ISSN print: 1224-5984
ISSN online: 2248-3942

Rom Biotechnol Lett. 2022; 27(1): 3167-3183
doi: 10.25083/rbl/27.1/3167-3183

Received for publication, October, 11, 2021
Accepted, January, 01, 2022

Original paper

Review of Heart Sound Analyses from Phonocardiogram Records

ALI FATIH GUNDUZ
Malatya Turgut Özal University, Akçadağ Vocational Highschool, 44600 Malatya, Turkey

Abstract Cardiovascular diseases are considered as one of the most common causes of death worldwide. Well-beings of people in the risk groups are monitored by various state-of-the-art tools in clinics and home-care units. Phonocardiograph is one of the them which captures sounds coming from the heart and gives high-quality graphical records (i.e., Phonocardiogram, PCG) of them for examination of pathologies. PCG records have been studied and interpreted in order to localize heart sound segments and classify abnormalities for decades. Moreover, there have been competitions for heart sound classification and researchers have developed successful solutions based on signal processing and machine learning approaches. Main steps of those studies are grouped as preprocessing, segmentation, feature extraction and classification. In this study we present a survey of proposed methods and used datasets. The features used in the literature are listed as time, frequency and time-frequency domains. Performances of different studies are presented and compared. From this perspective, it is concluded that there is still room for automated heart sound analysis. Larger open access PCG databases are required for testing state-of-the-art machine learning methods.

Keywords Phonocardiogram classification, heart sound databases, heart sound analysis, segmentation, feature extraction, cardiovascular monitoring

To cite this article: GUNDUZ AF. Review of Heart Sound Analyses from Phonocardiogram Records. *Rom Biotechnol Lett*. 2022; 27(1): 3167-3183. DOI: 10.25083/rbl/27.1/3167-3183.

*Corresponding author: ALI FATIH GUNDUZ, Malatya Turgut Özal University, Akçadağ Vocational Highschool, 44600 Malatya, Turkey.
E-mail: fatih.gunduz@ozal.edu.tr; fatih.gunduz.ali@gmail.com

Introduction

Heart, blood vessels and blood together form the cardiovascular system. Heart is located in the left of the middle chest surrounded with two lungs and diaphragm muscle. Blood vessels include veins (carrying blood from body to heart), arteries (carrying blood from heart to body) and capillaries (smallest vessels that enable material exchange between body cells and blood). Different sized vessels form a vascular network of approximately 96500 kilometers (60000 miles) [1] in which blood is consistently transported to supply vital elements to body cells. During this circulation process, body cells receive oxygen, nutrients, hormones and get rid of carbon dioxide and cellular waste products.

According to World Health Organization (WHO), diseases of cardiovascular system are the most prominent factors of death globally being the cause of an estimated 31% of all deaths worldwide [2]. In addition to having a high morbidity, they also deteriorate life quality of patients. Most common cardiovascular diseases are myocardial infarction, stroke, Kawasaki disease, high blood pressure (hypertension), high cholesterol, coronary artery disease, cardiomyopathy, rhythm disorders (arrhythmia) and heart structure related congenital defects. Hypertension, diabetes, obesity, smoking, use of alcohol, inadequate physical activity and family history of having CVDs are counted as risk factors. Controlling the risk factors and following a healthy lifestyle are very important for coping with CVDs.

Heart is the core organ of the cardiovascular system pumping and propelling the blood through vessels. Under normal conditions, it usually beats from 60 to 100 times per minute to achieve this task. Depending on the condition of the individual, heart receives messages through hormones, sympathetic and parasympathetic nervous system. According to the needs indicated by those messages, the heart can pump less or more blood than usual. During rest or sleep heart rate decreases while during periods of physical exercise heart rate increase.

Heart is made up of four chambers which are entitled according to their locational properties like left/right and upper/lower. The two upper chambers are called as atria and the two lower chambers are called as ventricles. A muscular wall (septum) divides the heart into left and right parts. In a healthy heart, blood of left side cannot be mixed with the blood of right side. The left atrium is separated from left ventricle by mitral valve and the right atrium is separated from the right ventricle by tricuspid valve. Those two valves are called as atrioventricular valves. Similarly, two valves separate ventricles from blood vessels that carry the leaving blood. Pulmonic valve is placed between right ventricle and pulmonary artery while aortic valve is located between left ventricle and aorta.

Having a specific type of muscle kind (cardiac muscle) heart is under involuntary control. Functioning of the heart is rhythmic and regular but non-stationary. The period from one heart beat to another heart beat is called as cardiac cycle. Heart beats are controlled by a system of electrical signals generated in the heart. Sinoatrial node (SA, a small tissue in the wall of right atrium) is called as the pacemaker of the heart. The electrical signals generated by SA sets the rate of heartbeats and triggers the heart to contract in rhythm. Contraction of the heart starts from atria and then ventricles follow them.

One complete cardiac cycle is divided into two phases: pumping phase (namely systole) and filling phase (namely diastole). In the systole phase, the ventricles contract in order to pump the blood through the vessels to the body. Backward blood flow into atria is prevented by atrioventricular valves. Those valves close creating the first heart sound (S1). When the contraction of the ventricles end, this time backward blood flow into ventricles is stopped by the pulmonic and aortic valves. Those valves close immediately creating the second heart sound (S2). Then the ventricles relax and they are refilled with blood coming from the atria. This phase is called as diastole during which the heart gets ready for the following heartbeat. Generally, one cardiac cycle lasts 0.8 seconds at a normal heart rate but some factors like gender and age can change cardiac cycle period smoothly [3].

Each heartbeat consists of characteristic electrical and mechanical events. Those events occur in accordance with each other during the cardiac cycle. This relation between those mechanical and electrical events is defined as the dynamics of the heartbeat. Basic concepts of this topic have been defined by Wiggers (1923), Lewis (1925) and their colleagues [4]. Changes at aortic pressure, heart chambers' volume, arterial flow and heart sounds are some of the recurrent attributes of cardiac cycle. Carefully observing those events help clinicians to understand and diagnose various CVDs.

Healthcare professionals use many modalities for recognizing CVDs such as Electrocardiography (ECG or EKG), Echocardiography (echo), Phonocardiography (PCG), Magnetic Resonance Imaging (MRI), Ballistocardiography (BCG), Impedance Cardiography (ICG) and etc. Those medical monitoring tools works with different kind of signals such as electrical, acoustic, seismic, optical and radio-frequency [5]. Hospitals, clinics, treatment centers and home-care units have many medical devices and equipment. Thanks to advances in the technology of communication and information, today patients can access many forms of body sensors as well. With increasing access to monitoring devices, healthcare information about blood pressure, heart rhythm, heart rate variability, respiration rate and many other features can be followed regularly.

Although there are plenty of alternatives, many clinicians' first choice for examining circulatory and respiratory systems would be auscultation (i.e listening to the body sounds by using a stethoscope). Stethoscope was invented by Rene Laennec in 1816 but it is still a valid

method for diagnosis. Auscultation is cost-effective, simple, noninvasive, practical and fast therefore stethoscopes have been the very symbol of medical profession for two centuries. Being consisted of a diaphragm, two earpieces and rubber tubing, stethoscopes basically convert vibration signal into acoustic signal.

Internal body sounds can be heard from lungs, abdomen, heart and major blood vessels. For cardiac examination, there are four main regions to put the diaphragm of the stethoscope from which the valves can be best heard. Those regions are aortic region (centered at the second right intercostal space), pulmonic region (in the second intercostal space along the left sternal border), tricuspid region (between the 3rd, 4th, 5th, and 6th intercostal spaces at the left sternal border) and mitral region (near the apex of the heard between the 5th and 6th intercostal spaces in the mid-clavicular line) [6].

Auscultation evidences can be interpreted successfully as long as clinicians have good listening skills and experience. When abnormal sounds are heard, they are graded on a 6-point level scale (Levine scale) by physicians [7]. This process can be hard for intern doctors and inexperienced physicians. Moreover, same sounds can be subjectively categorized and graded by distinct listeners since human ear has physical limitations [8, 9]. Permanent records are not kept by traditional stethoscopes which makes consultation for same hearing impossible. Those drawbacks of conventional stethoscopes paved the way of modern phonocardiography. A phonocardiogram (PCG) is a high-quality graphical record of heart sounds which are captured and stored in electronic environment with the help of the machine called phonocardiograph [10]. Digitalization of heart sounds allowed the transmission of recordings to computers, automated analysis, long-term storage and graphic visualization [11].

Computer aided auscultation improves recognition of pathological signs and it is preferred for many reasons. On one hand, temporal rate of various sound components can be seen better in PCG records by time-scaling. Medical interns find it useful to study internal structure of the heart sound signals. On the other hand, it can be supported by concurrently collected ECG signals. ECG signals are especially useful when they are used for segmentation of cardiac cycles since there is observable correlation between ECG and PCG [12].

Computerized heart sound analysis has been focused widely and many studies have been conducted with different datasets and state-of-the-art methods. Those studies have been reviewed, compared and discussed in the literature by researchers. In this study, we present a novel review study in which major steps of heart sound analysis is discussed comprehensively. Additionally, feature engineering and important heart sound features are looked through systematically. As a contribution to the literature, recent datasets are presented in detail. Finally, heart sound classification methods are evaluated in a consistent manner.

