mean and median values of the CQT coefficients across the time and frequency dimensions for each normal heartbeat. Finally, the mean and median values are calculated for the extracted features (e.g., mean or median CQT coefficients) across all the normal heartbeats as shown in figure 16. These statistics provide quantitative measures of the time-frequency characteristics of normal heartbeats.

TABLE 3. Analysis of Confusion matrix in percentage for each class.

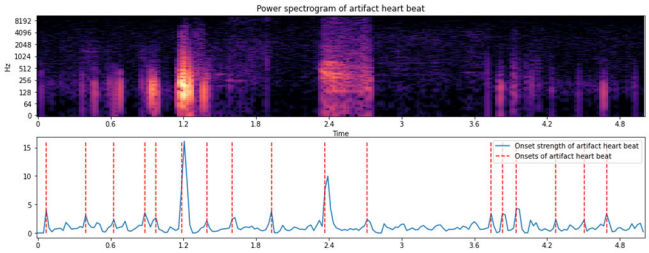
Class	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Artifact	100	89	94	95
Extrasystole	92	100	96	97
Murmur	90	100	95	97
Normal	100	94	97	97
Average Accuracy				**96.5**



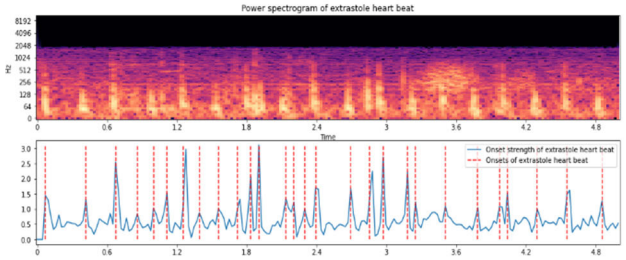
a) Power spectrum-based ONSET strength of normal heart beat



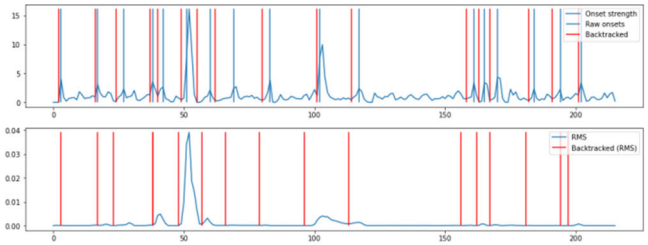
b) Power spectrum-based ONSET strength of murmur heart beat



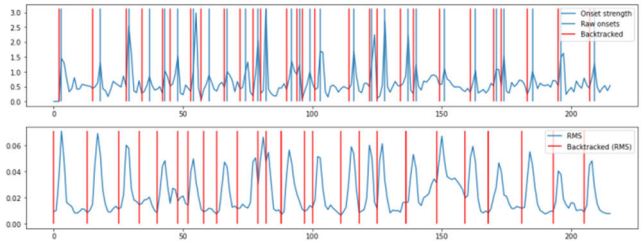
c) Power spectrum-based ONSET strength of artifact heart beat



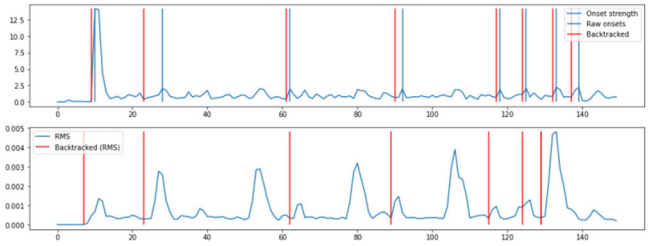
d) Power spectrum-based ONSET strength of Extrasystole heart beat



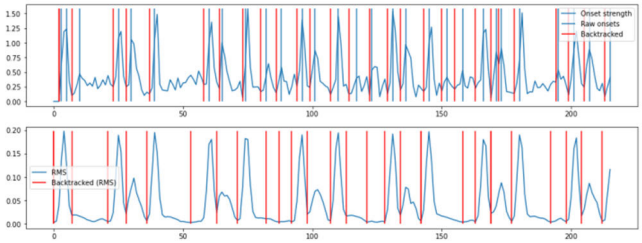
e) ONSET and it RMS value of Normal heart sound



f) ONSET and it RMS value of Murmur heart sound



g) ONSET and it RMS value of Artifact heart sound



h) ONSET and it RMS value of Extrasystole heart sound

FIGURE 15. **ONSET and RMS detection for Heart sound.**

The Confusion matrix with Artifact, Extrasystole, and Murmur and Normal values is as shown in figure 17. The comprehensive evaluation of the performance of a multi-class model is provided in the classification report shown in the Table 3. The explanation for the above report is briefed below.

a: PRECISION

Artifact: When the model is predicted the sample the as artifact in almost all instances then the Precision is 100%. But, the recall value 89% indicates that the prediction of artifact is wrong in some instances.

