IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 9, 2020, accepted August 14, 2020, date of publication September 9, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3023044

Automatic Recognition of Fundamental Heart Sound Segments From PCG Corrupted With Lung Sounds and Speech

K. AJAY BABU, (Graduate Student Member, IEEE), AND BARATHRAM RAMKUMAR, (Member, IEEE)
School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar 752050, India
Corresponding author: K. Ajay Babu (abk10@iitbbs.ac.in)

ABSTRACT Automated recognition of fundamental heart sound segments (FHSS) from Phonocardiogram (PCG) is the preliminary step before clinical parameters extraction to detect the presence of abnormality if any. PCG acquisition systems are usually based on microphones. These microphones apart from cardiac sounds will also pick up non-cardiac sounds like lung sounds and speech. The recognition of FHSS is challenging in the presence of non-cardiac events. Deep learning techniques like convolutional neural network (CNN) and recurrent neural network (RNN) are suitable for automated FHSS. However, it will be shown that their performance is degraded in the presence of interference like lung sounds, and speech. Hence in this work, a combination of conventional signal processing technique with deep neural network (DNN) is proposed to enhance the accuracy of automated FHSS. The conventional signal processing technique is based on EWT which can adaptively design the filter banks based on the type of interference. For DNN, U-Net is considered. The method involves the segmentation of PCG using EWT and recognition of FHSS using U-Net based DNN. Envelope features are extracted from the EWT based reconstructed signal and used for training the U-Net based DNN to recognize FHSS. To further improve the recognition accuracy of FHSS, delineation parameters obtained from EWT are incorporated for temporal modeling with the outcomes of U-Net based DNN. The performance of the proposed method is analyzed using both real-time signals and signals taken from standard databases like the Physionet database, and Littmann's lung sound library. Realtime PCG is acquired using an in-house developed PCG acquisition system. The proposed U-Net based DNN with the EWT method achieves FHSS recognition accuracy of 91.17% for PCG with lung sound interference and 90.78% for PCG with speech interference. The proposed method significantly improves the accuracy of FHSS recognition compared to long short term memory (LSTM), and gated recurrent unit (GRU).

INDEX TERMS Recognition, segmentation, fundamental heart sound, lung sound, speech, empirical wavelet transform (EWT), deep neural network (DNN).

I. INTRODUCTION

The blood flow mechanism in the heart will lead to vibrations and generates heart sounds. These heart sounds are used for diagnosis purpose and this technique is known as heart auscultation. Heart auscultation is a simple technique for cardiac diagnosis. In the activity of the heart, two important time intervals are corresponding to ventricular contraction and expansion known as systolic and diastolic periods respectively. The completion of one systolic and diastolic period is known as one heart cycle. The anatomy of the heart is shown in Figure 1 (a) and the location of the heart along with other vibration sources such as lungs and epiglottis are shown in Figure 1 (b). As shown in Figure 1 (a) the tensions generated on the mitral and tricuspid valves (AV valves) during the systolic period results in $S1$ sound. Similarly, the tensions generated on the aortic and pulmonary valves (semilunar valves) during the diastolic period results in $S2$ sound. $S1$ and $S2$ are the two heart sounds normally occurs in healthy adults [3]. For a normal functioning heart, $S1$ sound is longer in duration (of 150 milliseconds [3]) and low pitched whereas $S2$ sound is shorter in duration (60 milliseconds [3]) and high pitched [3]. Also, the diastolic period (the time duration from the $S2$ start point to next $S1$ start point) is longer than the

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

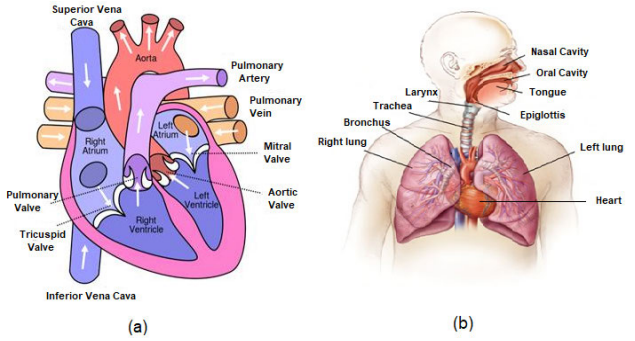

FIGURE 1. Illustrates (a) Heart anatomy [1] (b) Sources (lungs, epiglottis) of interference with heart sounds [2].

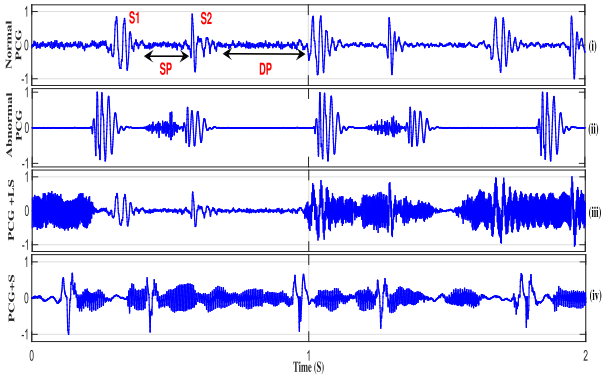

FIGURE 2. Illustrates (i) Normal PCG (SP: Systole pause, DP: Diastole pause) (ii) Abnormal PCG (iii) PCG with lung sounds (PCG+LS) (iv) PCG with speech (PCG+S).

systolic period (the time duration from the $S1$ start point to next $S2$ start point) in a normal condition of the heart [3].

Phonocardiogram (PCG) is a graphical representation of various sounds generated due to the cardiac activity of open and closure of heart valves. The rhythmic vibrations due to the open and closure of valves result in various heart sounds in PCG [3] like $S1$, $S2$, $S3$, $S4$, murmurs, splits, and ejection clicks, etc. As illustrated in Figure 1 (b), due to the proximity of the heart to other vital organs like lungs, and epiglottis, the acquired PCG signals are subjected to interference with non-cardiac vibrations especially from lung sounds and speech. PCG signal with normal cardiac sounds, abnormal cardiac sounds, and interference with non-cardiac sounds are shown in Figure 2. As shown in Figure 2 (i) generally PCG consists of $S1$, systole pause (SP), $S2$, and diastole pause (DP). As shown in Figure 2 (ii) an abnormal PCG consists of murmurs in addition with the fundamental $S1$, and $S2$ sounds. $S1$ and $S2$ sounds are generally low frequency (ranges 30Hz-200Hz [3]) signals and murmurs are randomly varying low amplitude high frequency (ranges up to 1KHz [3]) signal. Though the cardiac diagnosis by heart sounds is inexpensive, the identification of heart sounds in PCG is challenging in the presence of non-cardiac sounds such as lung sounds and speech. A normal PCG signal with the interference of lung sounds, and speech are shown in Figure 2 (iii) and Figure 2 (iv) respectively. Recognizing the fundamental heart sound segments such as $S1$, systole pause, $S2$, and diastole pause have more prominence as they correspond to the systolic and diastolic activities of the heart. The process of identifying the above events is considered as recognition of fundamental heart sound segments (FHSS). From Figure 2 (iii) and Figure 2 (iv), it can be seen that recognizing the FHSS from PCG signal corrupted with lung sounds and speech interference is challenging.

A. MOTIVATION

Biomedical signals such as Electrocardiogram (ECG), Photoplethysmogram (PPG), and Phonocardiogram (PCG) plays a key role in the assessment of cardiac-related issues. ECG provides the electrical activity of the heart, PPG provides the variations in the blood volume, and PCG provides the mechanical activity of the heart. Each biomedical signal has its own identity in cardiac diagnosis. However, ECG and PPG are more susceptible to noise due to the movement of patient. Hence, the first priority of any medical practitioner for cardiac diagnosis is to check the heart sounds using a stethoscope. Heart auscultation is a simple technique to estimate the heart condition. The demand for wearable and automated healthcare devices have been increased due to the availability of low-cost sensors, embedded processors, and communication modules. With the availability of low cost, less power consumption, and memory capable embedded processors the steps towards the design of automated digital stethoscope have been initiated [4]. Most of the PCG acquisition systems embedded in electronic stethoscopes consists of electret condenser microphone sensor of the frequency range 20Hz-20kHz [5], [6] and may also pick up the lower frequency range signal due to air leakage [7]. Frequency ranges of lung sound and speech overlap with that of fundamental heart sounds, hence eliminating them using sensors or filters at the acquisition level is not possible. Therefore it is required to choose an effective signal processing method that can reconstruct the fundamental heart sounds from the PCG corrupted with lung sounds, and speech. For the automation of a digital stethoscope, it is required to develop and test the robustness of intelligent learning algorithms. To the best of our knowledge, there is no particular study of recognizing fundamental heart sound segments from PCG corrupted with lung sounds and speech. Hence the motivation of this work is, limitations in PCG acquisition due to microphone sensors and lacking the study of automated algorithms for recognition of FHSS from PCG corrupted with lung sounds, and speech.