Acoustic properties of the heart

In essence, computer aided auscultation is similar to acoustic signal processing since the sounds heard from cardiovascular system have the same spectrum as that of audio signals [12]. PCG records are examined for abnormality detection, heart sound localization and classification. The most common challenges of automated heart sound analysis are poor recording quality and environmental noises such as breathing of patients and rustlings of the microphone [5]. Additionally, like other biologic signals, heart sounds are non-stationery and they show sudden frequency changes. Moreover, frequency bands of their internal components are very close.

Under normal conditions of cardiac cycle, heart generates a dominant pair of sounds namely S1 and S2. They are also known as fundamental heart sounds (FHS) and basically described as lub-dub sounds of the heart [13]. Both S1 and S2 consist of two components. Closure of mitral and tricuspid valves produces the M1 and T1 sounds which form the S1 together. Similarly, closure of aortic valve produces A2 sound and pulmonic valve produces P2 sound. A2 and P2 are the components of S2. In normal cases, the time interval between M1-T1 and A2-P2 is not to exceed 30 ms [14]. Having a larger time interval between FHS components is an anomaly and, in such cases, those split sounds can be heard separately. Frequency ranges of S1 and S2 are 10-140 Hz and 10-200 respectively [15].

In addition to FHS, gallop rhythms (namely S3 and S4) and heart murmurs can be detected in PCG records as well [16]. S3 and S4 are known as extra heart sounds and they do not exist in most adults. Even when they are occasionally found, they are difficult to be distinguished by auscultation [14]. S3 occurs at early diastole while S4 takes place at late diastole just before the onset of S1. Frequency ranges of both S3 and S4 are 20-70 Hz [15]. On the other hand, heart murmurs have a larger frequency range of 200-600 Hz [16]. In medicine literature, a heart murmur is a rustling sound made by abnormal turbulent blood. Common causes of heart murmurs can be grouped as septal defects, valve abnormalities and heart muscle disorders (cardiomyopathy). Regurgitation through valves, stenosis of valves, septal/valvular defects and perforations can make a whistling sound indicating abnormality in blood circulation. Intensity, frequency, location and duration of those sounds are carefully examined in PCG records for cardiovascular diagnosis.

Heart murmurs are categorized according to their occurrence during cardiac cycle. Firstly, systolic murmurs are seen at the beginning of S1 and ends before S2. Secondly, diastolic murmurs take place during diastole such as aortic and pulmonary valve regurgitations. Finally, continuous murmurs are seen through all cardiac cycle [14]. In Figure 1 various PCG records are shown.

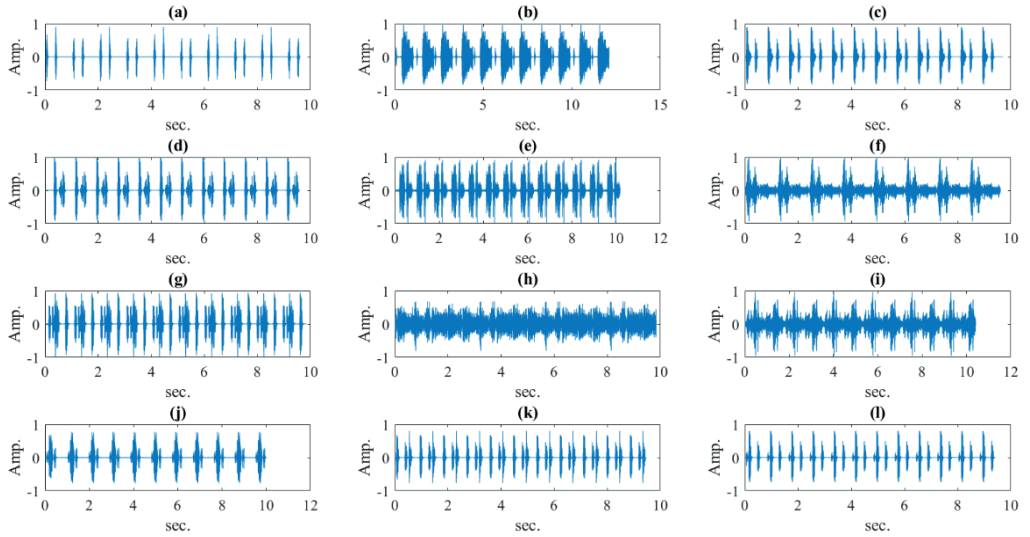

Figure 1. Pathologic and normal PCG records are exhibited. a: normal, b: aortic-insufficiency, c: aortic-stenosis-early, d: aortic-stenosis-late, e: atrial-septal-defect, f: mitral-regurgitation, g: mitral-stenosis, h: patent-ductus-arteriosus, i: pericardial rub, j: pulmonic-stenosis, k: s3 exists, l: s4 exists. Data obtained from University of Washington Heart Sound database in wav format and plotted as figure in MATLAB environment by us.

Heart sound localization

During auscultation, a clinician focuses on hearing the lubs and dubs of the heart to follow cardiac cycles. Determination of cardiac cycles and then localization of temporal positions of S1 and S2 is also known as segmentation. PCG records are divided into four segments: S1, systole, S2 and diastole [5, 12, 15, 17]. The time interval between ending of S1 to beginning of S2 is defined as systole and the time interval between the ending of S2 to onset of S1 is named as diastole. Temporal ratios of segments to each other and overall cardiac cycle is considered as an important feature in machine learning based studies. Moreover, envelopes, morphological and statistical properties of each segment can be examined separately. Expected temporal durations of those segments according to Schmidt et al. [18] is given in Table 1.

Table 1. Cardiac cycle segments and their durations

Segment	Mean (ms)	± Difference in 95% confidence interval (ms)
S1	122	32
Systole	208	80
S2	92	28
Diastole	523	311

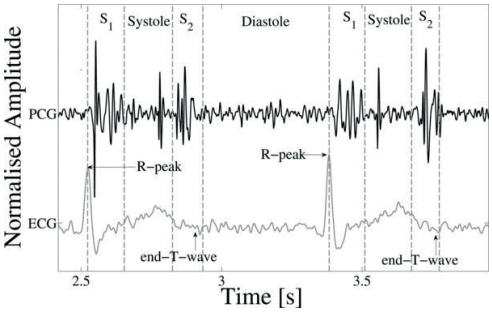

Figure 2. Synchronously collected ECG and PCG records example. Four states of PCG (S1, systole, S2 and diastole) are shown. Adapted from [17].

Heart sound datasets

In this section, we presented the datasets used in several studies. Those datasets are grouped into two categories: individual and public. Individual datasets belong to researchers, hospitals or small communities. They have been used in limited number of studies. On the other hand, public datasets are generally shared with larger communities and they have been used in researches world-wide. Some of them were used in contests and participants

developed state-of-art methods and tested them by using those PCG records.

PCG based computer aided diagnosis has been studied for decades. Early studies based on individual datasets and involved relatively small number of samples by which CNN based methods can be poorly trained. Another problem of working with PCG datasets is dissonance among the age range of the subjects. Datasets composed of children and teenagers is expected to show high heart rhythm variability [19]. Also, structural similarities between normal and abnormal PCG records make the classification task more challenging and hand-crafted studies based on selected features from small datasets are prone to overtraining [20]. Many studies present high accuracies but a direct comparison between those studies is not possible due to use of uncommon datasets. To overcome those constraints, larger public PCG datasets were prepared and made available for computation challenges like PhysioNet CinC/2016.

Researchers used different sampling frequencies during data acquisition for different datasets. 2 kHz can be accepted as the minimum limit of sampling frequency since there is not heart sound components with frequency higher than 1000 Hz [8, 21, 22].

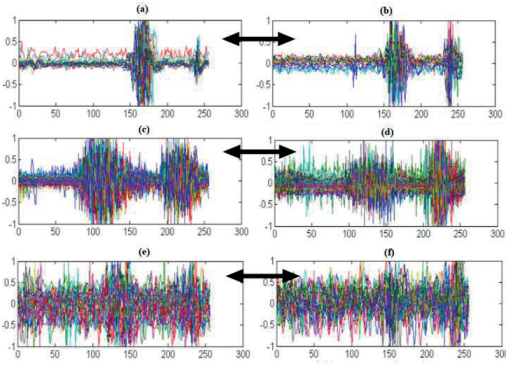

Figure 3. Structural similarities between normal and abnormal PCG records make heart sound classification more challenging. Normal PCG records: a, c and e. Abnormal PCG records: b, d and f. Similar records are shown by arrows. Adapted from [20]. Data from PhysioNet/CinC dataset.