Extrasystole: The proposed model apprehended the all actual Extrasystole cases by recall as 100% but the Precision is 92%, meaning that 92% of predicted Extrasystole instances are correct.

Murmur: Similar to Extrasystole the model has captures all murmur cases but prediction (Precision) is 90%, and recall is 100%, similar to Extrasystole, implying good performance in identifying Murmur.

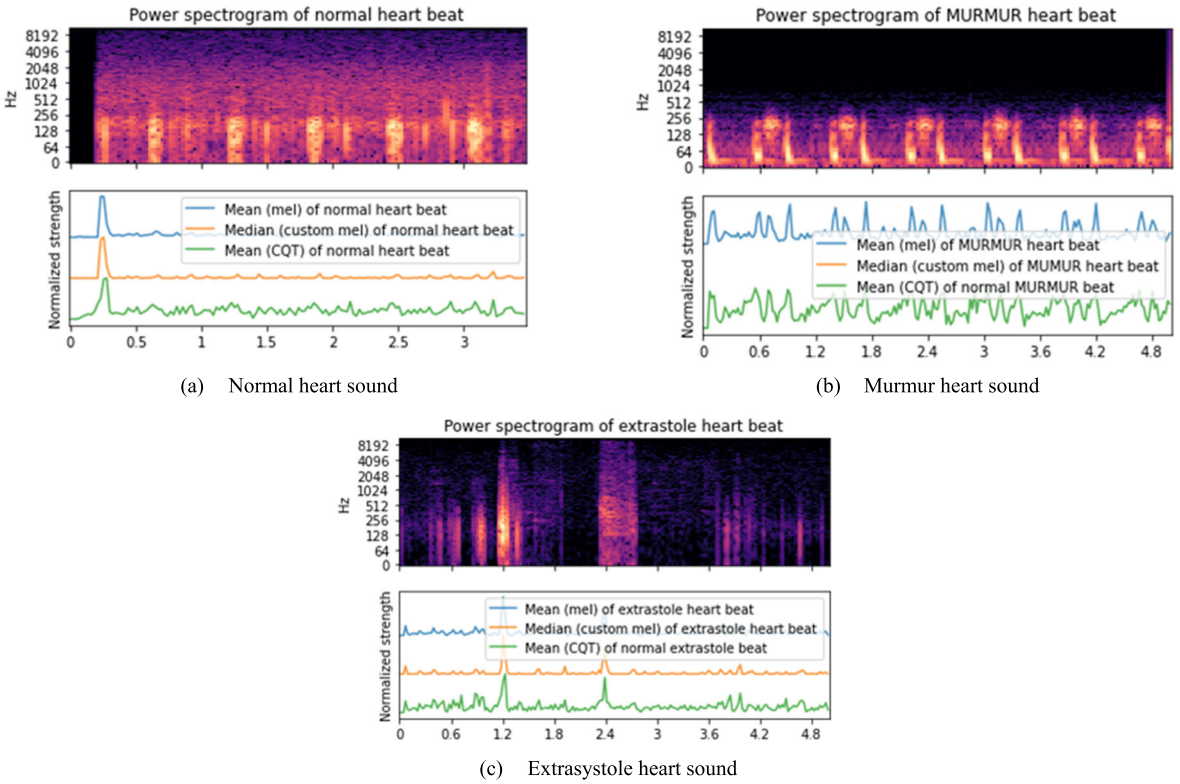
Normal: Compare to above three cases the prediction of normal cases is 94 % indicates that the model rarely miss classifies the normal heart sound.

b: RECALL

Artifact: 89% recall implies that the model correctly identifies 89% of all actual Artifact instances.

Extrasystole: 100% recall indicates that the model captures all true Extrasystole cases.

