IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 7, 2020, accepted August 27, 2020, date of publication September 1, 2020, date of current version September 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020806

Time-Frequency Analysis, Denoising, Compression, Segmentation, and Classification of PCG Signals

MD. TANZIL HOQUE CHOWDHURY[1], KHEM NARAYAN POUDEL[1], (Member, IEEE), AND YATING HU[2]

[1]Computational Science, Middle Tennessee State University, Murfreesboro, TN 37132, USA
[2]Engineering Technology, Middle Tennessee State University, Murfreesboro, TN 37132, USA

Corresponding author: Md. Tanzil Hoque Chowdhury (mc7d@mtmail.mtsu.edu)

ABSTRACT Phonocardigraphy (PCG) is the graphical representation of heart sounds. The PCG signal contains useful information about the functionality and the condition of the heart. It also provides an early indication of potential cardiac abnormalities. Extracting cardiac information from heart sounds and detecting abnormal heart sounds to diagnose heart diseases using the PCG signal can play a vital role in remote patient monitoring. In this paper, we have combined different signal processing techniques and a deep learning method to denoise, compress, segment, and classify PCG signals effectively and accurately. First, the PCG signal is denoised and compressed by using a multi-resolution analysis based on the Discrete Wavelet Transform (DWT). Then, a segmentation algorithm, based on the Shannon energy envelope and zero-crossing, is applied to segment the PCG signal into four major parts: the first heart sound (S1), the systole interval, the second heart sound (S2), and the diastole interval. Finally, Mel-scaled power spectrogram and Mel-frequency cepstral coefficients (MFCC) are employed to extract informative features from the PCG signal, which are then fed into a classifier to classify each PCG signal into a normal or an abnormal signal by using a deep learning approach. For the classification, a 5-layer feed-forward Deep Neural Network (DNN) model is used, and overall testing accuracy of around 97.10% is achieved. Besides providing valuable information regarding heart condition, this signal processing approach can help cardiologists take appropriate and reliable steps toward diagnosis if any cardiovascular disorder is found in the initial stage.

INDEX TERMS Classification, deep neural network, denoising, discrete wavelet transform, phonocardiogram, segmentation, Shannon energy envelope, TensorFlow, murmur, zero-crossing.

I. INTRODUCTION

Phonocardiography (PCG) is an automatic computer-aided diagnosis tool that is the graphical depiction of heart sounds. It provides information about the time duration, frequency, and other important parameters of heart sounds to determine the functionality and the current condition of the heart valves [1]. Identifying pathological symptoms by hearing heart sounds through a stethoscope is a very difficult skill and may take a long time to gain proficiency. Moreover, the human ear has the limitation in hearing heartbeats. So, we need more objective tools like PCG to extract informative characteristics of heart sounds that cannot be detected by the human ear. PCG helps to analyze heart sounds and to detect abnormalities in the heart, thereby improving overall diagnosis efficiency.

A normal PCG recording usually consists of two fundamental heart sounds called the first heart sound (S1) and the second heart sound (S2), which are generated due to the closure of the atrioventricular valves and semilunar valves, respectively. The interval from the starting point of S1 to the starting point of S2 is called the systole interval, and the interval from the starting point of S2 to the starting point of S1 is called the diastole interval [1], [2]. The diastole interval is usually longer than the systole interval [3]. Beside S1 and S2, two extra heart sounds known as the third and fourth heart sound (S3 and S4) can appear in both normal and pathological conditions. S3 appears just after S2, and S4 appears just before S1 [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

TABLE 1. The four basic heart sounds and their frequency range.

Sound	S1	S2	S3	S4
Frequency	30-100 Hz	Above 100 Hz	20-25 Hz	Below 30 Hz

FIGURE 1. A pathological PCG signal.

Table 1 shows the four basic heart sounds and their properties. Aside from these heart sounds, different kinds of murmurs may also be present in the PCG signal, which are produced due to turbulent flow of blood across the valves and are related to cardiac diseases. Murmurs usually have a higher frequency when compared to normal heart sounds [4]. Fig. 1 shows a PCG signal with murmurs and S3 with reference to S1 and S2.

While recording, PCG signal is usually interrupted by different kinds of noise and unnecessary information that may cause inaccurate clinical diagnosis. So, before segmentation, it is required to remove these redundant pieces of information from the PCG signal to evaluate the proper functionality of the heart. Fourier transform (FT) and Short-time Fourier transform (STFT) are commonly used tools for examining stationary signals, but they show limited performance in examining non-stationary signals, as they are unable to provide simultaneous time and frequency localization [5]. Non-stationary signals, like PCG, can be analyzed properly by using DWT, as DWT provides very good time-frequency localization [5]. By using a multi-resolution analysis technique based on DWT, the PCG signal can be decomposed into different sub-bands having different frequency ranges. The required sub-bands containing valuable clinical information can be picked up for further analysis. The other sub-bands, which contain murmurs, noise, or other unnecessary information, can be discarded [6], [7].

The location and duration of the four basic heart sounds, systole, and diastole intervals are important parameters to determine the cardiac function of a person [8]. To extract these important characteristics from the PCG signal, it is required to segment the PCG signal properly [9]. For segmentation, the normalized average Shannon energy is used to detect the components of the PCG signal by calculating the envelope of its energy [10], [11]. Different methods are available in scientific literature to detect the envelope of signals such as absolute value of signals, energy of signals, Shannon entropy, Shannon energy, and so on. The absolute value gives the same weight to all of the components, therefore it is difficult to separate high amplitude signals from low amplitude signals using this procedure. The energy (square) gives weight to high amplitude rather than low amplitude signals. The Shannon entropy gives more weight to low-intensity rather than high-intensity signals. Overall, the Shannon energy gives a better result when compared to the other methods by giving emphasis on the signals having medium intensity, which reduces the impact of low amplitude signals more than high amplitude signals. Thus, it is possible to detect the difference of the envelope intensity of the high and low amplitude sounds [10], [11]. Then, the zero-crossing algorithm can be used to detect the starting and stopping points of the basic heart sounds in the PCG signal [12]. Based on this information, the duration of each heart sound, systole, and diastole intervals can be obtained.

