IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2022, accepted March 1, 2022, date of publication March 8, 2022, date of current version March 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157305

Automated Heart Valve Disorder Detection Based on PDF Modeling of Formant Variation Pattern in PCG Signal

MONJUR MORSHED[1], (Graduate Student Member, IEEE), SHAIKH ANOWARUL FATTAH[1], (Senior Member, IEEE), AND MOHAMMAD SAQUIB[2], (Senior Member, IEEE)

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[2]Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA

Corresponding author: Shaikh Anowarul Fattah (fattah@eee.buet.ac.bd)

ABSTRACT Heart valve disorder (HVD) analysis from heart sound is being well known for a long period of time, and use of digital stethoscope gives opportunity to diagnose HVDs from phonocardiographic (PCG) signal. An automated HVD detection technique from PCG signal can play a key role as a first-hand diagnostic tool for the physicians. In this paper, in order to classify different HVDs, we propose to utilize the formant characteristic of the PCG signal, which is an acoustic property of the heart sound. PCG signals exhibit significant variations depending on different types of HVDs and thus conventional time frequency domain features or statistical features are extracted from PCG signal for disease classification. However, direct PCG signals are also used in sequential networks to classify HVDs. Similar to the formant peaks of voiced speech signal, the spectrum corresponding to the PCG signal exhibits distinguishable peaks, especially in the voiced part of the heart sound (lub-dub). Keeping this notable key point in consideration, Burg's autoregressive model is used to find the parametric spectrum of the PCG signal. The first two formants of the PCG signal, that carry the most informative acoustic properties of the heart sound, are estimated from the Burg's spectrum, and are used for feature extraction. The magnitude, frequency and phase of each formant are considered to evaluate these features. Instead of considering a long duration of PCG signal at a time, we consider the overlapping sub-frames, and extract formants from each sub-frame, which generates a temporal variation of the formants. Finally, we propose a PDF model fitting of the formant variation, and utilize the estimated model parameters along with some statistical features to classify the HVDs. Two famous publicly available PCG datasets are used to demonstrate the performance of the proposed method, that efficiently classify the binary/five classes of heart sounds. The results reveal that the proposed method has the overall accuracy values of 93.46% and 99.28% for the two datasets, which is better in comparison to other previously reported state-of-the-art techniques.

INDEX TERMS Classification, formant, feature extraction, heart valve disorder, model fitting, phonocadiogram.

I. INTRODUCTION

Early detection of heart valve disorder (HVD) may reduce almost one-third mortality rate of our planet due to various cardiac failures [1]. Every year, all over the world, this rate is increasing alarmingly. The treatment cost of cardiovascular diseases is predicted to be annually one trillion dollar by 2030 in the USA alone [2]. Stethoscopes are still used as a popular instrument to hear heart sounds from the chest to primarily diagnose the cardiac health. In case of a healthy person, two major heart sounds, namely S1 (lub) and S2 (dub) are clearly audible through a stethoscope. The first/fundamental heart sound (FHS), S1 (lub) occurs during the systole, because of ventricular contraction, the instant of mitral and tricuspid valves' closure [3]. The instant of the closure of the aortic and pulmonic valves generates a second FHS, S2 (dub) during the diastole. There are some very weak heart sounds, such as S3, S4, murmurs caused by turbulence

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

of blood flow in the arteries, ejection clicks (EC) during systole, opening snap (OS) during diastole, and mid-systolic clicks (MC) [4]. The assessment of various heart diseases through cardiac auscultation (typically heart sound heard by stethoscope) is one of the most primitive methods, especially to identify valve-related defects. The pitch and intensity of the heart sounds are overheard in different positions of the chest wall to make a quick assessment of normal/healthy class (N) and heart diseases like aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS) and mitral valve prolapse (MVP). Along with medical and family history, heart murmurs due to leaking of valves, clicking sounds, lung congestion etc. are heard by the physicians to primarily identify the valve related disorders [5]. Even in recent days, because of inexpensive and non-invasiveness, detection of valve related disorders from phonocardiographic (PCG) recordings of the heart sound has become a popular diagnosis method. With the advancement of digital signal processing kits, digital stethoscopes are getting popularity to acquire PCG, which helps to treat, analyze and interpret valve related disorders. PCG based automated methods are taken into account for developing intelligent stethoscopes with first hand decision making facilities [6].