B. STATE OF THE ART

In [8] different automated techniques for heart sound classification are summarized. In [9] Mel-frequency cepstral coefficients (MFCC) extracted from realtime recorded PCG and then obtained the refined features using the K-Means clustering algorithm. The obtained features fed to DNN classifier for segmentation of $S1$ and $S2$ sounds. In [10] four different envelograms extracted from PCG and fed to DNN. Hidden

semi Markov model-based temporal modeling is applied to the output of the DNN for classification of FHSS. In [11] seven different features extracted from PCG are fed to DNN for classification of FHSS. In [12] nine different feature selection algorithms used for choosing the effective features for classification of $S1$ and $S2$ sound. The obtained features are fed to DNN based stacked autoencoder classifier. In [13] the $S1$ and $S2$ scalograms are classified using DNN. The scalograms are obtained using a continuous wavelet transform. In [14] adaptive sojourn hidden semi Markov model (HSMM) based heart sound segmentation is performed. In [15] adaptive sojourn hidden semi Markov model (HSMM) based heart sound segmentation is performed. In [14] Markov switching autoregressive model used for model the raw heart sounds for heart sound segmentation. In [16] the features obtained from variational mode decomposition and Hilbert transformation are utilized with machine learning methods for identification of $S1$ and $S2$ heart sounds. But in all the existing methods there is no particular investigation on the performance of DNN classifiers when PCG corrupted with noises such as lung sounds, and speech.

In general to eliminate noise, conventional signal processing methods involve decomposing the acquired PCG signal into various time-frequency (TF) components using transforms like discrete wavelet transform (DWT) [17], [18], wavelet packet transform (WPT) [19], synchrosqueezing wavelet transform (SSWT) [20] and empirical wavelet transform (EWT) [21], [22]. Decomposition based techniques are proposed where PCG signal is decomposed into different modes using nonstationary decomposition techniques like empirical mode decomposition (EMD) [23], ensemble empirical mode decomposition (EEMD) [24], and variational mode decomposition (VMD) [25], [26]. Interference-free PCG signal is then reconstructed by eliminating the modes or time-frequency components that correspond to various noises and artifacts. Most of these works consider PCG signals corrupted with interference like additive white Gaussian noise (AWGN), baseline wander (BW), and murmurs. Only a few works in the state of the art have considered PCG corrupted with lung sounds. In [27] singular spectrum analysis (SSA) based method is analyzed for localizing heart sounds in respiratory signals. Adaptive line enhancement (ALE) method is presented in [28] for removal of wheeze sounds from PCG. Temporal feature-based methods are presented in [29], [30] for reconstructing $S1$ and $S2$ sounds. Also non-stationary signal decomposition techniques like EMD, and EEMD [23], [24] removed the lung sounds by proper selection of mode. The ALE based methods require simultaneous recording of lung sounds and heart sounds and therefore has synchronization issues. TF based methods employ fixed filter banks and hence are not effective to remove lung sounds and speech which have overlapping frequency content with PCG. The decomposition-based methods require proper selection of stopping criteria and also require further statistics to reject the lung sounds. To the best of our knowledge, there is no research work on eliminating lung sounds, and speech interference from PCG signal. Also there is no particular study on automated recognition of FHSS from PCG corrupted with lung sounds, and speech.

C. CONTRIBUTION

In this paper the combination of a conventional signal processing method and DNN is proposed for recognizing FHSS from PCG signal corrupted with lung sounds, and speech. Empirical wavelet transform is used as conventional signal processing method for FHSS and U-Net based DNN is used for recognition of FHSS. The effectiveness of using EWT for PCG corrupted with additive white Gaussian noise (AWGN), and murmurs are investigated in [21]. However, interference like lung sound and speech is not considered in [21]. In our previous contribution, the effectiveness of EWT in removing lung sounds for the reconstruction of FHSS is investigated in [22].

The motivation for using EWT is that it employs adaptive filter bank which is constructed based on the characteristics of the processing signal. Also, it provides high frequency resolution around the frequency ranges of $S1$ and $S2$ sound and hence provides the better reconstruction of the PCG signal. The main advantage of U-Net based DNN is that low-level features in the encoder part are concatenated with the corresponding high-level features in the decoder part which helps in the recognition of FHSS. The proposed method involves estimating the frequency ranges of clean PCG, PCG with lung sound, and PCG with speech. The dominating frequency ranges of $S1$ and $S2$ sounds are then incorporated into the EWT to construct the adaptive filter bank. The $S1$ and $S2$ sounds are reconstructed from the output of the filter bank. Then smoothed Shannon entropy envelogram is computed over the reconstructed signal. The smoothed signal is followed by adaptive thresholding to find the delineation parameters of $S1$ and $S2$ sounds. Four envelogram features obtained from reconstructed signal is used for training the U-Net based DNN and delineation parameters are utilized for temporal modeling. The performance of the proposed method is analyzed using both real-time signals and signals taken from standard databases like the Physionet database, and Littmann's lung sound library. Real-time PCG is acquired using in-house developed PCG acquisition system.

The rest of the paper is organized as follows: In section II, materials are presented. The proposed method is presented in section III. Results and discussion presented in section IV followed by the conclusion.

II. MATERIALS

In this paper, EWT and U-Net based DNN is used for effective recognition of FHSS. Brief description of the EWT and U-Net based DNN are presented in this section and the detailed description can be found in [31], [32].

A. BRIEF OVERVIEW OF EWT

EWT initially proposed in [31] has been applied to various fields including seismic data analysis [33],

electroencephalogram seizure detection (EEG) [34], power quality analysis [35], and PCG [21], [22]. EWT is similar to classical wavelet transform except that the scaling function ($\phi_1(\omega)$), and empirical wavelet function ($\psi_n(\omega)$) are adaptive in nature. That is the scaling function and empirical wavelet function are adaptively chosen according to the frequency content of the processed signal $x(t)$.

EWT decomposes the processed signal $x(t)$ into N modes. The detailed EWT coefficients for the $n^{th}(n = 1, 2, \ldots, N)$ mode is obtained as [31],

$$D_n(t) = \int x(\tau)\psi_n(\tau - t)d\tau = IFT(X(\omega) \times \psi_n(\omega)) \quad (1)$$

where $\psi_n(\omega)$ is the adaptive empirical wavelet which depends on the frequency content of $x(t)$.

The approximation coefficient is obtained by [31],

$$I_o(t) = \int x(\tau)\phi_1(\tau - t)d\tau = IFT(X(\omega) \times \phi_1(\omega)) \quad (2)$$

where $\phi_1(\omega)$ is the adaptive scaling function which depends on the frequency content of $x(t)$.

The adaptive $\psi_n(\omega)$ and $\phi_1(\omega)$ are given by [31],

$$\psi_n(\omega) = \begin{cases} 1, \ if (1+\gamma)\Omega_i \leqslant |\omega| \leqslant (1-\gamma)\Omega_{i+1} \\ \cos(\frac{\pi}{2}\beta(\gamma, \Omega_{i+1})), \\ \qquad if \ (1-\gamma)\Omega_{i+1} \leqslant |\omega| \leqslant (1+\gamma)\Omega_{i+1} \\ \sin(\frac{\pi}{2}\beta(\gamma, \Omega_i)), \ if \ (1-\gamma)\Omega_i \leqslant |\omega| \leqslant (1+\gamma)\Omega_i \\ 0, \ otherwise \end{cases} \quad (3)$$

and

$$\phi_1(\omega) = \begin{cases} 1, & if \ |\omega| \leqslant (1-\gamma)\Omega_1 \\ \cos(\frac{\pi}{2}\beta(\gamma, \Omega_1)), & if \ (1-\gamma)\Omega_1 \leqslant |\omega| \leqslant (1+\gamma)\Omega_1 \\ 0, & otherwise \end{cases} \quad (4)$$

where γ is overlap parameter, and β is given by [31],

$$\beta(\gamma, \omega) = \begin{cases} 0, & if \ (\gamma, \omega) \leqslant 0 \\ 1, & if \ (\gamma, \omega) \geqslant 1 \\ \beta(\gamma, \omega) + \beta(1 - (\gamma, \omega)), & if \ (\gamma, \omega) \in [0, 1] \end{cases} \quad (5)$$

Ω_i is the boundary parameter given by [31],

$$\Omega_i = \frac{\omega_i + \omega_{i+1}}{2} \ for \ \ 1 \leqslant i \leqslant N - 1 \quad (6)$$

where ω_i, ω_{i+1} are dominant frequencies present in $x(t)$.