Due to the significance of the PCG signal to detect cardiac abnormalities, different automatic PCG classification models have been developed using deep learning approaches such as Deep Neural network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and so on. All of these classification models use feature extraction methods to remove redundant features and to increase the classification accuracy. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are some of the most common and widely used algorithms for feature extraction. By analyzing the spectrogram of the PCG signal, we found that the PCG signal has the same properties as the speech signal. Mel-scaled power spectrogram and MFCC have been shown to be more effective compare to PCA and LDA to extract important features and to differentiate between different speech signals. Therefore, Mel-scaled power spectrogram and MFCC are used in our study to extract informative features from the raw PCG signal, which are then passed through a 5-layer feed forward DNN model trained by Keras. The performance of this PCG classification model has been compared with 12 other state-of-the-art PCG classification models, and our PCG classification model has outperformed those models by a large margin.

II. METHODS

A. MULTI-RESOLUTION ANALYSIS USING DISCRETE WAVELET TRANSFORM

DWT can analyze the signal with different resolutions at different frequencies. At high frequencies, a good time and poor frequency resolution can be achieved. Similarly, at low frequencies, a good frequency and poor time resolution can be achieved by using DWT [5]. The time-frequency analysis of a signal x by using DWT can be accomplished by passing the signal through a series of high-pass filters to analyze the high frequencies and low-pass filters to analyze the low frequencies [5]. A half-band low-pass filter removes all frequencies that are above half of the highest frequency in the signal. Similarly, a half-band high-pass filter removes all frequencies that are below half of the highest frequency in the signal. Using this procedure, the desired frequency band can be picked up for further processing. First, the samples of the signal x are decomposed simultaneously using a low-pass filter of impulse response g and a high-pass filter of impulse

FIGURE 2. Block diagram of the filter analysis.

response h, which results in the convolution of the signal x with the filters. After the process of filtering, as half of the frequencies have been removed, half of the total samples can be rejected according to the Nyquist rule. So, the signal is down-sampled by 2 to discard half of the total samples from the signal. This process is called one level decomposition and mathematically can be expressed as:

$$Y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k] * g[2n-k] \tag{1}$$

$$Y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k] * h[2n-k] \tag{2}$$

where $Y_{low}[n]$ and $Y_{high}[n]$ are the outputs of the low-pass and high-pass filters, respectively, after down-sampled by 2. The output from the low-pass and high-pass filters are called the approximation coefficients and detail coefficients, respectively [5]. The block diagram of the filter analysis is shown in Fig. 2.

Decomposition can be repeated if it is further required to get the desired frequency spectrum from the input signal.

B. SHANNON ENERGY ENVELOPE AND ZERO-CROSSING ALGORITHM

The Shannon energy is used in our research to accurately identify the boundaries of all the basic heart sounds (S1, S2, S3, and S4). The mathematical expression of this procedure is given below:

$$Shannon\ Energy : E = -x^2 log x^2 \tag{3}$$

where x is the denoised signal. The average Shannon energy is calculated in 0.02 second continuous segments of the whole PCG signal, with segment overlap of 0.01 second. The average Shannon energy can be represented as:

$$E_s = -\frac{1}{N}\sum_{i=1}^{N} x^2(i) log x^2(i) \tag{4}$$

where x is the denoised PCG signal, N is the signal length, and i is an integer. Lastly, the normalized average Shannon energy is calculated to get the Shannon energy envelop of the signal. Mathematically, the normalized average Shannon energy can be written as:

$$E_n(t) = \frac{E_s(t) - M(E_s(t))}{max(|E_s(t)|)} \tag{5}$$

where $E_n(t)$ is the normalized average Shannon energy, $E_s(t)$ is the average Shannon energy, $M(E_s(t))$ is the mean value of $E_s(t)$, and $max(|E_s(t)|)$ is the maximum absolute value among all the coefficients of $E_s(t)$, respectively. After calculating the boundary of each heart sound accurately, a zero-crossing algorithm can be used to know the starting and stopping points of each heart sound by calculating the points where the sign of the boundaries change from positive to negative or vice versa.

C. FEATURE EXTRACTION

Feature extraction is a process of deriving a compact and useful representation of the information from the signal [13], [14]. The heart sound signals in the database are redundant in nature. 1 second of data contains 2,000 samples with a sampling frequency of 2,000 Hz. Therefore, we need to extract the necessary and meaningful features from the signal to train the model. Mel-scaled power spectrogram and MFCC are used in our research to extract important features from the PCG signal.

1) MEL-SCALED POWER SPECTROGRAM

Time vs. Frequency representation of a signal is called the spectrogram of the signal. A spectrogram visually represents the change of the frequency of a signal with respect to time, which helps the model to recognize the sound accurately [15]. The Mel-scale aims to mimic the non-linear human ear perception of sound, by being more distinctive at lower frequencies and less distinctive at higher frequencies. The Mel-scaled filters are non-uniformly placed in the frequency axis to simulate human ear properties. Thus, there are more filters in the low-frequency region and fewer filters in the high-frequency region. A Mel-scaled power spectrogram of a signal can be found by applying Mel-scaled filters to the power spectrum of a signal and the neural network works much better if the Mel-scaled power spectrogram is used instead of the spectrogram [15]. The process of obtaining the Mel-scaled power spectrogram of a signal is shown in Fig. 3.