Until now many attempts in many ways have been taken to classify HVDs. By the emergence of machine learning and deep neural network (DNN), many new sequential networks have been proposed with heavy memory costing [7]–[10]. An in a nutshell view of the earlier works are depicted here to have a comparison with our proposed method. Heart sound classification and segmentation methods are broadly categorized into three groups: (i) Envelope/Envelogram based methods (ii) Feature based methods and (iii) Sequential networks.

Envelope/envelogram based methods were mostly used for both segmentation and classification purposes of the PCG signals. Shannon energy envelogram [11], Shannon envelogram on wavelet decomposed signal [12], Shannon envelogram on S-transformed signal [13], envelogram from Hilbert transformed signal [14], moment waveform envelogram [15], squared energy envelogram [16] etc. are the most commonly used available methods. In all these methods, total number of recordings are less than 80 subjects with a variety of sampling rates from 8000-44100 Hz, and are recorded in various conditions with different sorts of devices. The overall accuracy (OA) reported for these PCGs are as high as 90-100% [3]. In these methods, having small datasets, the authors specially focus on FHSs with reduced noise components.

Features from amplitude and frequency [17], high frequency components [18], instantaneous phase [19], fast wavelet decomposition, autocorrelation [20], complexity-based features [21], multilevel wavelet decomposition coefficients [22], ensemble empirical mode decomposition (EEMD), kurtosis-based etc. are proposed by many authors, where 55 to 120 in house recordings are available in these papers for evaluating valve defects. The OA revealed is around 84%. By selecting non-Gaussian intrinsic mode functions from EEMD from 11 normal and 32 patients an OA of 83.05% is found [23]. For lub-dub recognition only 120 recordings are used, where 97% accuracy is found for lub and 94% is found for dub [24]. Wavelet synchrosqueezing transform (WSST) [25] and chirplet transform (CT) [26] are used to find magnitude and phase features from time-frequency matrix (TFM) of the PCGs to detect multi-class HVDs of 800 subjects of the github dataset [27]. The time-frequency matrix evaluated from WSST and CT is used to classify N, AS, MR and MS, by dropping MVP from the dataset [25], [26]. Features from mel-frequency cepstral coefficients (MFCC), discrete wavelet transform (DWT) and combination of the former two are also checked on the same dataset [28]. An OA of 91% to 98% are reported for these multiclass HVD identification. Different statistical features of tunable Q-wavelet transform (TQWT) are applied to least square-support vector machine (LS-SVM), where an OA of 94.01% is achieved [29]. The fast and adaptive multivariate empirical mode decomposition (FA-MVEMD) and TQWT are reported in [6] to decompose the PCG signal and its first derivative into frequency subbands. Shannon energy envelope is then applied to extract the characteristic envelope of the first two intrinsic mode functions (IMFs) and an OA of 98.48% is achieved after classifying PCG system dynamics for normal/abnormal classes with deterministic learning theory (DLT). Efforts are also taken to identify S1 and S2 with hidden Markov models (HMM) on 80 subjects and even on very less numbers of recordings [30], [31]. Hidden semi Markov model (HSMM), that uses logistic regression with modified Viterbi algorithm for state selection on 112 patients for segmentation of S1 and S2 under noisy condition is also implemented. Approximately, 96% average F1 score is testified so far for HSMM based segmentation [32]. In [6], a time–frequency feature based method is proposed to classify heart sounds into normal and abnormal classes. From [33], only 944 recordings are taken to extract 18 time–frequency features, and are used as inputs to a binary SVM-based classifier, achieving 86% for 10-fold cross validations.

Machine learning (ML) algorithms are widely used to detect the anomalies in the heart valves of unknown subjects utilizing features extracted from the 1-d PCG heart sounds. It has become a state-of-the-art technique for today and days to come for high fidelity detection of HVDs. By the grace of ML and ease of availability of advanced graphics processing units (GPUs) many architectures are proposed. The K-means clustering and homomorphic filtering are widely used for segmenting the cardiac cycle [34]. In addition to this, feature vectors extracted from diversified domains are applied to various machine learning tools, such as, SVM, random forest (RF), multiclass composite classifier (MCC) and k-nearest neighbor (KNN) to solve multiclass diseases [25], [26], [28]. Deep learning based networks are also used for PCG signal segmentation as well as abnormal heart sound detection [9], [35]–[37]. In [8], the raw PCG signal is down-sampled at