The reconstructed signal is given by [31],

$$\hat{x}(t) = I_0(t) \star \phi_1(t) + \sum_{n=1}^{N} D_n(t) \star \psi_n(t)$$
$$= IFT(I_0(\omega)\phi_1(\omega) + \sum_{n=1}^{N} D_n(\omega)\psi_n(\omega)) \quad (7)$$

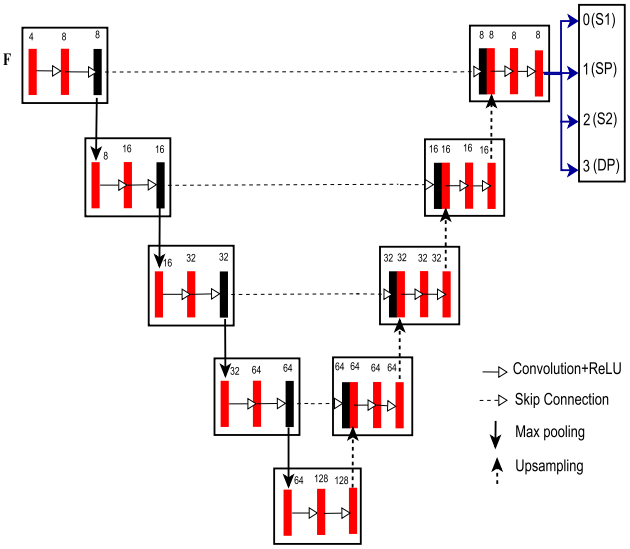

FIGURE 3. Illustrates the 1D variant of U-Net based DNN.

The application of EWT to a processed signal $x(t)$ involves proper choosing of the boundaries Ω_i, overlap parameter, and the number of modes.

B. U-NET BASED DEEP NEURAL NETWORK

Two-dimensional U-Net based DNN is a powerful segmentation model for various biomedical image segmentation [32]–[36]. In a [10] 1D variant of U-Net based DNN is used for the segmentation of FHSS. U-Net based DNN consists of an encoder-decoder structure with a bottleneck layer as shown in Figure 3. For the convenience of representation, a group of convolution layers are kept in a block and named it as mass block. Each convolution layer (shown as red and black colored vertical rectangle bars) consists of 1D convolutions of input with filters of different dimensions. The number of channels for each convolution layer is indicated on the top of the rectangle bars in Figure 3. To restrict in getting higher values of activations, batch normalization (BN) is used. ResNet blocks are used in U-Net based DNN to get the smoother surface of the loss landscape and hence it is easy to perform optimization. The output of each convolution layer is followed by a rectified linear unit (ReLU) activation function which helps in eliminating the negative values. In U-Net based DNN after every mass block, max-pooling operation is performed. Max-pooling will reduce the dimensions of input by a factor of 2 from one mass block to next lower mass block and hence helps in the compact representation of the input. Encoder and decoder parts are connected by the bottleneck layer which contains most of the information of input. To get more efficiency with the model in recognition of input events, skip connections are established by adding the final layer of each mass block in the encoder part with the decoder part as shown in Figure 3. Basically, skip connections in U-Net based DNN are supplying the additional information to the network. Now at the output part, it is required to maintain the same input dimension. This is done with upsampling by a

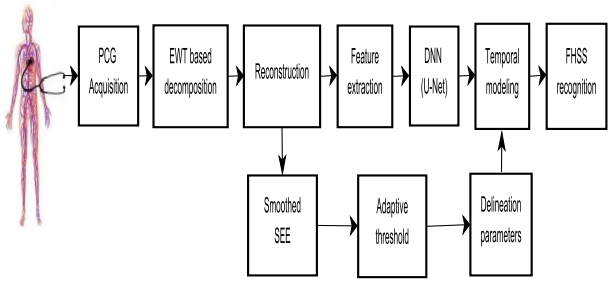

FIGURE 4. Illustrates the block diagram of the proposed method.

factor of 2. Upsampling layer contains transpose convolution followed by ReLU activation. The advantage of the U-Net based DNN is low-level features in the encoder path are concatenated with corresponding high-level features in the decoder path and hence increase the efficiency of recognition of events in the input signal.

III. PROPOSED METHODOLOGY

The block diagram of the proposed method is shown in Figure 4. It consists of EWT based reconstruction, detection of delineation parameters, and U-Net based recognition of FHSS. In the following subsections, the role of individual blocks to attain the objective of recognizing FHSS is explained in detail.

A. PCG ACQUISITION

The experimental setup for real-time recording is shown in Figure 5. A microphone is placed into one of the ear-tips of the stethoscope to get the electrical signal. The obtained electrical signal is passed through a high-pass filter (with cut-off frequency of 1Hz) to remove the DC component. The filtered signal is amplified (with the gain of 101) up to acceptable level using a non-inverting amplifier. The amplified signal is essentially PCG. The PCG signal acquisition without speech disturbance is shown in Figure 5 (a) and the PCG signal corrupted with speech is shown in Figure 5 (b). The obtained PCG signal is connected to the analog input pins of Arduino Uno. The analog to digital converter (ADC) of ATmega 328 micro-controller with 10 bit resolution, 16 MHz clock speed, 32 KB flash memory, 2 KB static random access memory (SRAM), and 1 KB electrically erasable programmable read-only memory (EEPROM) is used for digitizing the analog PCG signal. The digitized PCG signal is given to a computer system and saved the data in a text file using Arduino Uno software. In the pre-processing step, amplitude of the acquired signal is normalized.

1) EWT BASED DECOMPOSITION AND RECONSTRUCTION OF S1 AND S2 SOUNDS

As mentioned earlier, applying EWT to PCG signal $P_c[n]$ requires proper selection of the number of modes, frequency boundaries (Ω_i), and overlap parameter (γ). In order to do that the spectrum of $P_c[n]$ (with various interference) is estimated using fast Fourier transform (FFT) and is denoted

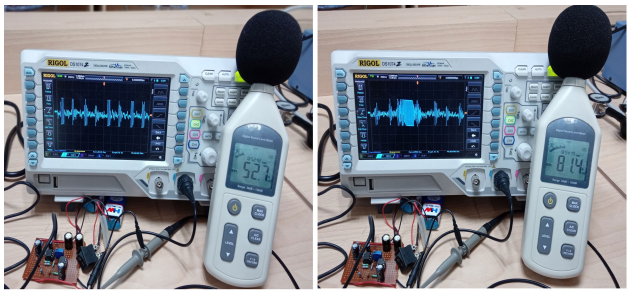

FIGURE 5. Illustrates the experimental setup of the PCG acquisition. (a) PCG acquisition without speech interference. (b) PCG acquisition with speech interference.