First, the signal is divided equally into small sections of short duration (20 to 30 ms) known as frames. Then, each frame is multiplied by the Hamming window. The Hamming window can be expressed as:

$$w[n] = 0.54 - 0.46 cos(\frac{2\pi n}{N-1}) \tag{6}$$

where $0 \leq n \leq N-1$, and N is the window length. Then, the Discrete Fourier transform (DFT) is applied to convert the signal from the time domain to the frequency domain. The Mel-scale filter-banks are computed as follows:

$$m = 2595\ ln(\frac{f}{700} + 1) \tag{7}$$

where f is the frequency in the linear scale, and m is the resulting frequency in Mel-scale. Now, the Mel-scaled power spectrogram of the signal is obtained by applying Mel-scale filter-banks to the power spectrum of the signal and the log

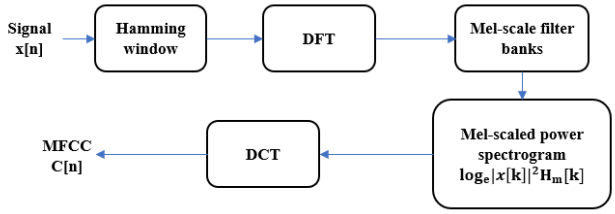


FIGURE 3. a) Feature extraction process by using Mel-scaled power spectrogram and MFCC.

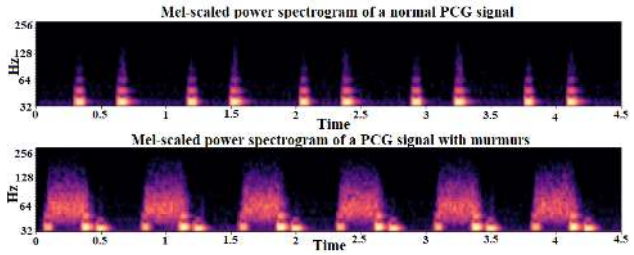


FIGURE 4. a) Mel-scaled power spectrogram of a normal PCG signal. b) Mel-scaled power spectrogram of a PCG signal with murmurs.

of the energy output of each filter. This can be expressed as:

$$S[m] = log(\sum_{k=0}^{N-1} |x[k]|^2 H_m[k]) \tag{8}$$

where $H_m[k]$ is the filter-banks, and m is the number of the filter-bank. Fig. 4 shows the Mel-scaled power spectrogram of a normal signal (without murmurs) and an abnormal signal (with murmurs), respectively.

2) MEL-FREQUENCY CEPSTRAL COEFFICIENTS

Mel-frequency cepstrum (MFC) is the compressed representation of the Mel-scaled power spectrogram, which can be found by taking the Discrete Cosine Transform (DCT) of a log power spectrum on a nonlinear Mel-scale of frequency [15]. The process of obtaining the Mel-scaled power spectrogram of a signal is shown in Fig. 3. The DCT of the spectrum to obtain the MFCC can be represented as:

$$c[n] = \sum_{m=0}^{M-1} S[m]\, cos(\frac{\pi n}{M}(m - \frac{1}{2})), \quad n = 0, 1, 2, \ldots, M \tag{9}$$

where M is the total number of filter banks. Fig. 5 shows the MFCC of a normal signal (without murmurs) and an abnormal signal (with murmurs), respectively.

The features of the Mel-scaled power spectrogram and the MFCC are biologically inspired and resemble the resolution of the human auditory system, which (features) are proven to be more efficient to discriminate between two different sound signals [15].

D. DEEP NEURAL NETWORK

A DNN is the network of artificial neurons with multiple hidden layers between input and output layers. These neurons

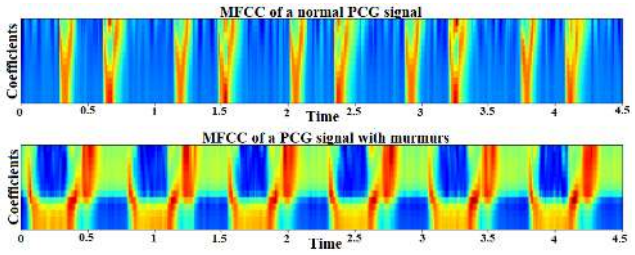


FIGURE 5. a) MFCC of a normal PCG signal. b) MFCC of a PCG signal with murmurs.

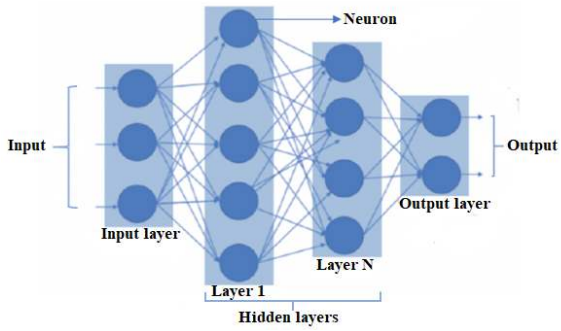


FIGURE 6. A deep neural network with N hidden layers.

usually create a complex network of different layers. Neurons from one layer pass signals to other neurons in the next layer. Fig. 6 represents a DNN of N hidden layers.