500 Hz and processed through Savitzky–Golay filter to suppress the high-frequency noises by data point smoothening. One-dimensional convolutional neural network (1d-CNN) and feed-forward neural network (FNN) are used to classify these smoothed data and claimed an OA of 85.65% [8]. At the same time, improvisation of convolutional and recurrent neural networks [38] are also used as the state-of-the-art tool to achieve highest level of accuracies without considering model complexity and computational cost. In [39], deep CNN are used on the preprocessed dataset [33], where heart sounds are segmented using U-net. Later on, classification of heart sounds is done through CNN and AdaBoost on 301 data. The OAs are reported as 87.30% and 76.1%, respectively. In [40], both Electrocardiogram (ECG) and PCG are used for feature extraction through long short-term memory (LSTM), and genetic algorithm (GA) is used for feature selection. In [41], the authors achieve 91.50% OA by setting a threshold (gradient-based method) to maximize prediction error, introducing a self-engineered mathematical formula for classification on [33]. For the same normal/abnormal classes, use of deep learning and ensemble learning with a Savitzky-Golay filter on the whole dataset generate an OA of 86.05% [42].

However, the deep-learning based heart disease classification methods mostly deal with the architectural modifications to get better performance without focusing on explainable characteristics for getting differences among various classes. It is very important to identify distinguishable characteristics among various heart diseases and obtain physiological relevance corresponding to those extracted characteristics. One possible way could be to analyze the acoustic properties of various sounds corresponding to different heart diseases and find out suitable features that follow the physiological aspects of these diseases. Automatic heart disease classification from PCG signal based on proper acoustic modeling of the given heart sound is still a very challenging task and rarely attempted.

In this paper, we propose a method that considers acoustic features (formants), computed by PDF model fitting along with statistical features and time series data. The detailed of our method is described in the next sections with proper justification and jurisdiction. The rest of the texts are arranged in the following manner. A concise yet clearly described step-by-step procedure of the proposed technique has been presented in Section III. In Section IV, we present and compare the results obtained from different platforms by providing a rational justification. Finally, concluding remarks are stated in Section V with future probable progress.

II. DATASETS

For classification and segmentation algorithms, reliable datasets like PhysioNet/CinC 2016 [33], and datasets developed by [27] have been used since their inception. In PhysioNet/CinC 2016, the dataset is composed of heart sound recordings, collected from different sources of the world by multiple research groups. Both healthy and diseased subjects are included in the dataset comprising children and adults. The recordings are sampled at 2 kHz, and shared in a standard .wav file format. In total 3240 heart sound recordings from 764 subjects/patients are available in six training sets. The duration of the recordings vary from 5 s to 120 s. This dataset provides the detailed descriptions of the population (764 subjects), age and sex of some subjects (out of 2766 trials, 2284 male and 482 female and the age range varies 10-88 years). In the second dataset, one thousand (1000) PCG recordings of five different classes of heart sounds: one normal or healthy class (N) and four common types of heart valve disorders, namely aortic stenosis (AS), mitral regurgitation (MR), mitral stenosis (MS) and mitral valve prolapse (MVP) are sampled at 8 kHz, and are also available in standard .wav format. For each of these five classes, there are 200 recordings. The duration of the PCG signals varies in between 1.156 s to 3.993 s, and each recording contains 3 cardiac cycles. Thus, from 1000 recordings there are in total 3000 cardiac cycles available for the purpose of analysis. A detailed description of the datasets can be found in [4] and [28], respectively.

TABLE 1. Duration and frequency of the heart sounds.

Sound	Duration (ms)	Frequency (Hz)
S1	70–150	20–150
S2	60–120	50–200
S3	40–100	50–90
S4	40–80	50–80

III. PROPOSED METHOD

Major steps involved in the proposed scheme are presented in the block diagram shown in Fig. 1. It involves three major operations: data preparation and preprocessing, feature extraction to form a feature dictionary and supervised classification. In the preprocessing stage, filtering, normalization and windowing operations are performed. The feature extraction stage is designed based on estimating the formants from the PCG signal and utilizing the temporal variation of the extracted formants. Finally, binary/five-class classification is performed using a supervised classifier. In what follows, the step-by-step procedure is described.