Individual datasets

Bhatikar et al. [23] collected a total of 241 PCG records with murmurs (pathologic or innocent) from pediatric patients of cardiology clinic of The Children's Hospital, Denver, Colorado. Durations of the data were in the range of 10-15 seconds and sampling frequency was 44.1 kHz. Cardiac cycles were segmented manually by clinicians. Ahlstrom et al. [24] used a database recorded at the Department of Internal Medicine at Ryhov County Hospital, Jönköping, Sweden and at the Department of Clinical Physiology, University Hospital, Örebro, Sweden. Tang et al. [22] collected concurrent ECG and PCG data in their laboratory and used sampling frequency as 2 kHz accepting dominant frequency of heart sounds do not exceed 600 Hz. Zhang et al. [25] studied heart sounds after medical heart valve replacement operation. They used a database of 150 heart sounds which are generated by artificial (mechanical prosthetic) heart valves. Dataset was separated into five classes and sampling frequency was used as 8000 Hz. Moukadem et al. [26] used a database of 80 records and half of them were pathological. Gharehbaghi et al. [27] tried to separate innocent murmurs from pathological ones by using four data sets consisting of 10 second duration concurrent ECG and PCG. Those data sets were collected from volunteered patients at Linköping University hospital and Tehran Children Medical Center. Springer et al. [17] used 405 synchronous PCG and ECG records of 30-40 seconds which are collected from 123 adult patients at Massachusetts General Hospital. Karar et al. [28] used a database provided by CliniSurf, Faculty of Medicine, University of Bern, Switzerland. The database consists of 19 abnormal and 3 normal heart sounds. Each record has got about 15 cycles and sampling frequency is 44.1 kHz. Othman and Khaleel [29] used a database of 9 PCG records consisting of 3 normal, 3 abnormal with mitral regurgitation and 3 abnormal with mitral stenosis. Bozkurt et al. [30] used UoC-Murmur database (466 records) belonging to University of Crete, Greece. The concurrent PCG and ECG records were collected from children of 8-years-old age and classified by pediatric cardiology experts either as normal or as pathologic. Durations of them change in the range of 4-10 seconds and sampling frequency is 44.1 kHz. Aziz et al. [31] built their own PCG dataset by using their own data acquisition system at Rawalpindi Institute of Cardiology, Rawalpindi, Pakistan. From 56 subjects, they collected 140 normal and 140 pathologic (85 arterial septal defect, 55 ventricular septal defect) PCG records which were labeled by expert cardiologists into three classes (Normal, ASD, VSD). Sampling frequency was 44100 Hz and the duration of each record is 5 seconds. Yaseen et al. [32] build a database from public premade sources. They randomly selected 1000 PCG records from five different classes (200 records per each class). Those records were labeled by experts as normal, aortic stenosis, mitral regurgitation, mitral valve prolapse and mitral stenosis heart sounds. PCG records were resampled to 8000 Hz and stored in wav format. Safara et al. [33] used a dataset consists of 59 heart sounds (16 normal, 43 pathological).

Publicly available datasets

Texas Heart Institute Heart Sound Series

The database produced by the Robert J. Hall Heart Sounds Laboratory of Texas Heart Institute at St. Luke's Episcopal Hospital in 2009. The database includes 44 types of heart sounds [34] but it is not available now.

E-General Medical

A cardiac auscultation database of size 64 records was provided by eGeneral Medical Inc. It used to require payment for all database [15] but a part of it was free [12]. Database consisted of normal sounds, S3, S4 and different

pathologic cases. It has been used in several studies [16] and today it is not available online.

Frontiers in Bioscience

A total of 25 records available from int-prop.lf2.cuni.cz/heart_sounds/h14/sound.htm. Records are in wav format and their durations change in the range of 1.50 to 4.26 seconds.

Thinklabs Heart Sound Library

105 heart sounds in the range of 10-50 seconds. Available from Thinklabs' youtube channel https://www.youtube.com/c/Thinklabs1.

University of Washington

Different kinds of murmurs, split heart sounds and normal records present in wav format. Available from https://depts.washington.edu/physdx/heart/tech1.html.

3M Littmann Heart and Lung Sounds Database

PCG files are divided according to auscultation area. Records can be listened from browser. A total of 51 records available from http://www.3m.com/healthcare/littmann/mmm-library.html.

Heart Sound & Murmur Library Open Michigan

University of Michigan Health Systems provided an online database to educate undergraduate medical students and teach them clinical auscultation skills. The Michigan heart sound and murmur database (MHSDB) includes 23 records in mp3 format. It is available from www.med.umich.edu/lrc/psb_open/html/repo/primer_heartsound/primer_heartsound.html.

Cardiac Auscultatory Recording Database (CARD)

John Hopkins Outpatient Center Pediatric Cardiology Clinic digitized their clinic examinations in 1997. CARD is built from simultaneous PCG and ECG records of volunteered patients. CARD also includes clinical data for each case described by responsible cardiologist. 20 second duration recordings obtained from more than 1200 patient forms the CARD [35]. Audio files are stored in wav format and sampled at 4 KHz. It is available from http://murmurlab.com/card6/ after registration.

PASCAL (Pattern Analysis, Statistical modelling and ComputAtional Learning)

PASCAL [36] heart sound database was shared publicly in 2011 for two challenges. First competition was for heart sound segmentation and the second competition was for heart sound classification. The data were gathered from two sources and grouped into two (A and B). Group A was gathered by patients using iStethoscope Pro iPhone app. On the other hand, samples of group B were collected in hospital environment by physicians [37, 38]. The first group consists of four categories while the second group has only three categories of heart sounds. Sampling rate is 44100 Hz and the samples are stored in wav format. Durations of the samples vary in the range of 1 to 30 seconds [15] [36]. It has been used in several researches and it is still available online [39]. Details of it is given in the Table 2.

Table 2. PASCAL heart sound dataset details. Categories and number of samples in them are shown. Resource [36]

	Train				Test
Group	**Normal**	**Murmur**	**Extra heart sound**	**Artifact**	**Unlabeled**
A	31	34	19	40	52
B	319	93	46	-	195

Physionet CinC/2016

Similar to PASCAL dataset, PhysioNet heart sound dataset was built for a challenge in 2016. It was assembled from nine independent heart sound databases which were collected by seven distinct groups. Therefore, data acquisition hardware, data quality, labeling details and sampling frequencies were slightly different. In order to overcome this heterogeneity problem, all samples were resampled to 2 KHz and provided as wav files after anti-aliasing filtering [15]. Avoiding from specific diagnosis details, only three labels were assigned to PCG records: normal, abnormal and unsure.

The dataset is divided into train and test sets. While there is open access to train set, the test set is kept private. Submissions of the 2016/CinC challenge participants were ranked according to their performance on the test set. Moreover, poor signal quality and good signal quality were weighted differently during competition. Signal quality information is not available for train set. Details of the samples are given in the Table 3.

Table 3. PhysioNet database details. Resource [15]

Set	**Patients**	**PCGs**	**Abnormal**	**Normal**	**Unsure**
Train-A	121	409	276	116	17
Train-B	106	490	73	295	122
Train-C	31	31	20	7	4
Train-D	38	55	26	26	3
Train-E	356	2054	146	1781	127
Train-F	112	114	31	78	5
Total	764	3240	572	2303	365
Test-B	45	205	32	100	73
Test-C	14	14	9	4	1
Test-D	17	24	11	11	2
Test-E	153	883	59	763	61
Test-G	44	116	21	95	0
Test-I	35	35	21	12	2
Total	308	1277	153	985	139

Currently, the PCG dataset of PhysioNet is the largest one among other publicly available databases [40]. PCG records have different durations in the range of 5-120 seconds. Records are mono channel with 16-bit resolution in little-endian format. In addition to PCG, there is also simultaneously recorded ECG data for each sample of Train-A set. Train part of the database is available from https://physionet.org/content/challenge-2016/1.0.0/.

Materials and Methods

Research area of heart sound examination is an intersection of pattern recognition and signal analysis. Computer aided auscultation studies generally focus on either heart sound segmentation or on classification of heart sounds for detecting abnormalities [41]. But many classification studies [42] involve segmentation as a sub-step even though their main target is classification of heart sounds. Segmentation is not a fundamental prerequisite and, in the literature, there are classification studies which do not involve segmentation at all such as [43, 44]. On the other hand, localization of heart sounds and segmentation is the main purpose of many studies [17]. The general structure of the proposed methods in the literature is as follows: preprocessing, segmentation, feature extraction and classification.

Preprocessing

The main processes applied on the signal during preprocessing phase include filtering, resampling, baseline removal, denoising, decimation and normalization. Filtering is generally the first step of preprocessing. It can be done by amplitude or frequency-based filters. Potes et al. [45] used a band-pass filter of 25-400 Hz. In several studies [46, 47] Butterworth bandpass filter of 25-400 Hz was applied to remove high frequency noise and low frequency artifacts. In [48] fourth order Butterworth high-pass and low-pass filters were used with 600 and 25 Hz cutoffs while in [49] 57th order Butterworth lowpass filter with 900 Hz was used. In [50] 3th order median filter and 10th order Butterworth low-pass filter with 150 Hz cutoff were applied. In another study [51] a model was proposed for separating breathing sounds from heart sounds based on Kalman filter. They used synthetic data which were prepared by adding Gaussian noise (representing the respiratory sounds) to real acquired heart sound signal. In order to obtain segments from pathological records, Atbi and Debbal [52] used a low-pass FIR filter with 100 Hz cutoff frequency eliminating all high frequency components.