From Fig. 6, we can see that the input data is fed into the neurons of the input layer. The output of the input layer works as input to the first hidden layer. This process will continue until the final layer. The output of the final layer will give the final prediction. Each layer can have one or more neurons and each neuron uses a threshold value in the form of an activation function to pass the signal to the next connected neuron. Two neurons of consecutive layers are connected with a parameter called weight. The function of the weight is to transform the input data within the hidden layers. While training the model, DNN uses a backpropagation algorithm to provide feedback to the network based on the output. The goal of the backpropagation algorithm is to update each of the weights several times step-by-step, thereby minimizing the error and gradually increase the overall accuracy [16]. After nth iteration the error at the output of neuron p can be expressed as:

$$e_p(n) = d_p(n) - a_p(n) \tag{10}$$

where $d_p(n)$ and $a_p(n)$ are the desired and actual output of neuron p, respectively. The instantaneous error energy at the output layer is defined as:

$$E(n) = \frac{1}{N} \sum_{p=1}^{N} e_p^2(n) \tag{11}$$

The above error can be reduced by using gradient descent method. The gradient descent is the widely used optimization method to update the weights by calculating the derivative

of the error with respect to the weights of the network. This process can be expressed as:

$$\Delta w_{p,j}(n) = -\eta \frac{\partial E(n)}{\partial w_{p,j}(n)} \tag{12}$$

where η is known as the learning rate. Learning rate is a hyper-parameter, which determines the adjustment of the weights with respect to the loss gradient. The range of the learning rate is between 0 to 1. This process of updating the weights will continue until the loss function is minimum. The final updated value of $\Delta w_{p,j}(n)$ can be expressed as:

$$\Delta w_{p,j}(n+1) = w_{p,j}(n) + \Delta w_{p,j}(n) \tag{13}$$

Thus, by minimizing the error we can find the optimal values for the weights of each neuron that will give the best model performance.

E. PERFORMANCE EVALUATION OF DNN

Classification accuracy of any learning model can be evaluated by investigating the confusion matrix. Two parameters known as sensitivity and specificity are used to analyze the prediction accuracy of any binary classification models. The sensitivity indicates the true positive rate and measures the proportion of the correctly identified actual positives. The specificity indicates the true negative rate and measures the proportion of the correctly identified actual negatives. Both of these parameters can be calculated by using the confusion matrix. Another important metric used to evaluate the classification model is known as accuracy, which is the number of correctly predicted data points out of all the data points. Sensitivity, specificity, and accuracy can be calculated by using these formulas:

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

$$Specificity = \frac{TN}{FP + TN} \tag{15}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

where TP (True Positive) is the number of sick people correctly identified as sick, TN (True Negative) is the number of healthy people correctly identified as healthy, FP (False Positive) is the number of healthy people incorrectly identified as sick, and FN (False Negative) is the number of sick people incorrectly identified as healthy.

III. RESULTS AND ANALYSIS

In this section, we verified the performance of the proposed denoising, compression, segmentation, and classification algorithms, respectively. An extensive simulation was carried out using the Python programming language to implement these algorithms.

A. DATASET

In this paper, the well-known University of Michigan Heart Sound and Murmur Library [17] and the 2016 PhysioNet Computing in Cardiology Challenge database [18] were used for evaluating the performance of different algorithms. There are 23 PCG signals in the University of Michigan Heart Sound and Murmur Library database including 5 normal and 18 pathological. The 2016 PhysioNet Computing in Cardiology Challenge database consists of 6 datasets (A through F) containing a total of 3,240 unique heart sound recordings. The recordings from these 2 databases were collected from both healthy people and patients with confirmed cardiac diseases.These 2 databases are not balanced. The imbalance ratio of normal heart sounds to abnormal heart sounds is 1:4. A total of 123 unique PCG signals were used from these 2 databases to validate our segmentation algorithm. For the classification, we used all the 3,240 PCG signals available in the 2016 PhysioNet Computing in Cardiology Challenge database. We used 90% of the data as the training set to develop the prediction ability of the model, and the remaining 10% of the data was used as the testing set to validate the model.

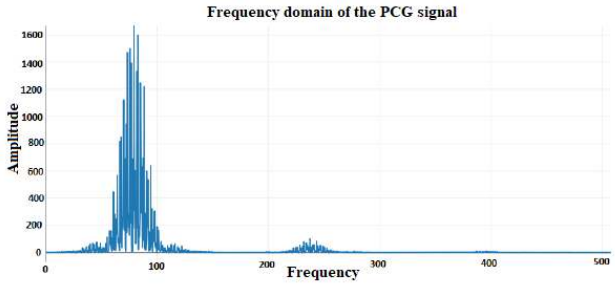


FIGURE 7. Frequency spectrum of a PCG signal.

TABLE 2. Different levels and their frequency range.

Level	Frequency range	Coefficients	Sub-bands
2	0 to 250 Hz	2,276	App. (A_2)
2	251 to 500 Hz	2,276	Detail (D_2)
1	501 to 1,000 Hz	4,517	Detail (D_1)

B. RESULTS OF DENOISING AND COMPRESSION

The presence of different kinds of noise and murmurs make it very difficult to extract the correct diagnostic information and features from the PCG signal. In some cases it is almost impossible to segment PCG signals because of the noise. Therefore, it is necessary to eliminate noise and isolate murmurs from the PCG signal before segmentation. These isolated murmurs can be picked up for further processing. The highest frequency of all the PCG signals in the database is 1,000 Hz. The signal was decomposed in such a way that its approximation band contained all of the information as well as the energy. Fig. 7 shows pathological information of a PCG signal stays within the 0-250 Hz frequency range. Hence, the signal was decomposed up to the 2nd level to cover 0-250 Hz by approximation band. Table 2 shows the frequency range of each sub-band.

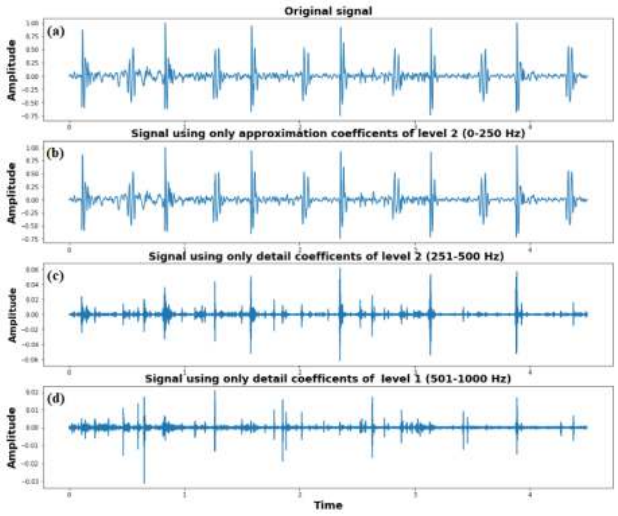


FIGURE 8. a) Original PCG signal. b) Signal using only approximation coefficients of level 2. c) Signal using only detail coefficients of level 2. d) Signal using only detail coefficients of level 1.