Murmur: 100% recall suggests that the model successfully identifies all actual Murmur instances.



(a) Normal heart sound (b) Murmur heart sound

(c) Extrasystole heart sound

FIGURE 16. Constant Q Transform (CQT), median and mean of heartbeat.

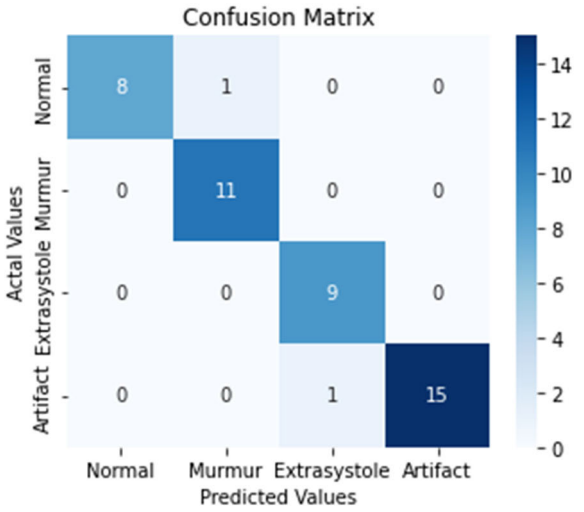


FIGURE 17. Confusion matrix.

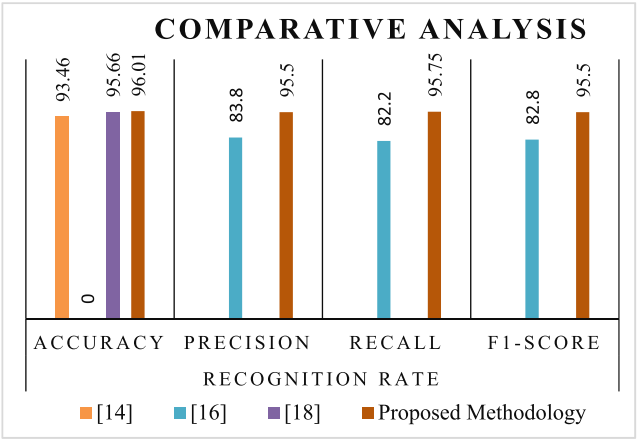


FIGURE 18. Comparative analysis of proposed methodology.

Normal: 94% recall means that the model correctly identifies 94% of all actual Normal instances.

c: F1-SCORE

F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall. All classes have high F1-scores, indicating a good balance between precision and recall for each class.

d: SUPPORT

The number of instances for each class in the dataset.

e: ACCURACY

The overall accuracy of the model across all classes is 96%, demonstrating its effectiveness in making correct predictions.

f: MACRO AVG AND WEIGHTED AVG

The macro average gives equal weight to each class, while the weighted average considers the number of instances for each class. Both macro and weighted averages of

precision, recall, and F1-score are high, reflecting the overall good performance of the model across all classes. Finally, the classification report justifies that the proposed model classifies the 4 classes effectively with high precision, recall, and F1-score.

The results of the proposed methodology for the recognition of irregularities from phonocardiogram signals is compared with results of [14], [16], and [18] only, because rest of the authors in the reference are concentrated more on ECG signal. From figure 18, it is clear that the proposed methodology is giving better results compare to previous methods in the field of PCH analysis. The accuracy of [18] and proposed methodology seems almost equal, but the process adapted in analysis in the proposed method is completely different and the time series analysis can be done accurately in the proposed work.