A. DATA PREPARATION AND PREPROCESSING

Generally, the frequencies of the heart sounds S1, S2, S3 and S4 lie within the range of 20-200 Hz [19]. The duration and range of frequencies are presented in Table 1. As a result, considering the general bandwidth of the PCG signal, all recordings are downsampled to 2000 samples/sec. Afterwards, the signals are passed through a bandpass filter (BPF) with cutoff frequencies (20, 450) Hz to have the frequency components of interest. The filtered signals are then amplitude normalized and scaled between -0.5 to $+0.5$. From a given PCG recording $s(n)$, an amplitude normalized signal $s_{\mathrm{N}}(n)$ is obtained as

$$s_N(n)|_{0\le|s_N(n)|\le 1} = \frac{s_A(n) - \min(s_A(n))}{\max[s_A(n) - \min(s_A(n))]} \quad (1)$$

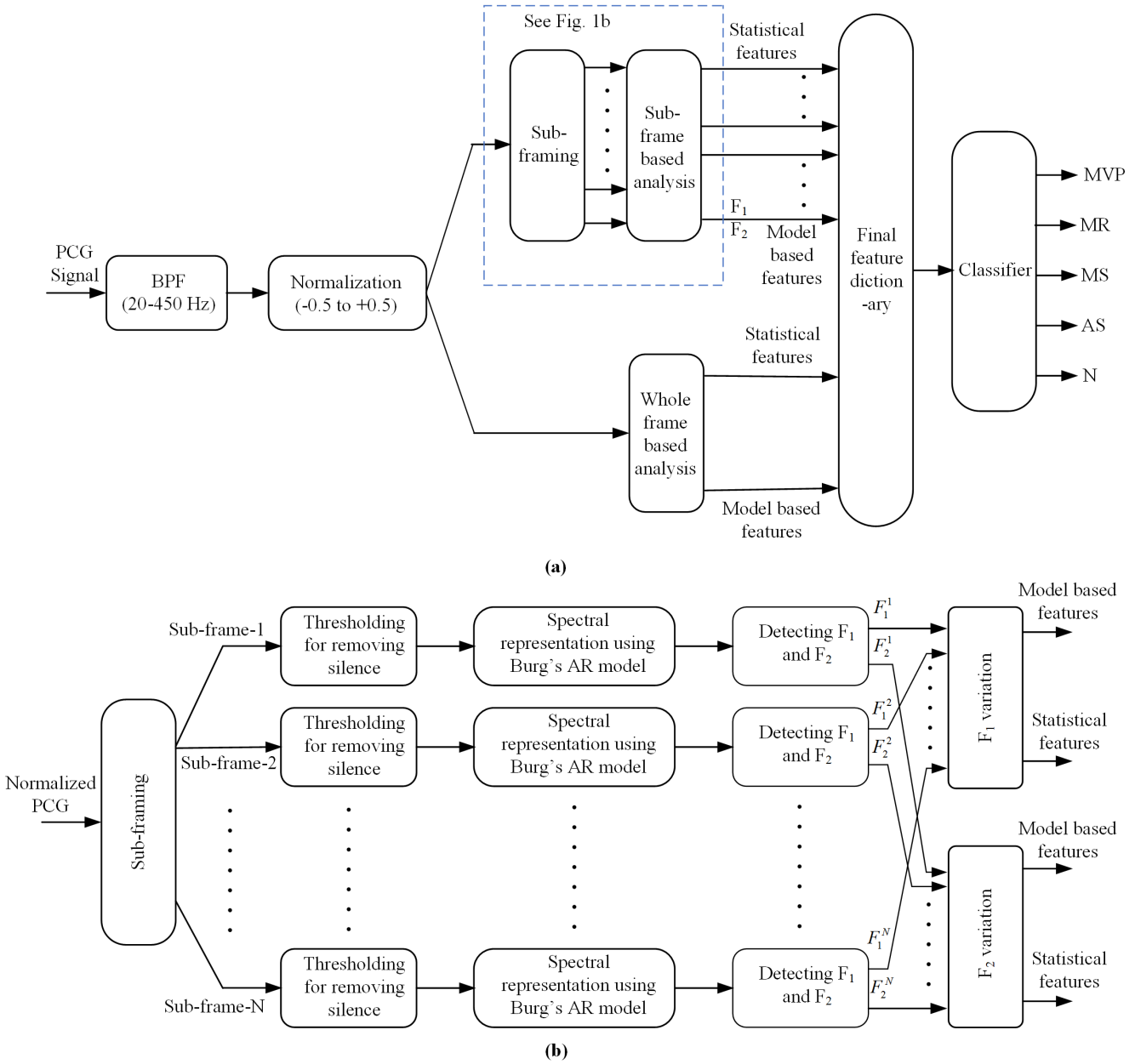

FIGURE 1. Block diagram of the proposed method. (a) Major steps involved in the proposed method to detect HVDs from PCG signal. (b) Detailed of sub-frame based analysis: model-based and statistical feature extraction from formant variations within a frame.

where, $s_A(n)$ is found via standardization i.e. $s_A(n) = (s(n) - \mu)/\sigma$, μ and σ are mean and standard deviation of $s(n)$, respectively.