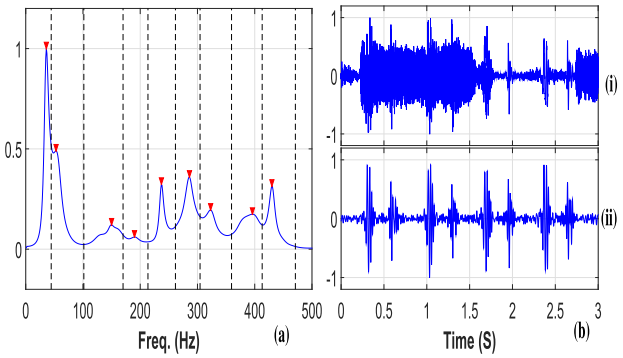

FIGURE 6. Illustrates the estimated frequency spectrum for PCG with lung sound and decomposition using EWT. (a) Detected boundaries on the average and smoothed spectrum of PCG corrupted with lung sounds. (b). (i) PCG with lung sounds. (ii) Reconstructed PCG using EWT.

as $P[\omega]$. The local maxima in the spectrum $P[\omega]$ are found based on amplitude thresholding suggested in [31]. Let $m = (m_i)_{i=1,2,...,N}$ denote set of local maxima and $\omega = (\omega_i)_{i=1,2,...,N}$ denote their corresponding frequency locations. The frequency spectrum is then segmented into N segments whose boundaries are denoted as $\Lambda_i = [\Omega_{i-1}, \Omega_i]$, where Λ_i is computed using (6). Thus the spectrum $P(\omega)$ is segmented into N modes whose boundaries are $[0, \Omega_1]$, $[\Omega_1, \Omega_2]$,...,$[\Omega_{m-1}, \frac{F_s}{2}]$. In each of these segments $\psi_n(\omega)$ and scaling function $\phi_1(\omega)$ are computed using (3) and (4) respectively. The value of γ is chosen by $\gamma = min_n(\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n})$ as suggested in [31].

An example of segmenting the averaged and smoothed spectrum of a PCG signal corrupted with lung sounds is shown in Figure 6 (a) The dashed vertical lines correspond to the boundaries (Ω_i). Similarly spectrum segmentation for PCG signal corrupted with speech is shown in Figure 7 (a). From the figures, it can be seen that the length of each segment is adaptive and hence the filter bank constructed using scaling function (ϕ_1) and empirical wavelet function (ψ_n) are also adaptive. For reconstructing the $S1$ and $S2$ sounds from the EWT decomposed PCG signal, it is necessary to have knowledge about the frequency ranges of fundamental heart sounds (FHS), murmurs, and various interference. In this work, the average smoothed spectrum for clean PCG, PCG

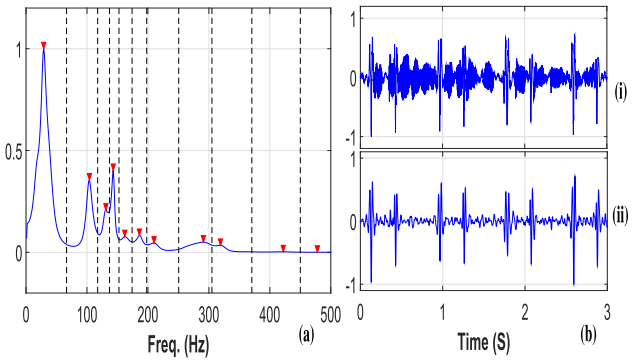

FIGURE 7. Illustrates the estimated frequency spectrum for PCG with speech and decomposition using EWT. (a) Detected boundaries on the average and smoothed spectrum of PCG corrupted with speech and (b). (i) PCG with speech. (ii) Reconstructed PCG using EWT.

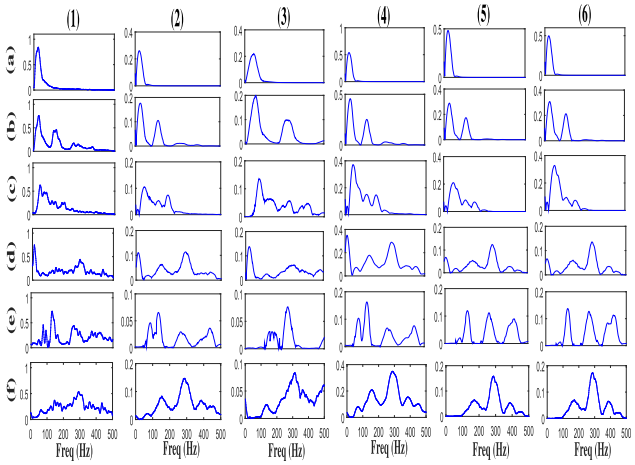

FIGURE 8. Illustrates the spectrum of heart sounds with different spectrum plotting methods such as (1) FFT, (2) PERIODOGRAM, (3) PWELCH, (4) PYULEAR, (5) PMUSIC, (6) PEIG for (a) clean PCG, (b) PCG with speech, (c) PCG with murmur, (d) PCG with lung sounds, (e) Only speech, (f) Only lung sounds.

with speech, PCG with murmurs, PCG with lung sound, is computed using different estimation techniques like periodogram, Welch, Yule-Walker, MUSIC, and Eigenvector. The spectrum analysis was carried out using both real-time signals and signals taken from the standard database and the results summarized in Figure 8.

The dominant frequency ranges of FHS, murmurs and other interference are obtained using energy-based thresholding (10% and 20%) and is reported in Table 1. From Table 1 it can be seen that FHS has frequencies around 10Hz-70Hz. It should be noted that the spectrum in Figure 6 (a) and Figure 7 (a) is the smoothed spectrum used for illustrating the adaptive filter bank. In practice, the spectrum will be non-smoothed (computed using FFT) and will have multiple peaks in the frequency range of interest (i.e 10-70Hz). Many of these peaks will correspond to interference like lung sounds, speech, and murmurs. However, reconstruction from the filter bank using peaks which has amplitude more than 50% of

TABLE 1. Estimated frequency ranges of clean PCG and PCG with different interference.

PCG Signal	Estimated freq. (Hz)		Existing freq. (Hz) study in literature
	10%	20%	
$S1$,$S2$	1-100	5-65	20-200 [37], 10-100 [38], 10-150 [24]
$S1$,$S2$,Speech	5-370	10-290	—
$S1$,$S2$,Murmurs	15-410	20-250	20-700 [37], 10-600 [24]
$S1$,$S2$,Lung sound	20-450	30-430	10-1000 [39],10-500 [40]

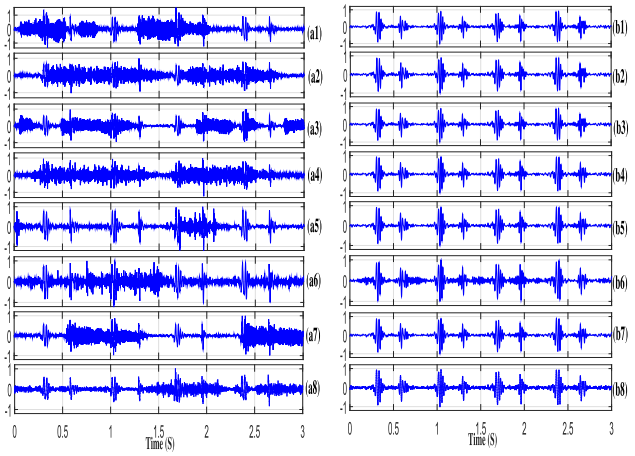

FIGURE 9. Illustrates the effective reconstruction of $S1$ and $S2$ heart sounds from PCG corrupted with different lung sounds. (a1)-(a8) PCG corrupted with different lung sounds. (b1)-(b8) EWT based reconstructed PCG with only $S1$ and $S2$ heart sounds.

the maximum is considered. This will eliminate interference whose frequency ranges overlap with the FHS. To reconstruct the FHS from EWT decomposition, only the modes whose frequencies are in the range of 10Hz-70Hz are considered. This is because the adaptive filter bank offers high resolution around the frequency of interest ($S1$ and $S2$ sounds). The set of modes whose frequency are in the range of 10Hz-70Hz is denoted as $S = (S_1, S_2, \ldots, S_p)(where\ P < N)$. Thus the reconstructed $S1$ and $S2$ sounds are given by,

$$\hat{x}(t) = IFT(I_0(\omega)\phi_1(\omega) + \sum_{i \in S} D_i(\omega)\psi_i(\omega)) \tag{8}$$

The reconstructed heart sound for PCG corrupted with lung sounds and speech is shown in Figure 6 (b) and Figure 7 (b) respectively. An example of efficiency in the reconstruction of FHS from PCG corrupted with different lung sounds using EWT is shown in Figure 12.