Many researchers normalized the signal into the [-1, 1] range by dividing with absolute maximum of the signal [16, 34, 53]. Some choose to normalize according to Equation 1 to obtain more visible peaks while weaking the noise [52, 54]. Another widely used normalization method [17, 39, 48] is z-score normalization. Safara et al. [33] normalized the signal according to Equation 2. Several studies [55, 56] involve baseline removal in this phase.

$$x_{norm} = \left(\frac{x}{\max(|x|)}\right)^2 \quad (1)$$

$$x_{norm} = \frac{x}{\sqrt{\sum x^2}} \quad (2)$$

In many studies, the classifier models need a uniform input shape. For example, inputs of CNN models and spectrograms of signals should have same dimensions to be processed. However, the datasets contain PCG records with variable length and/or different sampling frequencies. Features such as MFCC extracted from those signals differ in dimensions. Resampling and decimation are used to build a uniform input shape. Those steps beyond providing a uniform input shape also reduces computational cost. In studies [20, 45, 46, 48] PCG signals were down sampled from 2 kHz to 1000 Hz. Rubin et al. [57] decimated the PCG records to 3 second duration instances. In another study [58], first 5 seconds of the records were clipped and the remaining parts were discarded. Potes et al. [45] used 2.5 seconds decimation while shorter records were zero padded.

Another preprocessing step is denoising. Aziz et al. [31] used EMD for denoising. Gradolewski et al. [59] used Wavelet Packet Decomposition (WPD) to denoise PCG signals contaminated by white and pink noise. Similarly, Discrete Wavelet Transform was used in several studies [8] for denoising purposes by using thresholds. Threshold based denoising is either done by removing all samples below the threshold (hard thresholding) or it is done by producing a smoother transition over deleted values by subtracting threshold value from samples (soft thresholding) which are given in Equations (3-4).

$$y_{thr(T)}(t) = \begin{cases} y(t), & |y(t)| \geq T \\ 0, & otherwise \end{cases} \quad (3)$$

$$y_{thr(T)}(t) = \begin{cases} y(t) + T, & y(t) < -T \\ 0, & -T \leq y(t) \leq T \\ y(t) - T, & T < y(t) \end{cases} \quad (4)$$

Segmentation

Segmentation can be done by using simultaneously recorded ECG data [8, 24, 27]. Temporal segmentation of cardiac cycle is relatively easy by using ECG because its (PQRST) structure can clearly show the beginning and ending points of cycles. Moreover, it is more noise-free unlike PCG records. While interpreting ECG records, S1 is expected shortly after R peaks and S2 occurs at the end of the T wave [17]. However, the main disadvantages of this approach are needing auxiliary data and synchronizing it exactly with the PCG signal's timing.

Another segmentation approach depends solely on PCG. ECG-independent approach consists of various methods. Generally, envelope of the signal is extracted and used in this process. Envelope extracting approaches can be conducted with different mathematical properties of signals such as Shannon energy [37, 60, 61], Shannon entropy [62], variance fractal dimension [63], Hilbert-Huang transform [64] and autocorrelation [34]. Equations (5-9) can be used to map the original signal to non-negative domain for envelope extraction [29, 53].

Absolute value: $E = |S_i|$ (5)

Energy: $E = S_i^2$ (6)

Shannon entropy: $E = -|S_i| \log |S_i|$ (7)

Shannon energy: $E = -S_i^2 \log (S_i^2)$ (8)

Average Shannon energy: $E_{avg} = -\frac{1}{N}\sum_i^N S_i^2 \log (S_i^2)$ (9)

In the literature, there are methods based on amplitude thresholding such as [53, 65] and [66]. Similarly, peaks within predefined intervals are considered as S1 and S2 [9]. In an alternative approach, Ghosh et al. [61] segmented PCG records based on systolic and diastolic time intervals.

In several studies Hidden Markov Models (HMMs) are used for segmentation such as [67-69]. Schmidt et al. [18] proposed a hidden semi–Markov Model (HSMM) modeling the expected durations of the segments. Springer et al. [17] improved HSMM based approach of [18] by novel contributions such as adding probability of staying in a state for a defined duration, modifying Viterbi algorithm, applying Logistic Regression and then used a combination of homomorphic, Hilbert, PSD and Wavelet envelopes. In their study, the gold standard of the FHS positions in the PCG is derived from synchronously gathered ECG records. Abdollahpur et al. [48] proposed a novel method for cycle quality assessment build upon the work of [17]. After segmentation, original PCG signal is split into four distinct signals whose features are extracted and interpreted individually.

Feature engineering

PCG signals contain large number of samples and success of signal processing methods depend on extracting meaningful features from the signal. Features can be extracted from signals in time, frequency and time-frequency domains. In many studies, different features collected from distinct features are combined. However, when the number of extracted features increase dramatically, computational cost increases as well. In such cases dimension reduction methods Principal Component Analysis (PCA), Singular Value Decomposition (SVD) or Independent Component Analysis (ICA) are applied in several studies [70-72]. Another solution of this problem is to choose the most meaningful features and not to increase feature set sizes by using rest of the features.

Alternatively, deep learning-based classification methods do not need manual feature engineering. Those models can be applied directly on either PCG signals or on envelograms obtained from those signals without feature extraction [73]. The researchers can use those models as is or they can just take the generated features by those models to use with conventional classifiers [74].

Time Domain Features

Time domain features of a signal present the statistical attributes changing over time. PCG signals have various morphological characteristics which can be observed from time domain perspective. Most commonly used time domain features such as mean, standard deviation, median, signal energy, maxima, minima and zero-crossing rate are calculated directly from the signal itself. Moreover, signals can be converted into probability density functions and their entropies can be calculated by different methods. Some of the time domain features widely used in the researches are given in the Table 4.

Table 4. Time domain features

Feature	Formula	Ref
mean	$\bar{x} = \frac{1}{N}\sum_i^N x(i)$	[75-77]
Standard deviation	$s = \sqrt{\frac{1}{N}\sum_i^N (x(i) - \bar{x})^2}$	[75-77]
Skewness	$\left[\frac{1}{N}\sum_i^N (x(i) - \bar{x})^3\right] / s^3$	[75-77]
Kurtosis	$\left[\frac{1}{N}\sum_i^N (x(i) - \bar{x})^4\right] / s^4$	[75-77]
Median	Middle value or average of two middle values for arrays with even number of samples	[77]
Maxima, minima		[76]
Zero crossing rate	$\frac{1}{2N}\sum_{i=1}^{N} \lvert sgn[x(i)] - sgn[x(i-1)]\rvert$	[13, 76]
Shannon entropy	$-\sum_i^n \lvert p_i * \log (p_i)\rvert$	[76]
Avg Shannon enrgy	$-\frac{1}{N_{seg}}\sum_i^{N_{seg}} S_i^2 \log (S_i^2)$	[24]
Karcı entropy	$\sum_i^n \lvert(-p_i)^{\alpha} * \ln(p_i)\rvert$	[13]

Percentile rate	$\frac{25^{th} percentile}{75^{th}\ percentile}$	[77]
RMS	$\sqrt{\frac{1}{N}\sum_i^N x(i)^2}$	[76]
Zero cross rate	Rate of sign changes at signal	[76]
Total original signal power	$\sum_i^N \frac{x(i)^2}{N}$	[38]
Avg widths of segments		[71]
Ratios of segments' avg widths to each other		[78]

Frequency Domain Features

When a sinusoid signal is added with another one the result is another sinusoid signal but may be shifted in amplitude, phase and frequency. Assuming the general formula given in Equation 10 is valid for the signals of interest, amplitudes of different frequency components provide meaningful features. In the Equation 10, A is amplitude, f is frequency, t is time and θ is phase offset in radians.

$$y(t) = \sum_i A_i \sin(2\pi f_i t + \theta_i) \quad (10)$$

But frequency domain features lack the ability of indicating temporal positions of abnormalities. Moreover, suitability of them is often criticized due to the non-stationary structure of PCG signals. They are generally considered insufficient alone however they are used together or within other methods. Those features are used for band-pass filter banks and zero crossing analysis [40].

Fourier Transform (FT)

FT provides frequencies and their magnitudes of a signal and it is very useful for stationary signals. FT technique is used to examine harmonic components of a signal by transforming the signal from time domain to frequency domain. FT provides valuable information about frequency bands but it lacks the capacity of locating the frequency regions in time. Another major disadvantage of FT is that it cannot be applied on multi-channel signals. Debbal et al. [79] applied Fast Fourier Transform (FFT) on PCG data and detected FHS frequencies in the spectrum (gathered around 40-200 Hz) however they concluded that duration and transient variations cannot be detected by FFT. Bhatikar et al. [23] used 0-300 Hz energy spectrum obtained by Fast Fourier Transform.

$$e^{ix} = \cos(x) + isin(x) \quad (11)$$

$$F(f) = \int_{t=-\infty}^{\infty} f(t)\, e^{-i2\pi f t} dt \quad (12)$$

Direct Cosine Transform (DCT)

DCT can be used for audio and image signal compression. Discrete time domain signal is firstly converted into a sum of cosine functions with different frequencies. Their amplitudes are interpreted as the features. In essence, DCT is similar to FT but DCT uses only cosine functions for transformation and output values are real numbers.