Different signals in different sub-bands are shown in Fig. 8. From Table 2 and Fig. 8, it can be observed that only the approximation sub-band (0-250 Hz) contains all the pathological information with only 2,276 coefficients out of 9,069 coefficients. The two detail sub-bands of level 2 (251-500 Hz) and level 1 (501-1,000 Hz) with 6,790 coefficients contain redundant information including noise. In the case of abnormal PCG signal, detail sub-bands contain murmurs besides noise, which can be separated and analyzed further. As the pathological information of the signal usually stays within the frequency range (0-250 Hz) of the approximation band, only the approximation sub-band was used to reconstruct the signal and to eliminate high-frequency noise and murmurs over 250 Hz. The reconstructed signal can be defined as:

$$X_r(n) = A_2(n) \tag{17}$$

where $X_r(n)$ and $A_2(n)$ are the reconstructed and approximation signals, respectively. This process will not only denoise the PCG signal but will also compress the signal without losing any pathological information. It will save storage, so the PCG signal can be recorded for a long time.

C. RESULTS OF SEGMENTATION

Segmentation of the PCG signal facilitates to get the exact position and duration of the basic heart sounds, systole, and diastole intervals. The segmentation technique, based on the Shannon energy envelope and the zero-crossing algorithm, effectively extracted all the important characteristics of the PCG signal. A threshold was set to discard the effect of the noise and the low amplitude signal. Fig. 9 and 10 represent the original signal, the signal after removing noise and murmurs, Shannon energy envelope, and zero-crossing of a normal and an abnormal PCG signal, respectively. The time duration of the 4 basic heart sounds, systole interval, diastole interval,

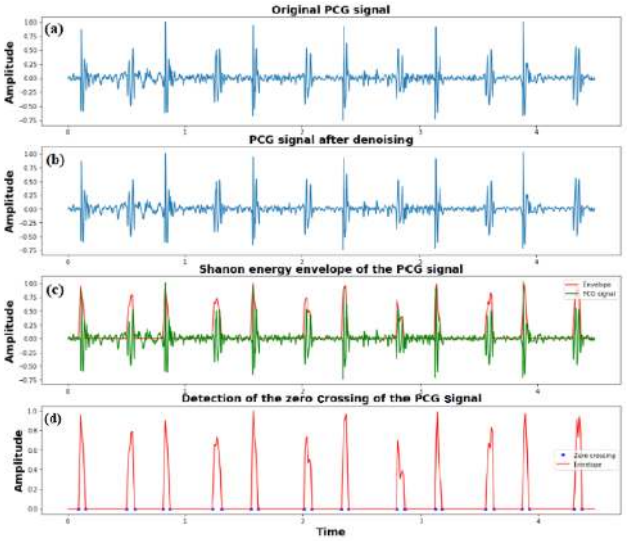


FIGURE 9. a) Original normal PCG signal. b) Reconstructed PCG signal after denoising. c) Shannon energy envelope of the PCG signal. d) Zero-crossing of the PCG signal.

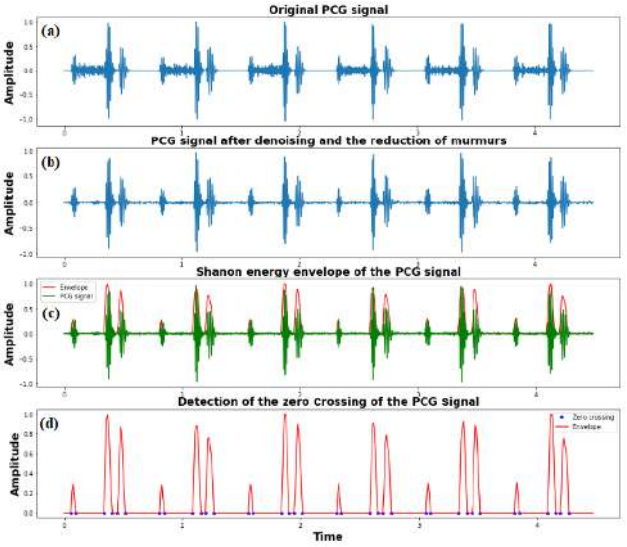


FIGURE 10. a) Original abnormal PCG signal. b) Denoising and isolation of the murmurs from the PCG signal after reconstruction. c) Shannon energy envelope of the PCG Signal. d) Zero-crossing of the PCG Signal.

one cardiac cycle, and heart rate information extracted from a normal and an abnormal PCG signal are shown in Table 3.

D. RESULTS OF CLASSIFICATION

A 5-layer sequential feed-forward DNN model trained by Keras was used in our research to classify the PCG signal into two categories, either normal or abnormal. Keras is the high-level API of TensorFlow, which we used to train our classifying model with great speed.

Mel-scaled power spectrogram and MFCC were used to extract meaningful features from each heart sound of the database. A total of 25 features were achieved from each of the PCG signals using these two methods, which were then

TABLE 3. Extraction of the cardiac information from PCG signals.