2) DETECTION OF DELINEATION PARAMETERS

The reconstructed $S1$ and $S2$ heart sound signal $\hat{P}_{S1,S2}[n]$ is subjected to a non-linear amplitude transformation to emphasize the informative amplitude content present in the signal. In this work, Shannon entropy is considered for non-linear transformation. Shannon entropy is chosen because

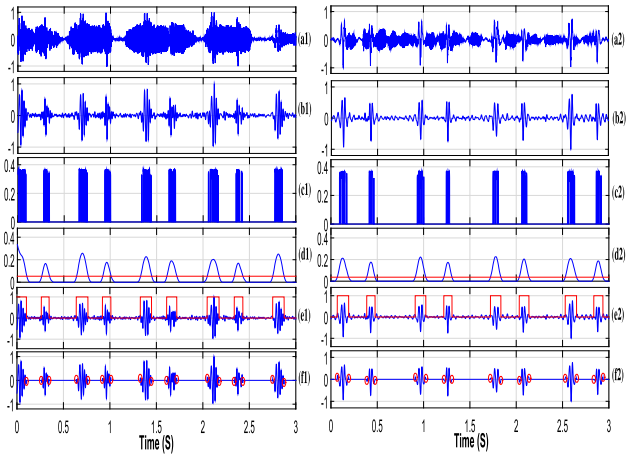

FIGURE 10. Illustrates the detection of $S1$ and $S2$ sounds from PCG with interference of lung sound, and speech using EWT decomposition method. (a1)-(a2) PCG corrupted with lung sound, and speech. (b1)-(b2) The reconstructed PCG using EWT for PCG with lung sound, and speech. (c1)-(c2) Shannon entropy envelograms. (d1)-(d2) Smoothed Shannon entropy envelograms. (e1)-(e2) Gating envelopes over FHS. (f1)-(f2) Fundamental heart sound detection using EWT for PCG corrupted with lung sound, and speech.

it enhances the informative low amplitude segments of the heart sound and is shown in Figure 10 (c1)-(c2) for PCG corrupted with lung sound, and speech respectively. Since this feature will also enhance the low amplitude noise, $\hat{P}_{S1,S2}[n]$ is subjected to a fixed threshold to suppress the noise. The threshold signal $\hat{P}_{th}[n]$ is given as,

$$\hat{P}_{th}[n] = \begin{cases} \hat{P}_{S1,S2}[n], & if\, \hat{P}_{S1,S2}[n] > \gamma_{th} \\ 0, & otherwise \end{cases} \tag{9}$$

The value of γ_{th} is chosen as 0.1 by considering the $S1$ and $S2$ amplitude levels. The Shannon entropy envelope (SEE) is computed as

$$P_{Sh}[n] = -[|\hat{P}_{th}[n]|].\log(|\hat{P}_{th}[n]|) \tag{10}$$

The smoothen Shannon entropy envelope (SSEE) $\hat{P}_{Sh}[n]$ is obtained by smoothening $P_{Sh}[n]$ using a zero phase forward and reverse filter (for filtering, a rectangular window of length 50 ms with an overlap of 1 ms is used) and is shown in Figure 10 (d1)-(d2) for PCG corrupted with lung sound, and speech respectively. Then the gated signal is computed as follows:

$$\hat{P}_g[n] = \begin{cases} 1, & if\, \hat{P}_{Sh}[n] > \gamma_{sh} \\ 0, & otherwise \end{cases} \tag{11}$$

where γ_{sh} is chosen as the mean value of $\hat{P}_{Sh}[n]$.

The gated signal computed from the SSEE is shown in Figure 10 (e1)-(e2) for PCG corrupted with lung sound, and speech respectively. In order to emphasize the large slope between consecutive points of the gating signal, it is subjected to a first-order derivative filter. The filtered signal is given by,

$$\hat{P}_{der}[n] = \hat{P}_g[n] - \hat{P}_g[n-1] \tag{12}$$

The filtered signal $\hat{P}_{der}[n]$ consists of alternative positive and negative impulses. Now the time instants of these impulses are projected onto the resultant PCG signal. The resultant PCG signal is obtained by multiplying the $\hat{P}_{S1,S2}[n]$ with the gated signal $\hat{P}_g[n]$. The projected time instants (shown in red circles) are shown in Figure 10 (f1)-(f2) for PCG corrupted with lung sounds and speech respectively. These time instants are known as delineation parameters of fundamental heart sound segments. To recognize the FHSS (whether the segments belong to $S1$, systole pause, $S2$, and diastole pause), U-Net based DNN is used and presented in the next subsection.

3) RECOGNITION OF FHSS USING U-NET BASED DNN

In this work 1D variant of U-Net based DNN is used with the motivation from [10]. Traditional U-Net based DNN is modified by using ResNet blocks and batch normalization as discussed earlier. From the state of the art, it is observed that the auto-correlation envelope, Hilbert envelope, homomorphic envelogram, and power spectral density (PSD) envelope are effective in localizing the segments of the heart sounds. Hence from the EWT based reconstructed signal these four envelograms are computed and used as four channels for training the U-Net. An input matrix 'F' of batch size 64 with 4 channels is created and applied to train the U-Net based DNN. As shown in Figure 3 various convolutional layers are used with filters of different dimensions. For the convolution process, a stride (τ) length of 8 is chosen from the state of the art and the input of the convolutional layers zero-padded to maintain output as the same size that of input. To update the filters, categorical cross-entropy is used as the loss function. Adam optimizer with differential learning rate is used for training the model. The upper and lower bounds used in differential learning rates are calculated using learning rate scheduler. As shown in Figure 3 and discussed earlier, U-Net based DNN constructed by mass blocks, max pooling, bottleneck layer, skip connections, and upsampling layers is effective in representation of the input EWT based reconstructed PCG signal. To enhance the recognition accuracy rate of FHSS, temporal modeling has been performed on the output of the U-Net with delineation parameters obtained from the SEE technique. As shown in Figure 3 the output sequences (shown in the block as $\{0, 1, 2, 3\}$) are the various states of the PCG signal which represents $S1$, systole pause (SP), $S2$, and diastole pause (DP).

IV. RESULTS AND DISCUSSION

The performance analysis of the proposed method is carried out on MATLAB 2014b, and Google colaboratory's open source platform (K80 GPU, 12GB RAM). MATLAB used for features extraction and Python (using PyTorch library) used for modeling U-Net based DNN to recognize FHSS.

A. DATABASE

To the best of our knowledge, there is no particular database available for PCG corrupted with lung sounds, and speech. Hence for PCG corrupted with lung sounds database, Littmann's lung sound library is used [41] and synthetically

added to Physionet dataset [42]. For creating the database of real-time PCG, 74 voluntarily participated male adults of age group ranging between $17-38$ years old subjects were recorded. An in-house developed PCG acquisition system is used for real-time PCG recordings with and without speech. The recordings are collected from subject in sitting position with 30 seconds of duration. For recording real-time PCG with speech, subjects were asked to speak a few words while recording their PCG. Also, their speech is simultaneously recorded using audacity software and synthetically added with Physionet database. Hence the list of PCG databases used for several experiments are the Physionet (PH), PH with lung sounds (PH+LS), PH with speech (PH+S), real-time (RT), RT with lung sounds (RT+LS), and RT with speech (RT+S).

To analyze various aspects of the proposed method, three experiments are considered. In the first experiment, the rationale for choosing the EWT for decomposition is explained in terms of quality parameters and computational complexity. In the second experiment, the robustness of the proposed EWT based method in segmenting $S1$ and $S2$ sounds from PCG corrupted with lung sounds, and speech is reported. In the third experiment, the performance of the proposed method for recognition of FHSS is presented.

1) EXPERIMENT I

The quality parameters such as root mean square error (RMSE), maximum absolute error (ME), and signal to noise ratio (SNR) of EWT based reconstructed PCG from PCG interfere with lung sounds, and speech is obtained by,

$$RMSE = \sum_{n=1}^{M}(\sqrt{(P_c[n] - \hat{P}_{S1,S2}[n])^2}) \tag{13}$$

$$ME = max_{n=1}^{M}(|P_c[n] - \hat{P}_{S1,S2}[n]|) \tag{14}$$

$$SNR = 10log_{10}(\frac{\sum_{n=1}^{M}(P_c[n] - \mu_o)^2}{\sum_{n=1}^{M}(P_c[n] - \hat{P}_{S1,S2}[n])^2}) \tag{15}$$

where μ_o is the mean of clean PCG $P_c[n]$, and M indicates the number of subjects.