V. CONCLUSION

The RNN-Bi LSTM based multi decision GAN approach presented in this study demonstrates a promising solution for the recognition of cardiovascular disease (CVD) from heartbeat sound which is the significant in the present medical field. The research focuses on feature optimization, aiming to enhance the accuracy and efficiency of CVD recognition. By utilizing the power of recurrent neural networks (RNNs) and bidirectional long short-term memory (Bi LSTM), the proposed approach effectively captures the temporal dependencies and long-term context in heartbeat sound data. This enables the model to learn intricate patterns and relationships that are crucial for CVD detection. Furthermore, the integration of a Multi Decision GAN framework adds an additional layer of sophistication to the approach. This framework promotes adversarial training and generates diverse decision outputs, leading to improved robustness and accuracy in CVD recognition. The feature optimization process conducted in this study plays a vital role in enhancing the overall performance of the model. By carefully selecting and engineering relevant features from heartbeat sound data, the approach achieves better discriminative capabilities, reducing the risk of misclassification and false negatives. The experimental results demonstrate the superiority of the proposed approach compared to traditional methods and other state-of-the-art techniques. The RNN-Bi LSTM based Multi Decision GAN achieves remarkable accuracy, sensitivity, and specificity in CVD recognition, making it a valuable tool for early detection and diagnosis of cardiovascular diseases. Finally, the study presents a novel and effective approach for CVD recognition by combining RNN-Bi LSTM architecture, Multi Decision GAN framework, and feature optimization techniques. The proposed model exhibits significant potential for improving the accuracy and efficiency of CVD diagnosis, ultimately contributing to better healthcare outcomes and saving lives. But the proposed work is limited one datasets and as future work this extended to test and validate various samples from different dataset and also for pulmonary sound analysis.

REFERENCES

[1] F. S. Butt, M. F. Wagner, J. Schäfer, and D. G. Ullate, "Toward automated feature extraction for deep learning classification of electrocardiogram signals," *IEEE Access*, vol. 10, pp. 118601–118616, 2022, doi: 10.1109/ACCESS.2022.3220670.

[2] Z.-H. Wang and Y.-C. Wu, "A novel rapid assessment of mental stress by using PPG signals based on deep learning," *IEEE Sensors J.*, vol. 22, no. 21, pp. 21232–21239, Nov. 2022, doi: 10.1109/JSEN.2022.3208427.

[3] T. H. Rafi and Y. Woong Ko, "HeartNet: Self multihead attention mechanism via convolutional network with adversarial data synthesis for ECG-based arrhythmia classification," *IEEE Access*, vol. 10, pp. 100501–100512, 2022, doi: 10.1109/ACCESS.2022.3206431.

[4] A. Nasim, D. C. Nchekwube, F. Munir, and Y. S. Kim, "An evolutionary-neural mechanism for arrhythmia classification with optimum features using single-lead electrocardiogram," *IEEE Access*, vol. 10, pp. 99050–99065, 2022, doi: 10.1109/ACCESS.2022.3203586.

[5] C. Zou, A. Müller, U. Wolfgang, D. Rückert, P. Müller, M. Becker, A. Steger, and E. Martens, "Heartbeat classification by random forest with a novel context feature: A segment label," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–8, 2022, doi: 10.1109/JTEHM.2022.3202749.

[6] A. Bhattarai, D. Peng, J. Payne, and H. Sharif, "Adaptive partition of ECG diagnosis between cloud and wearable sensor net using open-loop and closed-loop switch mode," *IEEE Access*, vol. 10, pp. 63684–63697, 2022, doi: 10.1109/ACCESS.2022.3182704.

[7] M. S. Islam, M. N. Islam, N. Hashim, M. Rashid, B. S. Bari, and F. A. Farid, "New hybrid deep learning approach using BiGRU-BiLSTM and multilayered dilated CNN to detect arrhythmia," *IEEE Access*, vol. 10, pp. 58081–58096, 2022, doi: 10.1109/ACCESS.2022.3178710.

[8] L.-H. Wang, Y.-T. Yu, W. Liu, L. Xu, C.-X. Xie, T. Yang, I.-C. Kuo, X.-K. Wang, J. Gao, P.-C. Huang, S.-L. Chen, W.-Y. Chiang, and P. A. R. Abu, "Three-heartbeat multilead ECG recognition method for arrhythmia classification," *IEEE Access*, vol. 10, pp. 44046–44061, 2022, doi: 10.1109/ACCESS.2022.3169893.

[9] T. Kitagawa, K. Yamamoto, K. Endo, and T. Ohtsuki, "Multi-beam Doppler sensor-based non-contact heartbeat detection using beam diversity," *IEEE Access*, vol. 10, pp. 16242–16253, 2022, doi: 10.1109/ACCESS.2022.3148426.

[10] X. Liang, H. Li, A. Vuckovic, J. Mercer, and H. Heidari, "A neuromorphic model with delay-based reservoir for continuous ventricular heartbeat detection," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 1837–1849, Jun. 2022, doi: 10.1109/TBME.2021.3129306.