Cardiac parameter	Value in a normal PCG signal	Value in an abnormal PCG signal
Duration of S1	0.07 second	0.04 second
Duration of S2	0.06 second	0.07 second
Duration of S3	–	0.07 second
Systole interval	0.31 second	0.28 second
Diastole interval	0.41 second	0.47 second
One cardiac cycle	0.72 second	0.75 second
Heart rate	83 beats per minute	80 beats per minute

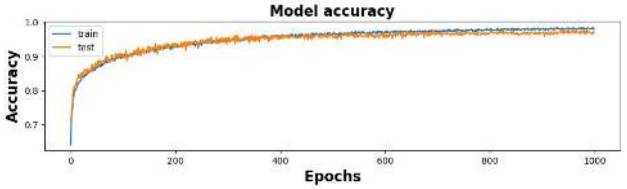


FIGURE 11. Training and testing accuracy with respect to epochs.

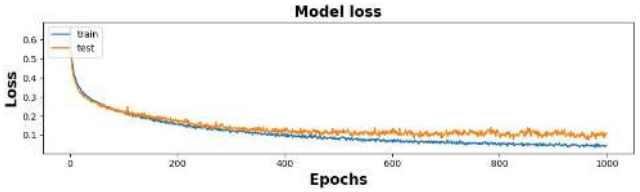


FIGURE 12. Training and testing loss with respect to epochs.

fed into the DNN to train the model. Afterwards, 5 hidden layers with 256, 512, 768, 1,024, and 1,280 filters were implemented with the ReLU activation function for non-linearity. In the output layer, the sigmoid activation function was used to get the probability distribution, which we applied on the cross-entropy cost function. The cross-entropy cost function was used to measure how far apart the output of the model was from that of the desired or target output. The Adam optimizer was used to minimize the cost function. The training started with a learning rate of 0.0001 and continued until it reached the maximum number of epochs. The Dropout technique was used in the model to reduce independent learning among the neurons and to handle overfitting. After training the model its prediction capability was tested on the testing set. This proposed model can discriminate between normal and abnormal PCG signals with an excellent training accuracy of 98.25% and testing accuracy of 97.10%. The achieved sensitivity and the specificity of the model are 99.26% and 94.86%, respectively. Fig. 11 shows the accuracy of the model and Fig. 12 shows the reduction of the cost with respect to epochs.

Based on the result, this model can be a promising solution to detect early-stage heart diseases by picking up potential abnormal PCG signals from a series of normal PCG signals.

IV. DISCUSSION

The main goal of the 2016 PhysioNet Computing in Cardiology Challenge [18] was to build a robust intelligent system that can detect anomaly in the PCG signal and can classify a PCG signal as normal or abnormal based on its features. The best overall accuracy achieved in the official phase of the 2016 PhysioNet Computing in Cardiology Challenge was 86.02% with sensitivity and specificity of 94.24% and 77.81%, respectively. Table 4 shows the comparison of our proposed PCG classification model with 12 other state-of-the-art PCG classification models. All these models used the same dataset published by the 2016 PhysioNet Computing in Cardiology Challenge. Presently, this is the largest database of PCG signals in the world. As shown in Table 4, the classification accuracy achieved from the previous models varied between 79.00% to 97.00%, whereas the range of the sensitivity and specificity varied between 77.00% to 98.33% and 77.81% to 98.00%, respectively. It is noteworthy to mention that the AdaBoost-CNN model proposed by Potes *et al.* [19] was ranked 1st in the 2016 PhysioNet Computing in Cardiology Challenge. Nassaralla *et al.* [20] extracted time and frequency features of PCG signals to build a learning model using RFC and DNN. Nassaralla et al obtained a very good accuracy and specificity of 92.00% and 98.00%, respectively, but low sensitivity of 78.00%. On the other hand, Han *et al.* [23] reached an overall good accuracy and sensitivity of 91.50% and 98.33%, respectively, but with less specificity of 84.67%. They used complex segmentation of heart sounds and CNN to identify PCG signals. Krishnan *et al.* [30] also implemented segmentation of the cardiac cycle and achieved 85.65% accuracy. Sotaquirá *et al.* [26] used DNN and weighted probability comparison of each card cycle and got high accuracy of 92.60%. Langley *et al.* [22] obtained 79.00% accuracy without using complex segmentation technique. They used threshold-based classification tree for PCG classification. Singh *et al.* [27] initially applied KNN on unsegmented heart sounds recording and got 90.00% accuracy. Later, Singh *et al.* [29] improved the accuracy to 92.47% by applying a set of classifiers. Whitaker *et al.* [21], Tang *et al.* [24], and Nogueira *et al.* [28] employed SVM with different structures to build their models and achieved 89.26%, 88.00%, and 87.85% accuracy, respectively. Dominguez et. al. [25] attained a great accuracy of 97.00% by employing the modified version of the AlexNet model but this model has high computational complexity. Our PCG classification model outperforms these state-of-the-art models with a significant improvement in overall accuracy rate. The proposed model achieved an overall accuracy of 97.10% with sensitivity and specificity of 99.26% and 94.86%, respectively. Moreover, it has very low computational complexity with high speed for PCG classification. Hence, this model can overcome the limitation to classify PCG signals accurately with high speed.

Our automatic PCG analysis technique is a complete package of denoising, compression, segmentation, and classification. Until now, very little research has been done on analyzing PCG signals which covers all four major techniques. Clinically, it is essential to extract as many possible features from the PCG signal for the correct classification. However, not all of the features carry important information, and there

TABLE 4. Comparison of the proposed model with other state-of-the-art PCG classification models.