Time-Frequency Domain Features

Short Term Fourier Transform (STFT)

STFT was developed to overcome time resolution problem of FT. STFT assumes that some portion of an input signal is stationary. Each stationary-accepted sub-region is applied FT and then all parts are added up. STFT formula is given in Equation 13 where w(t) is window function.

$$\mathrm{H(t,f)} = \int h(t) w(t-\tau) e^{-2\pi f \tau} d\tau \quad (13)$$

In STFT, there is a trade of between time and frequency resolution. Fixed size windows of STFT affect the time-frequency representation. As window size increases, frequency capturing performance rises but time resolution decreases. Conversely, if window size is kept small then time domain gets more accurate while losing frequency information.

Wavelet Transform (WT)

Being a better alternative than FT and STFT, WT was originally developed to optimize frequency dependent temporary resolutions [80]. WT can be defined as a short wave that has an average value approaching zero. The energy carried by the wave is condensed in time. WT provides good resolution both in time and frequency domains.

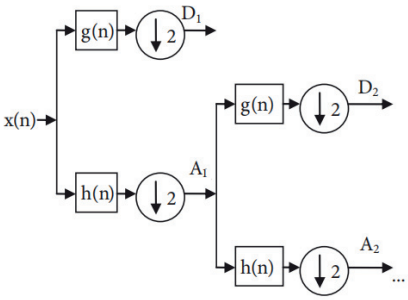

Figure 4. DWT decomposition scheme. D: detail, A: approximation, h: low-pass filter, g: high-pass filter

When an orthogonal basis function is used as wavelet, the WT is named as discrete wavelet transform (DWT). If a non-orthogonal basis function is used as wavelet, then the transform is called as continuous wavelet transform (CWT). Another effective wavelet-based feature extraction method is wavelet packet decomposition (WPT). Similar to STFT, its top level is good in time resolution and as level decomposition goes on temporal resolution decreases in the favor of frequency resolution. WT is used for denoising, data compression, obtaining sub-band features and visualizing spectral components as well. The general formula is given in Equation 14 where ψ(t) is wavelet function, ϕ(t) is scaling function, h(t) is low-pass impulse response and g(t) is high-pass impulse response, c_i and $d_{i,j}$ are coefficients [12].

$$y(t) = \sum_{j=-\infty}^{\infty} c_i \varphi_j(t) + \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d_{i,j} \psi_{i,j}(t) \quad (14)$$

Empirical Mode Decomposition (EMD)

EMD is proposed as a part of Hilbert-Huang Transform [72] and it has been used for many purposes such as denoising [31], detection of heart sounds [81] and segmentation [50].

EMD iteratively reduces the input signal into intrinsic mode functions (IMFs) and a residual. The process of extracting IMFs from the raw signal is known as sifting. The original signal can be expressed in terms of IMFs and residual signal as given Equation 15. An IMF must satisfy two requirements. First, in the whole dataset, the number of extrema and the number of zero-crossings must either be equal or differ at most by one. Second, at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

$$y(t) = \sum_{k=1}^{N} h_k(t) + r(t) \quad (15)$$

After IMFs are obtained, they can be examined individually. Generally, the first IMF is discarded since it contains high frequency components. Time and frequency domain features can be extracted from IMFs.

The main steps of sifting are:

1. Calculating all of the local minima and maxima from the signal y(t)
2. Cubic spline interpolation is applied on local minima and maxima in order to obtain envelopes $e_{min}(t)$ and $e_{max}(t)$
3. Mean of upper and lower envelopes are calculated according to Equation 16.

$$a(t) = (e_{max}(t) + e_{min}(t)) / 2 \quad (16)$$

4. Mean envelope is subtracted from the original signal y(t) to obtain ith IMF $h_i(t)$ according to Equations 17, 18.

$$h_i(t) = y(t) - a(t) \quad (17)$$

$$r_i(t) = y(t) - h_i(t) \quad (18)$$

5. Treat $r_i(t)$ as the new signal and repeat steps 1-4 until residual signal contains no more IMF

Mel-Frequency Cepstral Coefficients (MFCC)

MFCC is a well-known technique used in speech recognition and speaker identification. It has also found usage in PCG analysis as well [45]. Human ears do not perceive pitch linearly. Mel scaling aims to mimic human auditory systems by mapping the frequencies below 1000 Hz linearly and by mapping the frequencies above 1000 Hz logarithmically [12].

To obtain MFCC features, pre-emphasizing is the first step in which high frequencies are amplified. Then the quasi-stationary signal is divided into short frames across which the signal is assumed to be stationary. Generally consecutive frames overlap a pre-defined amount of time. Then a window (such as Hamming, Hanning or etc.) is applied on the frames to reduce edge effects and smooth the edges. Then Discrete Fourier Transform is applied on the windowed frames to compute the periodogram. Then the Fourier transformed signal is passed through Mel-filter bank (a set of bandpass filters). This phase results in non-linear frequency resolution. It is given in Equations (19-20) where f is physical frequency and f_{MEL} is its Mel-frequency representation.

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{\frac{-j2\pi nk}{N}};\ 0 \le k \le N-1 \quad (19)$$

$$f_{MEL} = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (20)$$

Now Mel spectrum is fit into log format in which most of the signal information is represented by the first few coefficients. M is total number of Mel weighting filters and $H_m(k)$ is the weight given to kth energy spectrum bin according to Equation 21.

$$P_{filt} = \sum_{k=0}^{N-1} [|X(k)|^2 H_m(k)];\ 0 \le m \le M-1 \quad (21)$$

Finally, MFCCs are obtained by taking a discrete cosine transform. This process converts the Mel spectrum to finite sequence of cosine functions oscillating at different frequencies. In Equation 22, *MFCC(t,k)* is kth cepstral feature of tth time frame and $P_{filt}(t,n)$ is filtered power at time frame t for nth filter bank. The number of MFCCs for each frame is C and zeroth coefficient can be excluded since it represents the average log energy of the input signal.

$$MFCC(t,k) = \sum_{n=0}^{N-1} \log\left(P_{filt}(t,n)\right) \cos\left(\frac{k\pi}{N}(n - 0.5)\right);\ k = 0,1,2 \ldots, C-1 \quad (22)$$

Classification

In this section we present a list of proposed solutions with year, dataset and performance results. It is not an exhaustive list and involves results from previous challenges.

Classification is done with many machine learning methods including support vector machine, k nearest neighbor, multilayer perceptron, decision trees, convolutional neural networks and their ensembles. Ensemble classifiers reduces the error rate significantly and usually their performances are better than their base classifiers [37]. Performance comparisons of PhysioNet challenge are done according to Equations (28-30) where w_{a1}, w_{a2}, w_{n1} *and* w_{n2} are weights of good signal quality of abnormal, poor signal quality of abnormal, good signal quality of normal and poor signal quality of normal records. Rules for determining the overall classification result of the challenge are given in Table 5. On the other hand, the rest of the comparisons are done according to well-known formulas given in the Equations 23-27.

Table 5. Rules of PhysioNet CinC/2016 scoring

Label	Quality	Weight	Predicted Abnnormal	Predicted Unsure	Predicted Normal
Abnormal	Clean	w_{a1}	A_{a1}	A_{q1}	A_{n1}
	Noisy	w_{a2}	A_{a2}	A_{q2}	A_{n2}
Normal	Clean	w_{n1}	N_{a1}	N_{q1}	N_{n1}
	Noisy	w_{n2}	N_{a2}	N_{q2}	N_{n2}

Classification performance is highly dependent on the preprocessing, segmentation and feature extraction phases. Several studies are shown in Table 6 with details about used dataset, deployed method, extracted features and performance results. Comparing them, it is seen that classification accuracies change in the range of 75 to 99 percent. Especially on private datasets, accuracies above 90 percent are reported generally. On the other hand, challenge datasets such as PhsyioNet evaluated performance results over hidden test sets avoiding overfitting. Diversity of those datasets prevent us from direct comparison between them but general trends and validation methods give insight about the success of the proposed methods and usefulness of the selected features.

In studies [45, 70, 90, 91], distinct frequency subbands of the signal were decomposed and used to extract features or envelopes. S1 and S2 intervals are searched for in low frequency subbands in [90]. In [45], more features were obtained from decomposed subbands. Koçyiğit [70] used discrete wavelet transform to get signal subband and then applied PCA and ICA dimension reduction methods on it to extract features. The extracted features were given to Naïve Bayes classifier and 99.8% accuracy was obtained.

The HSMM-based segmentation method proposed by Springer [17] was used by many Physionet 2016/CinC contestants in their studies [42, 45, 57, 74-76, 87]. Features were extracted from four segments of the detected cardiac cycle and variations in the durations of them were interpreted as the signs of anomalies. On the other hand, some studies [13, 43, 58, 70] processed the PCG signal as a whole without using any segmentation method on them. When PCG analysis approaches with segmentation-based and without segmentation methods are compared, it is seen that there is not much difference. For example, in PhysioNet 2016 CinC, Potes et al. [45] took the first place with a segmentation-based approach by obtaining a score of 86.02. However, Zabihi et al. [43] took the second place by obtaining 85.90 without applying segmentation. In [45] features were extracted from four states in time domain and frequency domain while in [43] desired features were extracted from whole signal alone. The small difference between those studies indicates that selection of features is more important for PCG classification than applying state-of-the-art segmentation methods.