B. FORMANTS OF THE PCG SIGNAL

In this paper, we propose a formant-based feature extraction method that involves formant estimation from the PCG data and modeling the temporal variations of the extracted formants (magnitude, frequency and phase). Effective spectro-temporal analysis of the PCG signals has great importance in diagnosing the valve-related disorders. The acoustic nature of the sound wave produced by the PCG signals needs special attention to explore the abnormalities. The spectral representation of a PCG signal of a healthy person generally exhibits spectral peaks similar to some speech signals (e.g. voiced speech), which inspires to utilize linear prediction model of a speech signal. In order to characterize the speech signals, fundamental and its harmonics play an important role. Specially for human voice analysis, for a long period of time, formants (fundamentals and harmonics) are playing a key role. Here the speech production system is first properly modeled and from the system's response, formants are generally estimated [43]. Heart sounds can also be treated in a similar fashion like the human voice and for

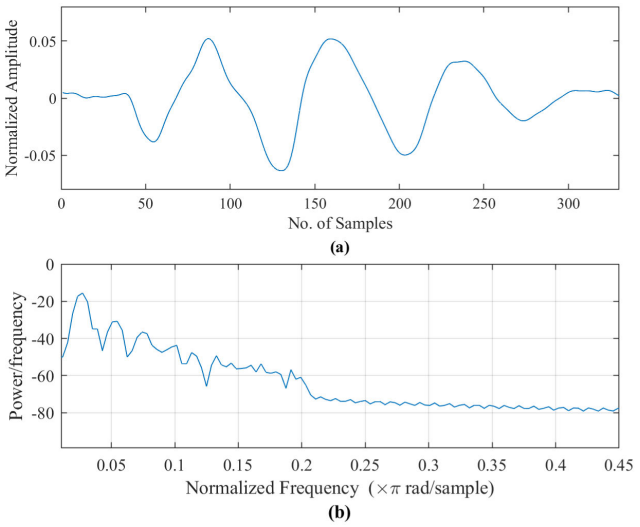

FIGURE 2. A typical PCG segment and its periodogram. (a) PCG signal representation in time domain and (b) Corresponding power spectral representation indicating formant peaks.

the estimation of fundamental and harmonic frequencies from the heart sound, linear prediction model can be used [3]. A typical PCG segment with its power spectral representation are shown in Fig. 2 for better visualization of the formants in a heart sound. Hence, similar to human voice, formant analysis of PCG signals can provide significant characteristics which may help in disease analysis. The physicians sometimes may overlook many vital audible portions during listening due to the usual limitations of human ear or in visual checkup of PCG features. However, the formant-based analysis may overcome the problem. A cardiac system entangled with PCG signal can be considered as an output of a linear time invariant (LTI) system, where the characteristics of the input to that system are unknown. Considering additive white Gaussian noise (AWGN) as an excitation to this auscultative process, this cardiac activity can be modeled with an LTI autoregressive (AR) system [3]. Moreover, the measurement of these signals requires real-time data analysis, and instantaneous evaluation of the frequency characteristics of the short time measured data are very vital. The processing of shorter data length is also effective for faster computation. In view of getting an accurate and low-energy spectral analysis of short data length, the Burg's method of parametric estimation is widely used [43]–[45]. From the estimated AR model parameters, a parametric spectrum is computed to estimate the formant peaks, and the first two formants are considered as key features to develop the feature dictionary. In addition to formant features, some conventional time-frequency features are also incorporated. Moreover, in the next subsequent section we will discuss the proposed statistical model based features extracted from the temporal variation of the formants.