Author	Approach	Sensitivity (%)	Specificity (%)	Accuracy (%)
Potes et al., (2016) [19]	Adaptive Boosting & CNN	94.24	77.81	86.02
Nassralla et al., (2017) [20]	Random Forest Classifier & DNN	78.00	98.00	92.00
Whitaker et al., (2017) [21]	Support Vector Machine	90.00	88.45	89.26
Langley et al., (2017) [22]	Classification Tree	77.00	80.00	79.00
Han et al., (2018) [23]	Threshold & CNN	98.33	84.67	91.50
Tang et al., (2018) [24]	Support Vector Machine	88.00	87.00	88.00
Dominguez et al., (2018) [25]	Modified AlexNet	93.20	95.12	97.00
Sotaquirá et al., (2018) [26]	Probability Comparison & DNN	91.30	93.80	92.60
Singh et al., (2019) [27]	K-Nearest Neighbors	93.00	90.00	90.00
Nogueira et al., (2019) [28]	Wavelet Trasnform & Suport Vector Machine	90.45	85.25	87.85
Sing et al., (2020) [29]	Ensemble of Classifiers	94.08	91.95	92.47
Krishnan et al., (2020) [30]	Segmentation & DNN	86.73	84.75	85.65
Our Study	**MFCC, MSPS & DNN**	**99.26**	**94.86**	**97.10**

can be some redundancy. Thus, denoising and compression techniques were used to remove redundant information from the PCG signal in our proposed method to classify each PCG signal correctly. This preprocessing technique played a crucial role in extracting all of the important cardiac parameters from the PCG signal and in increasing the overall accuracy of the classification and segmentation. During the acquisition, PCG signals usually suffer from distortion while passing from heart to sensor. This happens due to the time delay of different frequency components of the PCG signal during the propagation. For our research we did not use any data acquisition processes to collect PCG signals directly from patients in the clinical environment. Rather, we used a database to validate our proposed classification model. Therefore, minimizing the effect of time delay of the PCG signal during propagation is beyond the scope of this paper.

The main limitation of our proposed algorithm is the existence of murmurs within the frequency range of the basic heart sounds. In this case, the Shannon energy is affected by noise and it is very difficult to accurately identify the boundaries of each heart sound. In the presence of high intensity noise, the Shannon energy envelope is too noisy to read and will provide incorrect output. The DWT reconstructs the PCG signal by separating high-frequency murmurs from the low-frequency heart sounds. However, when the heart sounds and murmurs share the same frequency band, separating murmurs from the PCG signal will eliminate some of the major details of heart sounds. This will cause potential loss of the cardiac information. Additional research is needed to solve this problem.

V. CONCLUSION

PCG signals have been used for decades to detect cardiac abnormalities. The extraction of important cardiac information from the PCG signal and the detection of the abnormal PCG signal in the primary stage can play a vital role to decrease the death rate caused by cardiovascular diseases. An Automatic PCG signal analysis approach using DWT, Shannon energy envelope, zero-crossing, and feed-forward DNN is presented in this paper. This technique is not only able to denoise the signal efficiently but also separates the heart sounds from murmurs. Therefore, it can determine the duration of basic heart sounds as well as the duration of systole interval, diastole interval, and cardiac cycles properly. Classification of the PCG signal is also possible with the classification algorithm with great accuracy of 97.10%, which is better than many other state-of-the-art PCG classification methods. Based on the ability of this algorithm, an automatic wearable tool can be developed to detect the early symptoms of cardiac diseases, allowing for prompt intervention.

However, it should be noted that the proposed method requires a large amount of data to train the model. Therefore, in future work, it is necessary to evaluate the performance of our proposed model by using PCG signals from more subjects. Moreover, we will focus on exploring other important features to improve the classification performance. The efficiency of the segmentation algorithm becomes limited in the presence of a large number of murmurs overlapping with heart sounds. It is a very complex process to remove all the murmurs, and further research is required to overcome this limitation. In the future, other neural network models such as CNN, RNN, LSTM, and GRU will be used to increase the sensitivity, specificity, and accuracy of the model if possible, near 100%.

REFERENCES

[1] R. M. Rangayyan and R. J. Lehner, "Phonocardiogram signal analysis: A review," *Crit. Rev. Biomed. Eng.*, vol. 15, no. 3, pp. 211–236, 1987.
[2] J. Chebil and J. A. Nabulsi, "Classification of heart sound signals using discrete wavelet analysis," *Int. J. Soft Comput.*, vol. 2, no. 1, pp. 37–41, 2007.
[3] L. Hamza Cherif, S. M. Debbal, and F. BEREKSI-REGUIG, "Segmentation of heart sounds and heart murmurs," *J. Mech. Med. Biol.*, vol. 08, no. 04, pp. 549–559, Dec. 2008.
[4] X. Zhang, L. Durand, L. Senhadji, H. C. Lee, and J.-L. Coatrieux, "Time-frequency scaling transformation of the phonocardiogram based of the matching pursuit method," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 972–979, Sep. 1998.
[5] R. Polikar. *The Wavelet Tutorial*. [Online]. Available: http://engineering.rowan.edu/polikar/WAVELETS/WTtutorial.html
[6] M. Chowdhury, K. Poudel, and Y. Hu, "Phonocardiography data compression using discrete wavelet transform," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2018, pp. 01–03.
[7] M. T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Automatic phonocardiography analysis using discrete wavelet transform," in *Proc. 3rd Int. Conf. Vis., Image Signal Process.*, Aug. 2019, pp. 1–8.
[8] J. Zhong and F. Scalzo, "Automatic heart sound signal analysis with reused multi-scale wavelet transform," *Int. J. Eng. Sci.*, vol. 02, pp. 50–57, Jan. 2013.