The obtained quality parameters of the proposed method are compared with the EEMD and SSA methods which are reported in Table 2. From Table 2, it can be observed that the quality parameters of EWT based reconstructed PCG is better than the other methods. The rationale behind the enhancement of $S1$ and $S2$ sound in PCG corrupted with lung sounds, and speech is due to the inherent adaptive filtering nature of the EWT. As adaptive filter banks act as band pass filters and boundaries of segments are determined with the local information, high-frequency resolution can be achieved. Hence the reconstruction of PCG with FHSS using EWT results in high SNR, low RMSE, and low ME.

To find the computational complexity of different decomposition methods, MATLAB simulations are conducted on Intel(R) Core (TM) i5 3210M CPU @ 2.50 GHz, 4GB RAM computer. The computational complexity is obtained by averaging the 100 execution processing times of the decomposition of PCG corrupted with lung sound, and speech. The computational complexity of the proposed method is compared with the other methods like EEMD, and SSA and the same is reported in Table 3. From Table 3 it is observed that the proposed EWT based method is considerably less complex than the methods like SSA, and EEMD.

TABLE 2. Quality parameters of various methods.

Interference	EWT (Proposed)			SSA			EEMD		
	RMSE	ME	SNR	RMSE	ME	SNR	RMSE	ME	SNR
Lung sounds	0.07	0.59	17.89	0.14	0.79	10.93	0.27	0.95	2.84
Speech	0.07	0.57	17.77	0.15	0.80	10.79	0.28	0.98	1.82

TABLE 3. Computational complexity of decomposition methods.

Method	Run-time (Duration of 5 Sec.) (Avg. of 100 execution times) ($\mu \pm \sigma$)(Sec.)
EWT	$\mathbf{0.20 \pm 0.09}$
SSA	0.37 ± 0.18
EEMD	1.86 ± 3.81

2) EXPERIMENT II

The effectiveness of the proposed EWT based method in segmentation of $S1$ and $S2$ heart sounds from the PCG corrupted with lung sounds, and speech is demonstrated using the benchmark performance metrics like sensitivity (Se), positive predictivity (P_p), and overall accuracy (OA). The performance metrics are computed by,

$$Se = TP/(TP + FN) \times 100\% \tag{16}$$

$$P_p = TP/(TP + FP) \times 100\% \tag{17}$$

$$OA = TP/(TP + FP + FN) \times 100\% \tag{18}$$

where 'TP' indicates the true positive, 'FP' indicates the false positive and 'FN' indicates the false negative.

The obtained performance metrics of the proposed EWT based $S1$ and $S2$ heart sound from PCG corrupted with different interference is reported in Table 4. From Table 4 it is observed that the 'OA' for segmentation of realtime recordings are slightly higher than the Physionet database. This is because the Physionet database consist of normal, abnormal, and a few noisy PCG signals whereas realtime recordings consists of normal PCG. The proposed EWT based method is compared with the existing methods in detection of $S1$ and $S2$ sounds like SSA and EEMD. The proposed EWT based method achieves an average 'Se' of 99.36%, average 'P_p' of 99.16%, and an average 'OA' of 98.53% in the detection of $S1$ and $S2$ heart sounds from PCG corrupted with lung sound, and speech. The performance metrics of the proposed EWT based method are significantly improved than the methods like SSA, and EEMD.

TABLE 4. Performance metrics of the fundamental heart sound detection methods.

Method	Data	Seg.	TP	FP	FN	Se (%)	P_p (%)	OA (%)
EWT +SSEE +Th	PH	4000	3982	27	18	99.55	99.32	**98.88**
	PH+LS	4000	3973	38	27	99.32	99.05	**98.39**
	PH+S	4000	3968	51	32	99.20	98.73	**97.95**
	RT	4000	3990	12	10	99.75	99.70	**99.45**
	RT+LS	4000	3982	25	18	99.55	99.37	**98.93**
	RT+S	4000	3978	37	22	99.45	99.07	**98.53**
SSA +SSEE +Th	PH	4000	3962	59	38	99.05	98.53	97.61
	PH+LS	4000	3955	78	45	98.87	98.06	96.98
	PH+S	4000	3942	88	58	98.55	97.81	96.42
	RT	4000	3972	49	28	99.30	98.78	98.09
	RT+LS	4000	3967	67	33	99.17	98.33	97.54
	RT+S	4000	3958	77	42	98.95	98.09	97.08
EEMD +SSEE +Th	PH	4000	3907	81	93	97.67	97.96	95.73
	PH+LS	4000	3898	108	102	97.45	97.30	94.88
	PH+S	4000	3816	397	184	95.40	90.57	86.78
	RT	4000	3912	49	88	97.80	98.76	96.61
	RT+LS	4000	3887	87	113	97.17	97.81	95.10
	RT+S	4000	3853	287	147	96.32	93.06	89.87

B. EXPERIMENT III: RECOGNITION OF FHSS

Performance of the proposed U-Net based DNN with EWT method for recognition of FHSS from PCG corrupted with lung sounds and speech is analyzed and compared with long short term memory (LSTM), and gated recurrent unit (GRU) methods. To conduct the experiments, 792 subjects of Physionet database, 72 real-time recorded normal heart sounds with and without speech, 72 real-time recorded speech signals, 16 Littmann's lung sounds are considered. 72 speech signals and 16 lung sounds are for the synthetical addition to create a database of PCG with speech and PCG with lung sounds. For the recognition of FHSS using Physionet (PH), 80% of the Physionet database are used for the training and 20% of the Physionet database are used for the classification. For the recognition of FHSS when PCG interfere with lung sounds, a database is created in such a way that out of the 792 heart sounds of Physionet, 350 heart sounds are picked randomly and synthetically added with 16 different lung sounds. Remaining 342 heart sounds have remained the same. The generated database of Physionet with the interference of lung sounds are also included with the real-time recorded database of 72 normal heart sounds (without speech) synthetically added with 16 lung sounds. This new database is termed as 'PHN+LS'. For the recognition of FHSS when PCG interfere with lung sounds, 80% of the PHN+LS is used for the training and 20% of the PHN+LS is used for the testing. Similarly, for the recognition of FHSS when PCG interfere with speech, a database is created in such a way that out of the 792 heart sounds of Physionet, 350 heart sounds are picked randomly and synthetically added with 72 real-time recorded speech signals. Remaining 342 heart sounds have remained the same. The generated database of Physionet with interference of speech is also included with the real-time recorded database of 72 normal heart sounds with speech. This new database is termed as 'PHN+S'. For the recognition of FHSS when PCG interfere with speech, 80% of the PHN+S is used for the training and 20% of the PHN+S is used for the testing. For the cross-validation, 792 subjects of Physionet are synthetically added with 16 lung sounds and 72 speech signals are considered for training and the trained network is tested with 72 subjects of normal heart sounds recorded with speech are synthetically added with 16 different lung sounds. The database is termed as 'RT+LS+S'. For training the U-net, data is trained from zero learning, 10 fold cross-validation is performed, and a window length of 64 is considered. Performance metrics are computed using (16)-(18) where true positives are the estimation of $S1$ (or $S2$ or SP or DP) is the same as that of ground truth sequence of $S1$ (or $S2$ or SP or DP), all other estimations are false negatives, and false positives are the estimation of noisy segments as $S1$ (or $S2$ or SP or DP). The performance metrics are presented in Table 5. From Table 5 it can be observed that the performance of the U-Net, LSTM, and GRU methods for the classification of FHSS is degraded. To depict the rationale behind the degradation in classification accuracy, the effects of interference on four envelogram are shown in Figure 11. As shown in Figure 11, the features will not train the network accurately for the classification of FHSS. Hence, in the proposed method EWT is utilized for the removal of interference and then the features obtained from the EWT based reconstructed signal are feed for the training. From the table 5. it can be observed that the proposed U-Net based DNN with EWT method achieves significantly better recognition of FHSS than U-Net based DNN without EWT. Also, the proposed method outperforms other methods such as LSTM and GRU. The rationale behind the improvement in the performance of the proposed method is that the features do not get affected with noisy PCG as it is processed through

TABLE 5. Performance metrics of the FHSS recognition.