In earlier studies, different types of Artificial Neural Networks (ANN) have been applied [43, 48, 87] to solve heart sound classification problem. Features from time, frequency and time-frequency domains formed feature vectors of sizes 40 to 675. ANN models are trained with those feature vectors after applying dimension reduction. Accuracies higher than 80% percent are obtained by former ANN based approaches. It is also used for segmentation of heart sounds. Ghaemmaghami et al. [93] extracted and used 6 mel-frequency filterbank features to categorize temporal frames of audio recordings as S1, systole, S2, diastole and noise by using time-delay neural networks (TDNN). TDNN is good for detecting local correlations between segments and it is able to capture long term temporal correlations between cycles frames. Temporal events in heart sounds can be detected by TDNN as well.

Currently researchers' interest focused more on different types of Deep Neural Networks (DNN). For instance, Recurrent Neural Networks (RNN) have been used in recognizing sequential data for decades and it is also applicable on heart sounds since they have strong temporal correlation. Vanishing gradient problem of RNN is handled by structures like Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Khan et al. [94] used LSTM with MFCC features of unsegmented PCG data obtaining an AUC score of 91.4% and accuracy above 80%. Among the DNN based methods, LSTM models have relatively high complexity but they are capable of modeling temporal structures and dependencies.

An alternative of using DNN is to generate 2D images or image-like inputs from time series data by applying time-frequency spectrograms, MFCC or heat maps and then training Convolutional Neural Networks (CNN) [57, 82, 92]. There are also solutions [20, 46, 95] based on 1D CNN models which only perform simpler one-dimensional convolutions (scalar addition and multiplication) on PCG signals.

Both DNN and CNN methods require large datasets and long training time for better classification accuracies. To train those models in reasonable time periods, researchers generally take the advantage of using modern GPU hardware.

$$Accuracy = \frac{TP+TN}{All} \tag{23}$$

$$Recall = Sensitivity = \frac{TP}{TP+FN} \tag{24}$$

$$Specificity = \frac{TN}{TN+FP} \tag{25}$$

$$Precision = Positive\ Predictive\ Value = \frac{TP}{TP+FP} \quad (26)$$

$$F1\ Score = \frac{2*Presicion*Recall}{Presicion+Recall} \quad (27)$$

$$Modified\ Sensitivity(Msn) = \frac{w_{a1}\ A_{a1}}{A_{a1}+A_{q1}+A_{n1}} + \frac{w_{a2}\ (A_{a2}+A_{q2})}{A_{a2}+A_{q2}+A_{n2}} \quad (28)$$

$$Modified\ Specificity(Msp) = \frac{w_{n1}\ N_{n1}}{N_{a1}+N_{q1}+N_{n1}} + \frac{w_{n2}\ (N_{a2}+N_{q2})}{N_{a2}+N_{q2}+N_{n2}} \quad (29)$$

$$Modified\ accuracy(Macc) = \frac{Sensitivity+Specificity}{2} \quad (30)$$

Table 6. Performance results of studies in the literature

Ref	Year	Data	Classes	Features	Methods	Results (%)	
[47]	2020	PhysioNet	2	MFCC based features	Convolutional recurrent neural network	Acc: 98.34 Sen: 98.66 Spe: 98.01	
[31]	2020	private	3	MFCC 1D Local ternary patterns	SVM	Acc: 95.24	
[82]	2020	Dataset of Yaseen et al. [32]	5	Bispectrum images (full-image and contour)	CNN	**Full-img** Acc:98.70 Sen:98.70 Spe:99.67	**Contour** Acc:97.10 Sen:97.10 Spe:99.28
[61]	2020	Dataset of Yaseen et al. [32]	5	Time-frequency domain energy and entropy features	Multi class composite classifier	Acc: 98.33	
[20]	2020	PhysioNet	2	-	1D CNN	Sen: 89.67 Spe: 86.89 Ppr: 69.70	
[83]	2020	PASCAL	4 & 5	Time domain MFCC	SVM kNN	Acc: 99.25 Acc: 98.50	
[13]	2020	PhysioNet	2	Time domain Time-frequency domain MFCC	Ensemble of kNN, SVM, neural networks	Acc: 90.93 Sen: 98.00 Spe: 64.00	
[92]	2019	PhysioNet	2	STFT spectrogram Mel Spectrogram MFCC	CNN (VGGNet) Majority voting ensemble	Acc: 86.04 Sen: 86.46 Spe: 85.63	
[46]	2019	PhysioNet	2	Time-frequency MFCC based feature maps	Ensemble of 1D CNN and 2D CNN	Acc: 89.22 Sen: 89.94 Spe: 86.35	
[84]	2019	187 PCG records from PhysioNet	2	Cochleagram	Multilayer perceptron	Acc: 93.70 Sen: 84.50 Spe: 95.20 F1: 83.50	
[85]	2018	private	2	Time and frequency domains	ANFIS HMM	Acc: 98.70	
[86]	2018	Open Michigan Library PASCAL PhysioNet	2	Time-frequency MFCC STFT	Recurrent neural network CNN Bidirectional long short-term memory	Sen: 96.00 Spe: 100.00 F1: 98	
[30]	2018	UoC proprietary	2	Sub-band envelopes of segmented PCG	CNN	Acc: 81.50 Sen: 84.50 Spe: 78.50	
[32]	2018	individual	5	MFCC DWT	kNN SVM DNN	**kNN** Acc: 97.40 Sen: 97.60 Spe: 98.80 F1: 99.20 **SVM** Acc: 97.90 Sen: 98.20 Spe: 99.40 F1: 99.70	**DNN** Acc: 92.10 Sen: 94.50 Spe: 98.20 F1: 98.30

[78]	2017	PhysioNet	2	Time, frequency, sparse coding	SVM	Macc: 80.70 Sen: 84.30 Spe: 77.20
[49]	2017	Private	2	Time-frequency	kNN	Acc: 93.20
[48]	2017	PhysioNet	2	Time, time-frequency	Neural networks	Macc: 82.63 Mse: 76.96 Msp: 88.31
[42]	2017	PhysioNet	2	Time, frequency, time-frequency	Ensemble classifiers	Macc: 80.10 Mse: 79.60 Msp: 80.60
[93]	2017	Self-collected 128 records of 20 seconds duration	5	MFCC	Time Delay Neural Network	Acc: 95.80 Sen: 83.20 Spe: 99.20
[45]	2016	PhysioNet	2	Time, frequency	Adaboost & CNN ensemble	Macc: 86.02 Mse: 94.24 Msp: 77.81
[43]	2016	PhysioNet	2	Time, frequency and time-frequency	Ensemble of neural networks	Macc: 85.90 Mse: 86.91 Msp: 84.90
[87]	2016	PhysioNet	2	Time, frequency domain features, CWT features, MFCC	Neural networks	Macc: 85.20 Mse: 87.43 Msp: 82.97
[88]	2016	PhysioNet	2	Time-frequency	LR SVM kNN	Macc: 84.54 Mse: 86.39 Msp: 82.69
[76]	2016	PhysioNet	2	Time, frequency and time-frequency domain features	Random Forest LogitBoost	Macc: 84.48 Mse: 88.48 Msp: 80.48
[57]	2016	PhysioNet	2	MFCC heat map images	CNN	Macc: 83.99 Mse: 72.78 Msp: 95.21
[89]	2016	private	2	MFCC based features	kNN GMM LR SVM DNN	Acc: 78.11 Acc: 86.98 Acc: 87.57 Acc: 90.53 Acc: 91.12
[75]	2016	PhysioNet	2	Statistical features of WT applied data	SVM	Acc: 74.60 Sen: 64.40 Spe: 84.90
[90]	2016	PhysioNet	2	Statistical features Frequency domain	Fuzzy logic (PROBAfind)	Acc: 95.00 Sen: 93.00 Spe: 97.00
[33]	2013	private	4	Time-frequency	SVM	Acc: 97.56
[21]	2009	Students' training CD	15	Time-frequency	Divergence analysis	Acc: 99.00
[91]	2009	Littman and Frontiers in Bioscience datasets	5	Time-frequency	SVM with several kernel functions (Best results obtained with Gaussian Radial Basis Function kernel)	Acc: 91.43 Sen: 87.50 Spe: 94.74

Conclusion and suggestions

In this study we reviewed PCG analysis methods and existing databases. Feature extraction techniques and methodological approaches are presented and compared. Heart sound analysis is an interesting topic and it is still challenging. The fundamental heart sounds corrupted with various pathological factors. Mitral stenosis, mitral regurgitation, aortic insufficiency, aortic regurgitation, valve disorders, septal defects and gallop rhythms cause heart murmurs that differ from each other with respect to frequency and location. Main tasks of heart sound analysis focus on detecting murmurs and segmentation. Besides, classifying distinct types of murmurs has been targeted by several studies. On the other hand, heart sound segmentations are done by detecting peak values, using

systolic/diastolic temporal assumptions, threshold-based methods and using external signals (commonly ECG). Environmental and digital noises are tried to be remove from heart sounds. For this purpose, signal denoising is applied as the first step of automated PCG analysis. It is followed by feature engineering in conventional approaches but CNN-based methods do not involve feature extraction. Those features are grouped in time, frequency and time-frequency domains. Wavelet based features are most preferred time-frequency domain features and as it is seen in Table 6, the most commonly used features belong to time-frequency domain. Similar to Wavelet transform, EMD is also good at representing the sound signal in time-frequency domain. After this step, PCG records are classified by several algorithms such as HMM, SVM, decision trees and neural networks.