[11] L.-H. Wang, L.-J. Ding, C.-X. Xie, S.-Y. Jiang, I.-C. Kuo, X.-K. Wang, J. Gao, P.-C. Huang, and P. A. R. Abu, "Automated classification model with Otsu and CNN method for premature ventricular contraction detection," *IEEE Access*, vol. 9, pp. 156581–156591, 2021, doi: 10.1109/ACCESS.2021.3128736.

[12] E. Maghawry, T. F. Gharib, R. Ismail, and M. J. Zaki, "An efficient heartbeats classifier based on optimizing convolutional neural network model," *IEEE Access*, vol. 9, pp. 153266–153275, 2021, doi: 10.1109/ACCESS.2021.3128134.

[13] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," *IEEE Access*, vol. 9, pp. 103452–103464, 2021, doi: 10.1109/ACCESS.2021.3098986.

[14] H. Yang and Z. Wei, "A novel approach for heart ventricular and atrial abnormalities detection via an ensemble classification algorithm based on ECG morphological features," *IEEE Access*, vol. 9, pp. 54757–54774, 2021, doi: 10.1109/ACCESS.2021.3071273.

[15] J.-S. Kim and K. Lee, "Untact abnormal heartbeat wave detection using non-contact sensor through transfer learning," *IEEE Access*, vol. 8, pp. 217791–217799, 2020, doi: 10.1109/ACCESS.2020.3042643.

[16] H. Dang, Y. Yue, D. Xiong, X. Zhou, X. Xu, and X. Tao, "A deep biometric recognition and diagnosis network with residual learning for arrhythmia screening using electrocardiogram recordings," *IEEE Access*, vol. 8, pp. 153436–153454, 2020, doi: 10.1109/ACCESS.2020.3016938.

[17] F. Qiao, B. Li, Y. Zhang, H. Guo, W. Li, and S. Zhou, "A fast and accurate recognition of ECG signals based on ELM-LRF and BLSTM algorithm," *IEEE Access*, vol. 8, pp. 71189–71198, 2020, doi: 10.1109/ACCESS.2020.2987930.

[18] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020, doi: 10.1109/JBHI.2020.2981526.