Method	Data	Without EWT			With EWT		
		Se	+ P	**OA**	Se	+ P	**OA**
U-Net	PHN	91.29	92.74	**90.27**	92.82	94.81	**92.82**
	PHN+LS	88.21	90.41	**87.11**	91.72	93.86	**91.17**
	PHN+S	87.92	89.94	**86.87**	89.21	92.34	**90.78**
	RT+LS+S	88.54	90.23	**87.91**	90.87	93.14	**91.86**
LSTM	PHN	89.63	90.52	**88.14**	91.32	92.87	**90.03**
	PHN+LS	87.72	88.23	**86.17**	89.91	90.47	**89.23**
	PHN+S	86.92	87.95	**85.05**	89.24	90.81	**87.96**
	RT+LS+S	87.23	88.31	**86.55**	90.24	91.05	**88.32**
GRU	PHN	88.64	89.27	**84.21**	90.97	92.14	**89.73**
	PHN+LS	83.12	85.61	**82.23**	89.80	91.24	**87.83**
	PHN+S	82.79	84.93	**81.91**	88.61	90.12	**86.95**
	RT+LS+S	83.85	85.42	**82.57**	89.78	91.34	**88.09**

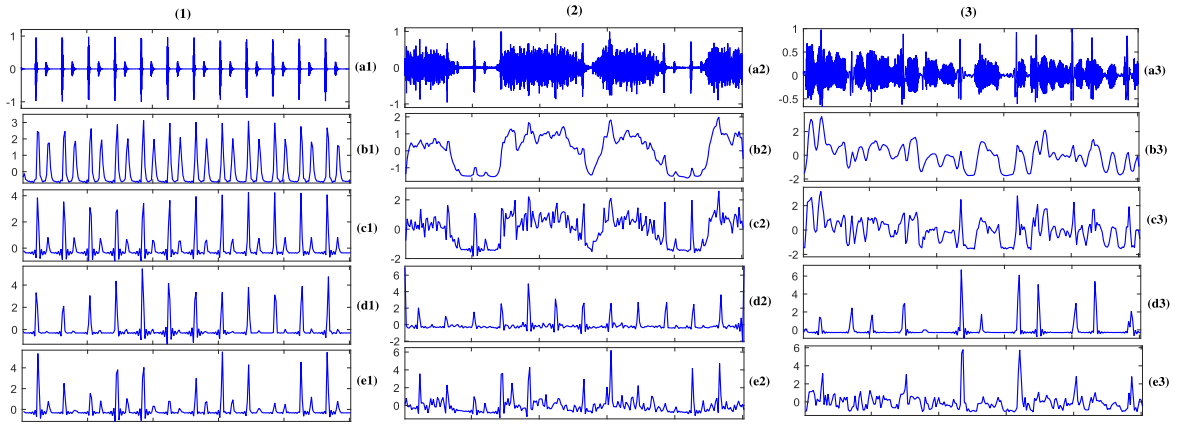

FIGURE 11. Illustrates the affects of lung sounds and speech on envelograms of PCG (5 Seconds of duration) (a1) PCG (a2) PCG with lung sounds (a3) PCG with speech. (b1)-(b3) Homomarphic envelograms of (a1)-(a3). (c1)-(c3) Hilbert envelograms of (a1)-(a3). (d1)-(d3) PSD envelograms of (a1)-(a3). (e1)-(e3) Wavelet envelograms of (a1)-(a3).

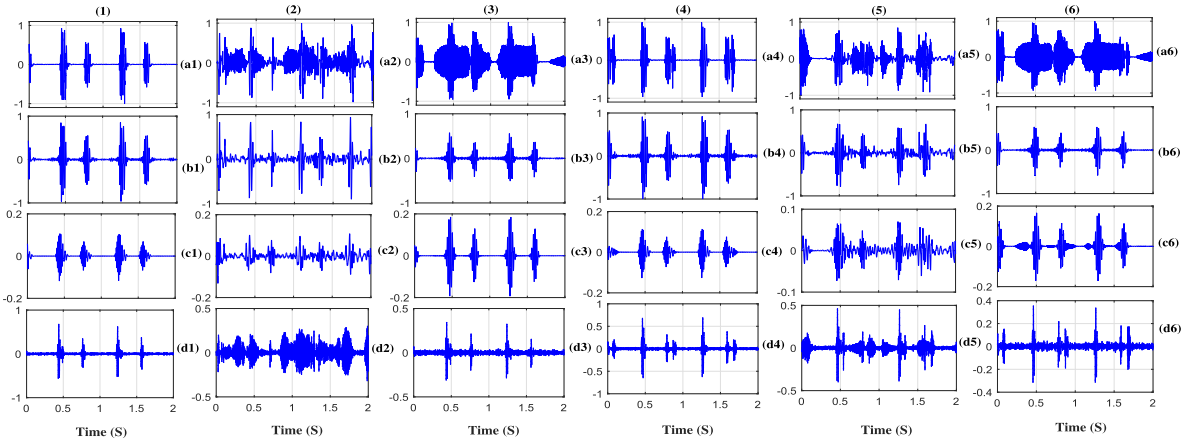

FIGURE 12. Illustrates the various reconstructed signals of normal PCG and abnormal PCG corrupted with lung sounds and speech. (a1) Normal PCG. (a2) Normal PCG corrupted with speech. (a3) Normal PCG corrupted with lung sound. (a4) Abnormal PCG. (a5) Abnormal PCG corrupted with speech. (a6) Abnormal PCG corrupted with lung sound. (b1)-(b6) EWT based reconstruction of (a1)-(a6). (c1)-(c6) SSA based reconstruction of (a1)-(a6). (d1)-(d6) EEMD based reconstruction of (a1)-(a6).

effective reconstruction using EWT and also due to the usage of delineation parameters for temporal modeling.

C. MERITS AND LIMITATIONS

The proposed EWT based DNN for classification of FHSS has significant merits of effective reconstruction and classification from the PCG corrupted with lung sounds, and speech. To demonstrate the effectiveness in reconstruction using EWT, six different cases are considered and are shown in Figure 12. As shown in Figure 12, EWT is effective in the reconstruction of FHS for all the different cases. As shown in Figure 12, reconstruction of FHS using singular spectrum analysis (SSA) and ensemble empirical mode decomposition (EEMD) are not effective for the PCG with the interference of lung sounds and speech.

The best example for the benefit of contributed work in medical practices are with the pandemic COVID 19. Several articles on COVID 19 are reporting that the virus is mysteriously affecting the lungs. Hence, in this case, we can expect that the interference of lung sounds as well as coughing sounds from the subject while examining the heart sounds using a stethoscope. Hence, if there is an automated system in digital stethoscope which can eliminate non-cardiac events and recognize the fundamental heart sounds, it will much beneficial for the medical practitioner to assess the cardiac condition of the patient. The technical significance of the proposed method is depicted in Figure 11. The recent advancement of automated digital stethoscopes are becoming powerful in assessing the heart condition with Artificial Intelligence methods. If the interference of the kind overlap with that of fundamental heart sounds are not removed, then there will be a reduction in classification accuracy of heart segments and may lead to more number of false alarms which result in the wrong identification of systolic and diastolic parts of the PCG. In smart pacemakers, it is required to give the timing information of systolic and diastolic periods to generate the electrical signal if an abnormality in the functionality of heart. Hence, the proposed method is useful to medical practitioners

in the automatic identification of systolic and diastolic activities of the heart.

The proposed method combines conventional signal processing method with artificial intelligence (AI) based technique. There is a significant improvement in performance compared to the usage of only AI-based techniques. However, there is an increase in computational complexity.

V. CONCLUSION

In this work, U-Net based DNN with EWT for recognition of fundamental heart sound segments from PCG corrupted with lung sounds, and speech is proposed. In the proposed method, the corrupted PCG signal is decomposed using adaptive filter banks of EWT. The estimated frequency range of fundamental heart sounds is incorporated into EWT for the effective reconstruction of the fundamental heart sounds. Delineation parameters of FHSS are obtained by using Shannon entropy. It is observed that the EWT based method offers better performance in the segmentation of FHS when compared with existing decomposition methods like EEMD and filtering based techniques like SSA. Four different envelogram features are extracted from the EWT based reconstructed signal. These features are used for training the U-Net model for recognition of FHSS. In the part of the work a new database of PCG with lung sounds, and real-time PCG with speech is created. For recording realtime PCG an in-house developed acquisition system is used. The proposed method achieves 91.17% and 90.78% for recognition of FHSS from PCG corrupted with lung sounds, and speech respectively. The proposed method is compared with U-Net based DNN without EWT, LSTM with and without EWT, and GRU with and without EWT. The results demonstrate that there is an average improvement of 3.83% accuracy in recognition of FHSS with the combination of conventional EWT and deep neural networks.