Generally, in the frame by frame analysis, the investigation of a signal is carried out at a time on the full duration. In this paper, a sub-frame based approach is proposed, where a frame of PCG signal is divided into sub-frames (windows) with a reasonable amount of sample overlap amongst succeeding sub-frames. In order to minimize the effect of leakage and data loss at the edge of the signal, we have applied overlapping Hanning windows to extract sub-frames. The frame length and frame shift are chosen in the same mode as auditory recognition system [43]. For example, a signal having N samples, with a sub-frame (window) length m and shifted it by q samples, the second window will consist from $(m+1)$-th to $(m+q)$-th samples. In this way, all the succeeding windows will be traced until the last sample of the signal is reached. Afterwards, the AR model parameters of each windowed PCG signal is predicted using Burg's method [45].

C. BURG'S SPECTRUM BASED FORMANT EXTRACTION

In Burg's AR method each output sample is predicted based on the superposition of the past samples. The output signal $y(n)$ can be represented as a linear combination of preceding values of the same signal plus white noise $w(n)$ as the input.

$$y(n) = -\sum_{i=1}^{P} \alpha_P(i) y(n-i) + w(n), \tag{2}$$

where P is the model order. The forward and backward linear prediction estimates of order m are

$$\hat{y}(n) = -\sum_{k=1}^{m} \alpha_m(k) y(n-k) \tag{3}$$

$$\hat{y}(n-m) = -\sum_{k=1}^{m} \alpha_m^*(k) y(n+k-m) \tag{4}$$

The corresponding forward and backward predictor errors are

$$f_m(n) = y(n) - \hat{y}(n) \tag{5}$$

$$b_m(n) = y(n-m) - \hat{y}(n-m) \tag{6}$$

where $\alpha_{\mathrm{m}}(k)$ are prediction coefficients. It should be noted that $\alpha_{\mathrm{m}}(0) = 1$, by the definition. The lattice of the FIR filter is represented by the set of recursive equations

$$f_m(n) = f_{m-1}(n) + k_m b_{m-1}(n-1) \tag{7}$$

$$b_m(n) = k_m f_{m-1}(n-1) \tag{8}$$

where the initializations of the residuals are $f_0(n) = b_0(n) = f(n)$ and k_{m} are the reflection coefficients of the m-th recursion step.

$$k_m = \frac{-2\sum_{n=P+1}^{N} [f_m(n)]}{\sum_{n=P+1}^{N} [f_{m-1}(n)]^2 + [b_{m-1}(n-1)]^2} \tag{9}$$

$$\alpha_m(n) = \alpha_{m-1}(n) + k_m \alpha_{m-1}(k-m) \tag{10}$$

where $\alpha_{\mathrm{m}}(0) = k_{\mathrm{m}}$.

It is advantageous to use Burg's AR method because it is computationally efficient, stable and has satisfactory frequency resolution. An optimum AR model order is selected to obtain distinguishable peaks of the spectrum of each PCG sub-frame [45]. The parametric spectrum is constructed using the estimated AR parameters. It is observed that the

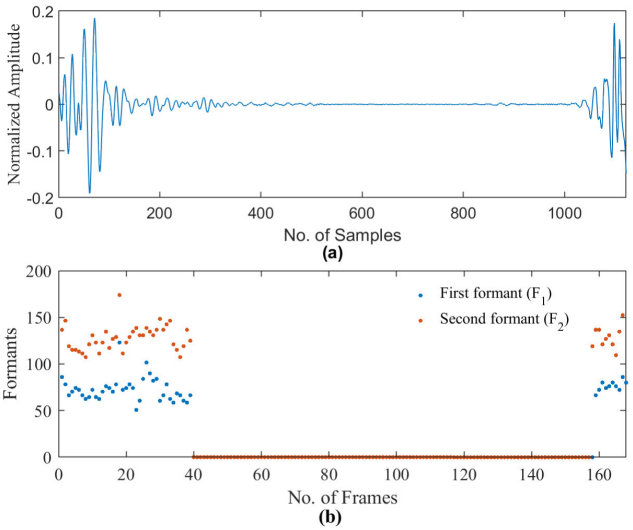

FIGURE 3. A PCG segment with first two formants after discarding the silence. (a) A portion of time series PCG data (b) Sub-frame wise variations of first formant (blue dots) and second formant (orange dots).