[19] H. Yang and Z. Wei, "Arrhythmia recognition and classification using combined parametric and visual pattern features of ECG morphology," *IEEE Access*, vol. 8, pp. 47103–47117, 2020, doi: 10.1109/ACCESS.2020.2979256.
[20] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020, doi: 10.1109/ACCESS.2020.2974712.
[21] J. Wu, F. Li, Z. Chen, Y. Pu, and M. Zhan, "A neural network-based ECG classification processor with exploitation of heartbeat similarity," *IEEE Access*, vol. 7, pp. 172774–172782, 2019, doi: 10.1109/ACCESS.2019.2956179.
[22] Q. Xie, S. Tu, G. Wang, Y. Lian, and L. Xu, "Feature enrichment based convolutional neural network for heartbeat classification from electrocardiogram," *IEEE Access*, vol. 7, pp. 153751–153760, 2019, doi: 10.1109/ACCESS.2019.2948857.
[23] J. Zheng, M. Liang, S. Sinha, L. Ge, W. Yu, A. Ekstrom, and F. Hsieh, "Time-frequency analysis of scalp EEG with Hilbert–Huang transform and deep learning," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1549–1559, Apr. 2022, doi: 10.1109/JBHI.2021.3110267.
[24] Y. Zhang, J. Li, S. Wei, F. Zhou, and D. Li, "Heartbeats classification using hybrid time-frequency analysis and transfer learning based on ResNet," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 11, pp. 4175–4184, Nov. 2021, doi: 10.1109/JBHI.2021.3085318.
[25] C. Lastre-Domínguez, Y. S. Shmaliy, O. Ibarra-Manzano, and M. Vazquez-Olguin, "Denoising and features extraction of ECG signals in state space using unbiased FIR smoothing," *IEEE Access*, vol. 7, pp. 152166–152178, 2019, doi: 10.1109/ACCESS.2019.2948067.
[26] X. Zhang, R. Li, H. Dai, Y. Liu, B. Zhou, and Z. Wang, "Localization of myocardial infarction with multi-lead bidirectional gated recurrent unit neural network," *IEEE Access*, vol. 7, pp. 161152–161166, 2019, doi: 10.1109/ACCESS.2019.2946932.
[27] E. K. Wang, X. Zhang, and L. Pan, "Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network," *IEEE Access*, vol. 7, pp. 182873–182880, 2019, doi: 10.1109/ACCESS.2019.2936525.
[28] R. Li, X. Zhang, H. Dai, B. Zhou, and Z. Wang, "Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and BiLSTM-attention neural network," *IEEE Access*, vol. 7, pp. 109870–109883, 2019, doi: 10.1109/ACCESS.2019.2933473.
[29] K. Wolk and A. Wolk, "Early and remote detection of possible heartbeat problems with convolutional neural networks and multipart interactive training," *IEEE Access*, vol. 7, pp. 145921–145927, 2019, doi: 10.1109/ACCESS.2019.2919485.
[30] T. Pokaprakarn, R. R. Kitzmiller, R. Moorman, D. E. Lake, A. K. Krishnamurthy, and M. R. Kosorok, "Sequence to sequence ECG cardiac rhythm classification using convolutional recurrent neural networks," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 2, pp. 572–580, Feb. 2022, doi: 10.1109/JBHI.2021.3098662.
[31] H. Dang, M. Sun, G. Zhang, X. Qi, X. Zhou, and Q. Chang, "A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals," *IEEE Access*, vol. 7, pp. 75577–75590, 2019, doi: 10.1109/ACCESS.2019.2918792.
[32] W. Sun, N. Zeng, and Y. He, "Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network," *IEEE Access*, vol. 7, pp. 67123–67129, 2019, doi: 10.1109/ACCESS.2019.2918361.
[33] B. Hu, S. Wei, D. Wei, L. Zhao, G. Zhu, and C. Liu, "Multiple time scales analysis for identifying congestive heart failure based on heart rate variability," *IEEE Access*, vol. 7, pp. 17862–17871, 2019, doi: 10.1109/ACCESS.2019.2895998.
[34] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018, doi: 10.1109/ACCESS.2018.2833841.
[35] P. Lameski, E. Zdravevski, S. Koceski, A. Kulakov, and V. Trajkovik, "Suppression of intensive care unit false alarms based on the arterial blood pressure signal," *IEEE Access*, vol. 5, pp. 5829–5836, 2017, doi: 10.1109/ACCESS.2017.2690380.
[36] V. C. Kagawade and S. A. Angadi, "Savitzky–Golay filter energy features-based approach to face recognition using symbolic modeling," *Pattern Anal. Appl.*, vol. 24, no. 4, pp. 1451–1473, Nov. 2021, doi: 10.1007/s10044-021-00991-z.
[37] S. K. Ghosh, P. R. Nagarajan, and R. K. Tripathy, "Heart sound data acquisition and preprocessing techniques," in *Handbook of Research on Advancements of Artificial Intelligence in Healthcare Engineering*. Hershey, PA, USA: IGI Global, 2020, pp. 244–264, doi: 10.4018/978-1-7998-2120-5.ch014.
[38] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 13, Feb. 2011, doi: 10.1186/1475-925x-10-13.
[39] T. Zan, Z. Liu, Z. Su, M. Wang, X. Gao, and D. Chen, "Statistical process control with intelligence based on the deep learning model," *Appl. Sci.*, vol. 10, no. 1, p. 308, Dec. 2019, doi: 10.3390/app10010308.
[40] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Long short-term memory for machine remaining life prediction," *J. Manuf. Syst.*, vol. 48, pp. 78–86, Jul. 2018, doi: 10.1016/j.jmsy.2018.05.011.
[41] C.-C. Kao, M. Sun, W. Wang, and C. Wang, "A comparison of pooling methods on LSTM models for rare acoustic event classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 316–320, doi: 10.1109/ICASSP40776.2020.9053150.
[42] Y. Wang, J. Li, and F. Metze, "A comparison of five multiple instance learning pooling functions for sound event detection with weak labeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 31–35, doi: 10.1109/ICASSP.2019.8682847.
[43] R. Avanzato and F. Beritelli, "Heart sound multiclass analysis based on raw data and convolutional neural network," *IEEE Sensors Lett.*, vol. 4, no. 12, pp. 1–4, Dec. 2020, doi: 10.1109/LSENS.2020.3039366.
[44] H. Murtaza, M. Ahmed, N. F. Khan, G. Murtaza, S. Zafar, and A. Bano, "Synthetic data generation: State of the art in health care domain," *Comput. Sci. Rev.*, vol. 48, May 2023, Art. no. 100546, doi: 10.1016/j.cosrev.2023.100546.
[45] H. Zhang, P. Zhang, F. Lin, L. Chao, Z. Wang, F. Ma, and Q. Li, "Co-learning-assisted progressive dense fusion network for cardiovascular disease detection using ECG and PCG signals," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122144, doi: 10.1016/j.eswa.2023.122144.
[46] G.-Y. Son and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, Nov. 2018, doi: 10.3390/app8122344.
[47] S. Abbas, S. Ojo, A. Al Hejaili, G. A. Sampedro, A. Almadhor, M. M. Zaidi, and N. Kryvinska, "Artificial intelligence framework for heart disease classification from audio signals," *Sci. Rep.*, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-53778-7.