REFERENCES

[1] *Heart Anatomy*. Accessed: Jan. 4, 2019. [Online]. Available: https://medschool.cuanschutz.edu/surgery/specialties/cardiothoracic/patient-care/heart-valve-repair-replacement

[2] *Heart, Lungs Locations*. Accessed: Jan. 4, 2019. [Online]. Available: https://stanfordhealthcare.org/medical-treatments/l/lung-transplant.html

[3] R. M. Rangayyan, *Biomedical Signal Analysis*, vol. 33. Hoboken, NJ, USA: Wiley, 2015.

[4] M. Elhilali and J. E. West, "The stethoscope gets smart: Engineers from Johns Hopkins are giving the humble stethoscope an AI upgrade," *IEEE Spectr.*, vol. 56, no. 2, pp. 36–41, Feb. 2019.

[5] D. Kouzoudis and C. A. Grimes, "The frequency response of magnetoelastic sensors to stress and atmospheric pressure," *Smart Mater. Struct.*, vol. 9, no. 6, pp. 885–889, Dec. 2000.

[6] W. Hernandez, J. De Vicente, O. Sergiyenko, and E. Fernández, "Improving the response of accelerometers for automotive applications by using LMS adaptive filters," *Sensors*, vol. 10, no. 1, pp. 313–329, Dec. 2009.

[7] Y.-N. Jeng, T.-M. Yang, and S.-Y. Lee, "Response identification in the extremely low frequency region of an electret condenser microphone," *Sensors*, vol. 11, no. 1, pp. 623–637, Jan. 2011.

[8] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds—A systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.

[9] T.-E. Chen, S.-I. Yang, L.-T. Ho, K.-H. Tsai, Y.-H. Chen, Y.-F. Chang, Y.-H. Lai, S.-S. Wang, Y. Tsao, and C.-C. Wu, "S_1 and S_2 heart sound recognition using deep neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 372–380, Feb. 2017.

[10] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2435–2445, Nov. 2019.

[11] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation—An event detection approach using deep recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, 2018.

[12] M. Mishra, H. Menon, and A. Mukherjee, "Characterization of S_1 and S_2 heart sounds using stacked autoencoder and convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3211–3220, Sep. 2019.

[13] A. Meintjes, A. Lowe, and M. Legget, "Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks," in *Proc. EMBC*, Jul. 2018, pp. 409–412.

[14] F. Noman, S.-H. Salleh, C.-M. Ting, S. B. Samdin, H. Ombao, and H. Hussain, "A Markov-switching model approach to heart sound segmentation and classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 705–716, Mar. 2020.

[15] J. Oliveira, F. Renna, T. Mantadelis, and M. Coimbra, "Adaptive sojourn time HSMM for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 642–649, Mar. 2019.

[16] M. Mishra, S. Pratiher, H. Menon, and A. Mukherjee, "Identification of S_1 and S_2 heart sounds using spectral and convex hull features," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4311–4320, Apr. 2020.

[17] A. Haghighi-Mood and J. N. Torry, "A sub-band energy tracking algorithm for heart sound segmentation," in *Proc. Comput. Cardiol.*, 1995, pp. 501–504.

[18] S. M. Debbal and F. Bereksi-Reguig, "Computerized heart sounds analysis," *Comput. Biol. Med.*, vol. 38, no. 2, pp. 263–280, Feb. 2008.

[19] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. A. Ramaiah, "Multi-level basis selection of wavelet packet decomposition tree for heart sound classification," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1407–1414, Oct. 2013.

[20] V. Nivitha Varghees, K. I. Ramachandran, and K. P. Soman, "Wavelet-based fundamental heart sound recognition method using morphological and interval features," *Healthcare Technol. Lett.*, vol. 5, no. 3, pp. 81–87, Jun. 2018.

[21] V. Nivitha Varghees and K. I. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3861–3872, Jun. 2017.

[22] K. A. Babu, B. Ramkumar, and M. S. Manikandan, "Empirical wavelet transform based lung sound removal from phonocardiogram signal for heart sound segmentation," in *Proc. IEEE ICASSP*, May 2019, pp. 1313–1317.

[23] S. Ari and G. Saha, "Classification of heart sounds using empirical mode decomposition based features," *Int. J. Med. Eng. Inform.*, vol. 1, no. 1, pp. 91–108, 2008.

[24] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1138–1152, Jul. 2014.

[25] S. Banerjee, M. Mishra, and A. Mukherjee, "Segmentation and detection of first and second heart sounds (S_1 and S_2) using variational mode decomposition," in *Proc. IECBES*, Dec. 2016, pp. 565–570.

[26] K. A. Babu, B. Ramkumar, and M. S. Manikandan, "Automatic identification of S_1 and S_2 heart sounds using simultaneous PCG and PPG recordings," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9430–9440, Nov. 2018.

[27] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3360–3367, Dec. 2011.

[28] T. Tsalaile and S. Sanei, "Separation of heart sound signal from lung sound signal by adaptive line enhancement," in *Proc. IEEE EUSIPCO*, Sep. 2007, pp. 1231–1235.

[29] A. Yadollahi and Z. M. K. Moussavi, "A robust method for heart sounds localization using lung sounds entropy," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 497–502, Mar. 2006.

[30] I. Y. Ozbek and H. Shamsi, "Heart sound localization in respiratory sound based on a new computationally efficient entropy bound," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 105–114, Jan. 2017.

[31] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*. Springer, 2015, pp. 234–241. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-319-24574-4_28

[33] W. Liu, S. Cao, and Y. Chen, "Seismic time–frequency analysis via empirical wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 28–32, Jan. 2016.
[34] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.
[35] K. Thirumala, A. C. Umarikar, and T. Jain, "Estimation of single-phase and three-phase power-quality indices using empirical wavelet transform," *IEEE Trans. Power Del.*, vol. 30, no. 1, pp. 445–454, Feb. 2015.
[36] U. Baid, S. Talbar, S. Rane, S. Gupta, M. H. Thakur, A. Moiyadi, S. Thakur, and A. Mahajan, "Deep learning radiomics algorithm for gliomas (drag) model: A novel approach using 3D U-Net based deep convolutional neural network for predicting survival in gliomas," in *Proc. MICCAI*. Springer, 2018, pp. 369–379. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-11726-9_33
[37] H. Naseri and M. R. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, Feb. 2013.
[38] P. Samanta, A. Pathak, K. Mandana, and G. Saha, "Classification of coronary artery diseased and normal subjects using multi-channel phonocardiogram signal," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 426–443, Apr. 2019.
[39] M. T. Pourazad, Z. Moussavi, F. Farahmand, and R. K. Ward, "Heart sounds separation from lung sounds using independent component analysis," in *Proc. IEEE EMBC*, Jan. 2005, pp. 2736–2739.
[40] G. Shah, P. Koch, and C. B. Papadias, "On the blind recovery of cardiac and respiratory sounds," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 151–157, Jan. 2015.
[41] *Littmann Lung Sounds Database*. Accessed: Jan. 9, 2018. [Online]. Available: http://www.3m.com/healthcare/littmann/mmm-library.html
[42] *Physionet Database*. Accessed: Jun. 15, 2016. [Online]. Available: https://www.physionet.org/content/challenge-2016/1.0.0/

K. AJAY BABU (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Kakinada, India, in 2009, and the M.Tech. degree in electronics and communication engineering from IIT Bhubaneswar, Odisha, India, in 2014, where he is currently pursuing the Ph.D. degree with the School of Electrical Sciences. His current research interests include biomedical signal processing, wearable healthcare monitoring, machine learning, the Internet of Things, and wireless communication.

BARATHRAM RAMKUMAR (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Idaho State University, Pocatello, ID, USA, in 2007, and Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2011. He is currently working as an Associate Professor with IIT Bhubaneswar. He has been the reviewer of a number of prestigious conferences and journals. His research interests include signal processing, cognitive radios, and wireless communication.