normalized and scaled PCG signal contains significant energy in some portions (mainly S1 (lub) and S2 (dub)), that are responsible for audible sound, and very negligible energy in some portion portraying silence. Hence, the regions in the normalized and scaled PCG signal containing negligible energy are eliminated to discard the silent portions and feature extraction is not performed in those regions. For a better understanding, a certain duration of normalized and scaled PCG signal is considered which consists both significant energy and negligible energy portions. Using the Burg's AR spectrum, the first and second formants are estimated from each sub-frames. In Fig. 3, the time domain PCG signal and corresponding formant estimations (first two formants F1 and F2) are shown. Consistent formant estimates are found in the significant energy region and formant estimation is not performed in the negligible energy region. It is further discussed in the results and discussion section. From the magnitude and phase spectrum the magnitude, frequency and phase of the 1st and 2nd formant of each frame are thus computed. In order to demonstrate the spectral peaks in the PCG, a sub-frame of a sampled PCG signal is considered. By using Burg's methods AR(12) spectrum is constructed and shown in Fig. 4. In the same figure magnitude and phase spectra are plotted using solid and dashed lines, respectively. The first two peaks indicate the 1st and 2nd formant respectively, which are our subject of interest. The magnitudes, frequencies and phases of the first two formants are thus evaluated for each of the overlapping sub-frames to develop the feature dictionary.

Next our objective is to demonstrate the temporal variation pattern of the extracted formants in a given PCG frame. For this purpose four PCG segments are considered which correspond to a healthy (N) class and three diseased classes, namely aortic stenosis (AS), mitral stenosis (MS) and mitral

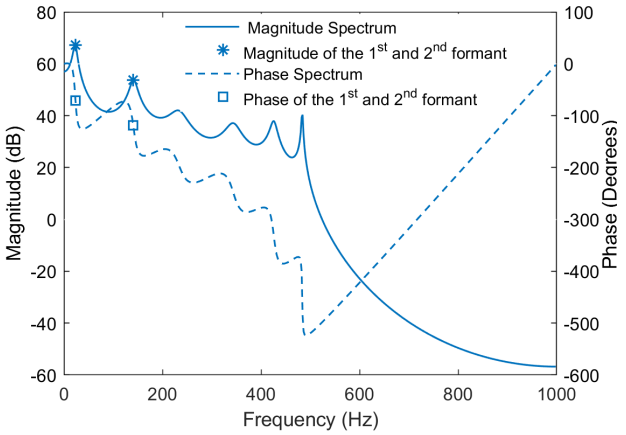

FIGURE 4. Magnitude and phase spectrum of a PCG segment obtained from AR Burg's algorithm.

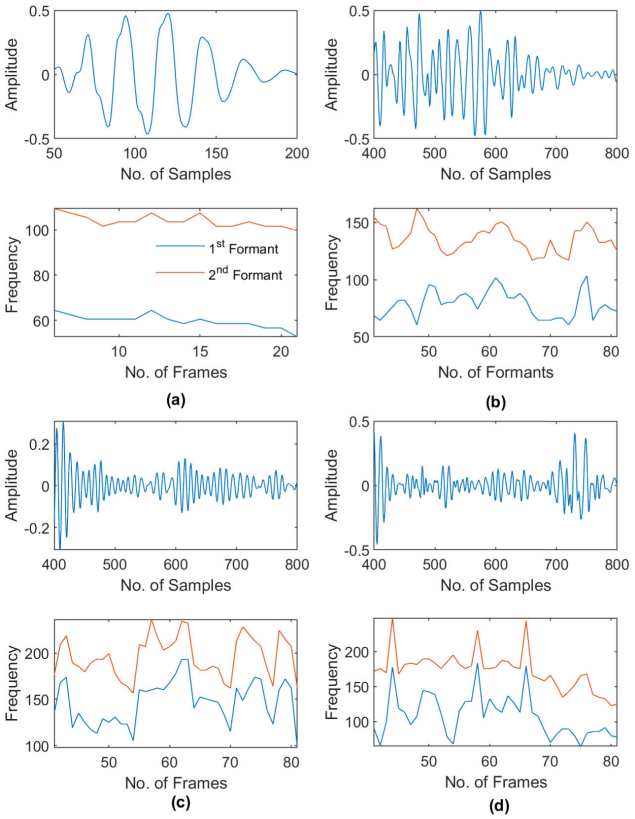

FIGURE 5. Temporal variation of first two formants along with the normalized and scaled PCG signal for healthy and diseased cases. In each case, higher value of formant frequency indicates the 2nd formant (red color) and the lower value (blue color) corresponds to 1st formant. (a) Healthy class, (b) Aortic Ste- nosis, (c) Mitral Stenosis and (d) Mitral Valve Prolapse.

valve prolapse (MVP). In Fig. 5, temporal variation of the first two formants along with the normalized and scaled PCG signals are shown for these four classes. As mentioned before, the values of the first two formant frequencies, extracted from each of the overlapping sub-frames in the region with sufficient energy, are investigated over a certain period.