N. A. VINAY (Member, IEEE) received the bachelor's degree in electronics and communication engineering from VTU Belgaum, India, and the master's degree in digital electronics from SAHE Tumkur. He is currently pursuing the Ph.D. degree with REVA University, Bengaluru, Karnataka, India. He is also an Associate Professor with the Department of Electronics and Communication Engineering, Dayanand Sagar College of Engineering, Bengaluru. His research interests include speech processing, image processing, and signal processing. Additionally, he holds the designation of an Associate Member of the Institution of Electronics and Telecommunication Engineers (AMIETE) in India.

K. N. VIDYASAGAR (Member, IEEE) received the bachelor's degree in electronics and communication engineering and the master's degree in digital electronics and communication from VTU Belgaum, India, and the Ph.D. degree from VTU, Bengaluru, Karnataka, India. Currently, he is a Faculty Member of the School of Electronics and Communication Engineering, REVA University, Bengaluru. His research interests include video processing, image processing, and signal processing. Additionally, he holds the position of a member of the Institution of Electronics and Telecommunication Engineers (MIETE-500838) in India.

S. ROHITH (Member, IEEE) received the joint Bachelor of Engineering degree from the Nagarjuna College of Engineering and Technology, Bengaluru, and in electronics and communication engineering from Visvesvaraya Technological University, Belgaum, in 2006, the Master of Technology degree in VLSI design and embedded systems from the Dr. Ambedkar Institute of Technology, Bengaluru, Visvesvaraya Technological University, Belgaum, in 2008, and the Ph.D. degree in digital image watermarking from Visvesvaraya Technological University, in 2020. He has 14 years of teaching experience and one year of industry experience. He is currently an Associate Professor with the Department of Electronics and Communication, NCET, Bengaluru. He is also working toward in the thrust research area, such as image processing, cryptography, information hiding, and VLSI design. He has published many papers in international journals and conferences. He is a Fellow Member of IETE and a Life Member of ISTE and IAENG.

DAYANANDA PRUTHVIRAJA (Senior Member, IEEE) received the M.Tech. degree from RVCE and the Ph.D. degree from VTU. He is currently a Professor in information technology with Manipal Institute of Technology Bengaluru, MAHE. In previous assignment, he was a Professor and the Head of the Department of Information Science and Engineering, JSSATE, Bengaluru. He has published many articles in national and international journals in the field of image processing and retrieval. He has got few research grants and consultancy into his account. His research interests include image processing and information retrieval.

S. SUPREETH (Member, IEEE) received the Ph.D. degree in computer science and engineering from Visvesvaraya Technological University, India. He is currently an Associate Professor with the School of CSE, REVA University, Bengaluru. He has been a Technology Educator for more than 11 years, teaching various subjects like theory of computation, virtualization and cloud computing, big data and data analytics, and image processing. He has published many papers in reputed international journals and conferences in the domain of cloud, fog computing, image processing, and machine learning. He is a life-time member of CSI, ISTE, and IAENG.

S. H. BHARATHI (Member, IEEE) received the Ph.D. degree in bio sensor's, VLSI and digital signal processing, in 2013. She has over 28 years of experience as a Teacher and a Researcher. The lady acquired her undergraduate and postgraduate training in electronics and communication with Bangalore University, Karnataka, India. She is a Fellow Member of IETE. As well, she belongs to the IEEE Women in Engineering.