Although the current studies make a good sum, there is still room for improvements. Firstly, there is need for a universally standardized and open access database. In the past, the lack of data was a problem for researchers. Researchers generally used private databases decades ago but today there are databases like PASCAL, CARD, PhysioNet and etc. Among those PCG databases, the largest one is PhysioNet with 665 abnormal and 2575 normal records having an imbalance ratio of 2575/665 = 3.87. Although it is the largest one, state-of-the-art CNN-based methods require larger databases with smaller imbalance ratio. Secondly, more information on data acquisition and auscultation locations should be given in those standardized databases. Heart sounds generally recorded from aortic area, pulmonic area, tricuspid area and mitral area. Depending on the position of the auscultation sensor, loudness of first and second heart sounds could be captured differently. Finally, data acquisition should be improved by avoiding from noise. Additionally, an extra classifier can be added to the systems in order to detect signal quality. Excluding records with significant environmental noise increases methods' accuracy and results in more meaningful conclusions about pathologies.

Performance of automated heart sound analysis is promising and it has many possible benefits. Development of computationally efficient methods paves the way of intelligent biomedical devices. Smart phones, wearable systems and other portable gadgets empower home health care systems. Moreover, those equipments could play crucial role in monitoring and diagnosis of CHDs in rural areas where it is hard to access expert clinicians' consultation.

References

1. Yang, Lucy, et al. "The role of epoxyeicosatrienoic acids in the cardiovascular system." British journal of clinical pharmacology 80.1 (2015): 28-44.
2. WHO. *Cardiovascular Diseases*. Accessed: January 2021. [Online]. Available: https://www.who.int/health-topics
3. Al Ahmad, Mahmoud, and Soha Ahmed. "Heart-rate and pressure-rate determination using piezoelectric sensor from the neck." 2017 4th IEEE international conference on engineering technologies and applied sciences (ICETAS). IEEE, 2017.
4. Coblentz, B., et al. "The relationship between electrical and mechanical events in the cardiac cycle of man. "British heart journal 11.1 (1949): 1.
5. Jain, Puneet Kumar, and Anil Kumar Tiwari. "Heart monitoring systems—A review." Computers in biology and medicine 54 (2014): 1-13.
6. "Auscultation Assistant". http://www.med.ucla.edu/wilkes/intro.html. Accessed 14 January 2021.
7. Silverman, Mark E., and Charles F. Wooley. "Samuel A. Levine and the history of grading systolic murmurs." The American journal of cardiology 102.8 (2008): 1107-1110.
8. Delgado-Trejos, Edilson, et al. "Digital auscultation analysis for heart murmur detection." Annals of biomedical engineering 37.2 (2009): 337-353.
9. El-Segaier, Milad, et al. "Computer-based detection and analysis of heart sound and murmur." Annals of Biomedical Engineering 33.7 (2005): 937-942.
10. Leng, S., San Tan, R., Chai, K. T. C., Wang, C., Ghista, D., & Zhong, L. (2015). The electronic stethoscope. Biomedical engineering online, 14(1), 1-37.
11. Montinari, Maria Rosa, and Sergio Minelli. "The first 200 years of cardiac auscultation and future perspectives." Journal of multidisciplinary healthcare vol. 12 183-189. 6 Mar. 2019, doi:10.2147/JMDH.S193904
12. Ismail, Shahid, Imran Siddiqi, and Usman Akram. "Localization and classification of heart beats in phonocardiography signals—A comprehensive review. "EURASIP Journal on Advances in Signal Processing 2018.1 (2018): 26.
13. Gündüz, Ali Fatih, and Ali KARCİ. "Heart Sound Classification for Murmur Abnormality Detection Using an Ensemble Approach Based on Traditional Classifiers and Feature Sets. "Bilgisayar Bilimleri 5.1: 1-13.
14. Felner JM. The First Heart Sound. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition. Boston: Butterworths; 1990. Chapter 22. Available from: https://www.ncbi.nlm.nih.gov/books/NBK333/
15. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(9).
16. Varghees, V. Nivitha, and K. I. Ramachandran. "A novel heart sound activity detection framework for automated heart sound analysis." Biomedical Signal Processing and Control 13 (2014): 174-188.
17. Springer DB, Tarassenko L, Clifford GD. Logistic regression-HSMM-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63(4):822-832.

18. Schmidt, Samuel E., et al. "Segmentation of heart sound recordings by a duration-dependent hidden Markov model." Physiological measurement 31.4 (2010): 513.
19. Oliveira, Jorge, et al. "Adaptive sojourn time HSMM for heart sound segmentation." IEEE journal of biomedical and health informatics 23.2 (2018): 642-649.
20. Kiranyaz, Serkan, et al. "Real-time phonocardiogram anomaly detection by adaptive 1D convolutional neural networks." Neurocomputing 411 (2020): 291-301.
21. Dokur, Zümray, and Tamer Olmez. "Feature determination for heart sounds based on divergence analysis." Digital Signal Processing 19.3 (2009): 521-531.
22. Tang, Hong, et al. "Segmentation of heart sounds based on dynamic clustering." Biomedical Signal Processing and Control 7.5 (2012): 509-516.
23. Bhatikar, Sanjay R., Curt DeGroff, and Roop L. Mahajan. "A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics." Artificial intelligence in medicine 33.3 (2005): 251-260.
24. Ahlstrom, Christer, et al. "Feature extraction for systolic heart murmur classification." Annals of biomedical engineering 34.11 (2006): 1666-1677.
25. Zhang, Di, et al. "Noninvasive detection of mechanical prosthetic heart valve disorder." Computers in biology and medicine 42.8 (2012): 785-792.
26. Moukadem, Ali, et al. "A robust heart sounds segmentation module based on S-transform." Biomedical Signal Processing and Control 8.3 (2013): 273-281.
27. Gharehbaghi, Arash, et al. "A novel method for discrimination between innocent and pathological heart murmurs. "Medical engineering & physics 37.7 (2015): 674-682.
28. Karar, Mohamed Esmail, Sahar H. El-Khafif, and Mohamed A. El-Brawany. "Automated diagnosis of heart sounds using rule-based classification tree." Journal of medical systems 41.4 (2017): 60.
29. Othman, Mazin Z., and Asmaa N. Khaleel. "Phonocardiogram signal analysis for murmur diagnosing using Shannon energy envelop and sequenced DWT decomposition." Journal of Engineering Science and Technology 12.9 (2017): 2393-2402.
30. Bozkurt, Baris, Ioannis Germanakis, and Yannis Stylianou. "A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection." Computers in biology and medicine 100 (2018): 132-143.
31. Aziz, Sumair, et al. "Phonocardiogram signal processing for automatic diagnosis of congenital heart disorders through fusion of temporal and cepstral features." Sensors 20.13 (2020): 3790.
32. Yaseen, G. Y. Son, and S. Kwon. "Classification of heart sound signal using multiple features." *Applied Sciences-Basel* vol. 8, no. 12, p. 2344, 2018.
33. Safara, Fatemeh, et al. "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification." Computers in biology and medicine 43.10 (2013): 1407-1414.
34. Kao, Wen-Chung, and Chih-Chao Wei. "Automatic phonocardiograph signal analysis for detecting heart valve disorders. "Expert Systems with Applications 38.6 (2011): 6458-6468.
35. Thompson, W. Reid, et al. "Artificial intelligence-assisted auscultation of heart murmurs: validation by virtual clinical trial." Pediatric cardiology 40.3 (2019): 623-629.
36. Bentley, P.J., Nordehn, G., Coimbra, M., Mannor, S.: The PASCAL classifying heart sounds challenge 2011 (CHSC2011) Results. www.peterjbentley.com/heartchallenge/
37. Balili, Christine C., Ma Caryssa C. Sobrepena, and Prospero C. Naval. "Classification of heart sounds using discrete and continuous wavelet transform and random forests." 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, 2015.
38. Chakir, Fatima, et al. "Phonocardiogram signals processing approach for PASCAL classifying heart sounds challenge." Signal, Image and Video Processing 12.6 (2018): 1149-1155.
39. Vrbancic, Grega, and Vili Podgorelec. "Automatic detection of heartbeats in heart sound signals using deep convolutional neural networks." Elektronika ir Elektrotechnika 25.3 (2019): 71-76.
40. Dwivedi, Amit Krishna, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. "Algorithms for automatic analysis and classification of heart sounds–a systematic review. "IEEE Access 7 (2018): 8316-8345.
41. Liu, Qingshu, Xiaomei Wu, and Xiaojing Ma. "An automatic segmentation method for heart sounds. "Biomedical engineering online 17.1 (2018): 1-22.
42. Homsi, Masun Nabhan, and Philip Warrick. "Ensemble methods with outliers for phonocardiogram classification. "Physiological measurement 38.8 (2017): 1631.
43. Zabihi, Morteza, et al. "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
44. Deng, Shi-Wen, and Ji-Qing Han. "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps." Future Generation Computer Systems 60 (2016): 13-21.
45. Potes, Cristhian, et al. "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds." 2016 computing in cardiology conference (CinC). IEEE, 2016.
46. Noman, Fuad, et al. "Short-segment heart sound classification using an ensemble of deep convolutional neural networks." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.
47. Deng, Muqing, et al. "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks." Neural Networks 130 (2020): 22-32.