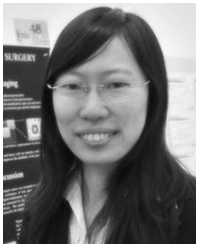
[9] H. Nazeran, "Wavelet-based segmentation and feature extraction of heart sounds for intelligent PDA-based phonocardiography," *Methods Inf. Med.*, vol. 46, no. 02, pp. 135–141, 2007.
[10] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart Sound Segmentation Algorithm Based on Heart Sound Envelogram," *Proc. IEEE Comput. Cardiol.*, vol. 24, pp. 105–108, 1987.
[11] N. Shankar and M. S. Sangeetha, "Analysis of Phonocardiogram for Detection of Cardiac Murmurs using Wavelet transform," *Int. J. Adv. Sci. Tech. Res.*, vol. 1, no. 3, pp. 350–357, 2013.
[12] K. Courtemanche, V. Millette, and N. Baddour, "Heart sound segmentation based on mel-scaled wavelet transform," in *Proc. 31st Conf. Can. Med. Biol. Eng. Soc.*, Montreal, Quebec, Canada, 2008, pp. 1–4.
[13] M. Nabih-Ali, E.-S.-A. El-Dahshan, and A. S. Yahia, "Heart diseases diagnosis using intelligent algorithm based on PCG signal analysis," *Circuits Syst.*, vol. 8, no. 7, pp. 184–190, 2017.
[14] E. F. Gomes and E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction," Tech. Rep., 2012.
[15] J. L. C. Loong, K. S. Subari, M. K. Abdullah, N. N. Ahmad, and R. Besar, "Comparison of MFCC and cepstral coefficients as a feature set for PCG biometric systems," *Int. J. Biomed. Biol. Eng.*, vol. 4, no. 8, pp. 335–339, Jan. 2010.
[16] B. Benuwa, Y. Z. Zhan, B. Ghansah, D. K. wornyo, and F. K. Banaseka, "A review of deep machine learning," *Int. J. Eng. Res. Afr.*, vol. 24, pp. 124–136, Apr. 2016.
[17] (Apr. 2014). *Heart Sound & Murmur Library*. [Online]. Available: https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library.
[18] C. Liu, D. Springer, Q. Li, and B. Moody, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Nov. 2016.
[19] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature:Based and deep learning:Based classifiers for detection of abnormal heart sounds," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 621–624.
[20] M. Nassralla, Z. E. Zein, and H. Hajj, "Classification of normal and abnormal heart sounds," in *Proc. 4th Int. Conf. Adv. Biomed. Eng. (ICABME)*, Oct. 2017, pp. 1–4.
[21] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, and D. V. Anderson, "Combining sparse coding and time-domain features for heart sound classification," *Physiol. Meas.*, vol. 38, no. 8, pp. 1701–1713, Jul. 2017.
[22] P. Langley and A. Murray, "Heart sound classification from unsegmented phonocardiograms," *Physiol. Meas.*, vol. 38, no. 8, pp. 1658–1670, Jul. 2017.
[23] W. Han, Z. Yang, J. Lu, and S. Xie, "Supervised threshold-based heart sound classification algorithm," *Physiol. Meas.*, vol. 39, no. 11, Nov. 2018, Art. no. 115011.
[24] H. Tang, Z. Dai, Y. Jiang, T. Li, and C. Liu, "PCG classification using multidomain features and SVM classifier," *BioMed Res. Int.*, vol. 2018, pp. 1–14, Jul. 2018.
[25] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, "Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 1, pp. 24–34, Feb. 2018.
[26] M. Sotaquirá, D. Alvear, and M. Mondragón, "Phonocardiogram classification using deep neural networks and weighted probability comparisons," *J. Med. Eng. Technol.*, vol. 42, no. 7, pp. 510–517, Oct. 2018.
[27] S. A. Singh and S. Majumder, "Classification of unsegmented heart sound recording using KNN classifier," *J. Mech. Med. Biol.*, vol. 19, no. 4, pp. 24–34, 2019.
[28] D. M. Nogueira, M. N. Zarmehri, C. A. Ferreira, A. M. Jorge, and L. Antunes, "Heart sounds classification using images from wavelet transformation," *Prog. Artif. Intell.*, to be published.
[29] S. A. Singh and S. Majumder, "Short unsegmented PCG classification based on ensemble classifier," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 875–889, Mar. 2020.
[30] P. T. Krishnan, P. Balasubramanian, and S. Umapathy, "Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 505–515, Feb. 2020.

MD. TANZIL HOQUE CHOWDHURY received the B.S. degree in electrical and electronic engineering from the Ahsanullah University of Science and Technology, Bangladesh, in 2014, and the M.S. degree in computer science from Middle Tennessee State University, Murfreesboro, TN, USA, in 2020, where he is currently pursuing the Ph.D. degree in computational science.

His research interests include bio-medical signal processing and deep learning in bio-medical applications.

KHEM NARAYAN POUDEL (Member, IEEE) received the B.E. and M.S. degrees in electronics and communication engineering and information and communication engineering from Tribhuwan University, Kathmandu, Nepal, in 2010 and 2013, respectively, the M.S. degree in electrical and computer engineering from The University of Utah, Salt Lake City, UT, USA, and the M.S. degree in computer science and the Ph.D. degree in computational science from Middle Tennessee State University, Murfreesboro, TN, USA.

His research interests include computational electromagnetics, dielectric multi-layer grating structure, and deep learning in bio-medical signal processing.



YATING HU received the B.S. degree in electrical engineering from Nankai University, in 2008, and the M.S. and Ph.D. degrees from Wayne State University, in 2011 and 2014, respectively.

She joined the Faculty Member of Middle Tennessee State University, in 2014. Her research interests include sensor and instrumentation design, high performance bio-medical sensors, energy harvesting for wireless sensors, and technology commercialization. In particular, she is interested in developing mobile health care related technologies, which enable ubiquitous patient monitoring and proactive health management. She had worked on several federal funded research projects, including wearable sensors for continuous heart and lung sound monitoring (NSF), piezoelectric vibration energy harvesting for self-powered smart sensors (NSF), bio-inspired 3-D tactile sensor for minimally invasive surgery (NSF), and micromachined piezoresistive accelerometers based on an asymmetrically gapped cantilever (NSF).