48. Abdollahpur, Mostafa, et al. "Detection of pathological heart sounds." Physiological measurement 38.8 (2017): 1616.
49. Ghaffari, Milad, et al. "Phonocardiography signal processing for automatic diagnosis of ventricular septal defect in newborns and children." 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2017.
50. Papadaniil, Chrysa D., and Leontios J. Hadjileontiadis. "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features." IEEE journal of biomedical and health informatics 18.4 (2013): 1138-1152.
51. Charleston, S., and Mahmood R. Azimi-Sadjadi. "Reduced order Kalman filtering for the enhancement of respiratory sounds." IEEE Transactions on Biomedical Engineering 43.4 (1996): 421-424.
52. Atbi, A., and S. M. Debbal. "Segmentation of pathological signals phonocardiogram by using the Shannon energy envelogram." Aditi Journal of Computational Mathematics 2.1-2 (2013): 1-14.
53. H. Liang et al., "Heart sound segmentation algorithm based on heart sound envelogram," in Proc. Comput. Cardiol., Lund, Sweden, 1997, vol. 24, pp. 105–108.
54. Sharma, Praveen Kumar, Sourav Saha, and Saraswati Kumari. "Study and design of a Shannon-energy-envelope based phonocardiogram peak spacing analysis for estimating arrhythmic heart-beat." International Journal of Scientific and Research Publications 4.9 (2014): 1-5.
55. Naseri, Hosein, and M. R. Homaeinezhad. "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric." Annals of biomedical engineering 41.2 (2013): 279-292.
56. Mondéjar-Guerra, V., et al. "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers." Biomedical Signal Processing and Control 47 (2019): 41-48.
57. Rubin, Jonathan, et al. "Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients." 2016 Computing in cardiology conference (CinC). IEEE, 2016.
58. Langley, Philip, and Alan Murray. "Heart sound classification from unsegmented phonocardiograms." Physiological Measurement 38.8 (2017): 1658.
59. Gradolewski, Dawid, et al. "A wavelet transform-based neural network denoising algorithm for mobile phonocardiography." Sensors 19.4 (2019): 957.
60. Moukadem, A., et al. "Comparative study of heart sounds localization. "Bioelectronics, Biomedical and Bio-inspired Systems SPIE N8068A-27, Prague (2011).
61. Ghosh, Samit Kumar, et al. "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals." Computers in biology and medicine 118 (2020): 103632.
62. Yadollahi, Azadeh, and Zahra MK Moussavi. "A robust method for heart sounds localization using lung sounds entropy. "IEEE transactions on biomedical engineering 53.3 (2006): 497-502.
63. Hasfjord, F. "Heart sound analysis with time dependent fractal dimensions. "Department of Biomedical Engineering at Linköpings university (2004).
64. Zhang, Di, et al. "Analysis and classification of heart sounds with mechanical prosthetic heart valves based on Hilbert-Huang transform." International journal of cardiology 151.1 (2011): 126-127.
65. D. Kumar et al., Detection of S1 and S2 heart sounds by high frequency signatures. in Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., New York, NY, USA, 2006, vol. 1, pp. 1410–1416.
66. S. Ari et al., A robust heart sound segmentation algorithm for commonly occurring heart valve diseases. J. Med. Eng. Technol., vol. 32, no. 6, pp. 456–65, Jan. 2008.
67. Gamero, L. G., and R. Watrous. "Detection of the first and second heart sound using probabilistic models." Proceedings of the 25th annual international conference of the ieee engineering in medicine and biology society (ieee cat. no. 03ch37439). Vol. 3. IEEE, 2003.
68. Ricke, Anthony D., Richard J. Povinelli, and Michael T. Johnson. "Automatic segmentation of heart sound signals using hidden Markov models." Computers in Cardiology, 2005. IEEE, 2005.
69. Gill, Daniel, Noam Gavrieli, and Nathan Intrator. "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model." Computers in Cardiology, 2005. IEEE, 2005.
70. Koçyiğit, Yücel. "Heart sound signal classification using fast independent component analysis." Turkish journal of electrical engineering & computer sciences 24.4 (2016): 2949-2960.
71. Goda, Márton Aron, and Péter Hajas. "Morphological determination of pathological PCG signals by time and frequency domain analysis." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
72. Krishnan, Sridhar, and Yashodhan Athavale. "Trends in biomedical signal feature extraction." Biomedical Signal Processing and Control 43 (2018): 41-63.
73. Renna, Francesco, Jorge Oliveira, and Miguel T. Coimbra. "Deep convolutional neural networks for heart sound segmentation." IEEE journal of biomedical and health informatics 23.6 (2019): 2435-2445.
74. Tschannen, Michael, et al. "Heart sound classification using deep structured features." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
75. Munia, Tamanna TK, et al. "Heart sound classification from wavelet decomposed signal using morphological and statistical features." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.

76. Homsi, Masun Nabhan, et al. "Automatic heart sound recording classification using a nested set of ensemble algorithms." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
77. Khaled, Sara, Mahmoud Fakhry, and Ahmed S. Mubarak. "Classification of PCG Signals Using A Nonlinear Autoregressive Network with Exogenous Inputs (NARX)." 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE). IEEE, 2020.
78. Whitaker, Bradley M., et al. "Combining sparse coding and time-domain features for heart sound classification." Physiological measurement 38.8 (2017): 1701.
79. Debbal, S. M., and Fethi Bereksi-Reguig. "Computerized heart sounds analysis." Computers in biology and medicine 38.2 (2008): 263-280.
80. Hanbay, Kazım, Muhammed Fatih Talu, and Ömer Faruk Özgüven. "Fabric defect detection systems and methods—A systematic literature review. "Optik 127.24 (2016): 11960-11973.
81. Mishra, Madhusudhan, et al. "Detection of third heart sound using variational mode decomposition." IEEE Transactions on Instrumentation and Measurement 67.7 (2018): 1713-1721.
82. Alqudah, Ali Mohammad, Hiam Alquran, and Isam Abu Qasmieh. "Classification of heart sound short records using bispectrum analysis approach images and deep learning." Network Modeling Analysis in Health Informatics and Bioinformatics 9.1 (2020): 1-16.
83. El Badlaoui, O., A. Benba, and A. Hammouch. "Novel PCG analysis method for discriminating between abnormal and normal heart sounds." IRBM 41.4 (2020): 223-228.
84. Das, Sangita, Saurabh Pal, and Madhuchhanda Mitra. "Supervised model for Cochleagram feature based fundamental heart sound identification." Biomedical Signal Processing and Control 52 (2019): 32-40.
85. Fahad, H. M., et al. "Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM." Microscopy research and technique 81.5 (2018): 449-457.
86. Alam, Shahnawaz, Rohan Banerjee, and Soma Bandyopadhyay. "Murmur detection using parallel recurrent & convolutional neural networks." arXiv preprint arXiv:1808.04411 (2018).
87. Kay, Edmund, and Anurag Agarwal. "Dropconnected neural network trained with diverse features for classifying heart sounds." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
88. Bobillo, Ignacio J. Diaz. "A tensor approach to heart sound classification." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
89. Chen2016: Chen, Tien-En, et al. "S1 and S2 heart sound recognition using deep neural networks." IEEE Transactions on Biomedical Engineering 64.2 (2016): 372-380.
90. Plesinger, Filip, et al. "Discrimination of normal and abnormal heart sounds using probability assessment." 2016 Computing in Cardiology Conference (CinC). IEEE, 2016.
91. Maglogiannis, Ilias, et al. "Support vectors machine-based identification of heart valve diseases using heart sounds." Computer methods and programs in biomedicine 95.1 (2009): 47-61.
92. Wu, Jimmy Ming-Tai, et al. "Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model." Applied Soft Computing 78 (2019): 29-40.
93. Ghaemmaghami, Houman, et al. "Automatic segmentation and classification of cardiac cycles using deep learning and a wireless electronic stethoscope." 2017 IEEE Life Sciences Conference (LSC). IEEE, 2017.
94. Khan, Faiq Ahmad, Anam Abid, and Muhammad Salman Khan. "Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. "Physiological measurement 41.5 (2020): 055006.
95. Krishnan, Palani Thanaraj, Parvathavarthini Balasubramanian, and Snekhalatha Umapathy. "Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network." Physical and Engineering Sciences in Medicine (2020): 1-11