In each of the four cases of Fig. 5, the upper figure of represents a portion of the normalized and scaled PCG signal and the lower figure represents the temporal variation pattern of the 1st and 2nd formant. In each pair of the figures, the x-axis of the lower figure represents sub-frame numbers corresponding to the time instances of the upper figure. It is observed from these figures that the temporal variation pattern of the formant frequencies exhibits distinguishable pattern for different cases, which also serves as a strong motivation behind utilizing this variation pattern as potential features.

D. MODELING THE TEMPORAL VARIATION OF PCG FORMANTS

In real-time and fast disease detection, this sub-frame based formant extraction may result in large numbers of features than that of conventional feature extraction methods. As an alternative, one may consider less number of sub-frames or even conventional frame based method. However, in that case temporal resolution for the features will be drastically reduced and there will be feature averaging over a certain time duration. Hence, we prefer to use the sub-frame based feature extraction. In this case, one may feed the classifier all the sub-frame based extracted features which will increase computational burden and time. In the proposed method, instead of directly feeding the sub-frame extracted formants to the classifier, we propose to obtain a suitable statistical distribution of the formant variation pattern, and utilize the distribution parameters in the classifier. This not only reduces the time complexity but also offers a flexibility of selecting most suitable distribution for the formant variation of the PCG signal. Moreover, the statistical distribution of the time domain PCG signal is also taken into consideration. In order to demonstrate statistical distribution, the same PCG signals used in Fig. 5 and corresponding formant variations (magnitude, frequency and phase) are taken into consideration. Different well known statistical distributions are tested on the PCG signal and extracted formant variations. It is observed that the fitting performance of the statistical distribution varies significantly for various cases. Based on the extensive experimentation, four different statistical distributions, namely t-location-scale, lognormal, Birnbaum-Saunders and logistic distribution are selected for the PCG signal, and variations of the formants magnitudes, frequencies and phases, respectively. In Table 2, name of the selected four statistical distributions along with their kernels and basic features is presented.

For the better visualization, in Fig. 6 statistical distributions of the PCG signal and formant variations (magnitude, frequency and phase) are shown. In this figure, histogram plots are shown using rectangles, and fitted distribution is shown by using green line. It is clearly observed from this figures that in each case a very satisfactory fitting performance is obtained by the chosen statistical distributions. The fitted model parameters as reported in Table 2 are considered as the features in the classification stage. From four different statistical models a 15-dimensional (15-d) feature dictionary can be formed as equation (11). Here F^3_{PCG} represents three model parameters obtained from the fitted model corresponding to the PCG signal. $F^{2\times2}$ refers to the two fitted model parameters extracted from each of the two formants corresponding to formants' magnitude, frequency and phase variation pattern.

$$F^{15}_{\text{pdf}} = F^{3}_{\text{PCG, pdf}(\Sigma,\eta,\nu)} + F^{2\times2}_{\text{Mag, pdf}(\mu,\sigma)} + F^{2\times2}_{\text{Freq, pdf}(\beta,\gamma)} + F^{2\times2}_{\text{Pha, pdf}(\kappa,\lambda)} \quad (11)$$

E. FORMATION OF THE FEATURE DICTIONARY

To form a feature dictionary for classification purpose, we are pensive on the evaluated spectrum of the sub-frames of the PCG signal. Apart from the various statistical measures, we consider six operations, namely mean absolute deviation (MAD), 1st quartile, 3rd quartile, inter quartile range (IQR), skewness and kurtosis. Considering two formants, these operations provide 12 parameters for each case of the formants (magnitude ($F^{12}_{\text{Mag, Stats}}$), frequency ($F^{12}_{\text{Freq, Stats}}$) and phase ($F^{12}_{\text{Pha, Stats}}$)). If we closely observe Fig. 6 the range of magnitude variation of 1st formant is a subset of magnitude variation found in case of 2nd formant. A very similar overlapping pattern is obtained in case phases of 1st and 2nd formants. On the contrary, the ranges of frequency values of the 1st and 2nd formants are exhibiting very negligible overlap, the two distributions are well separated. As a result, the features extracted from the formant frequency values are expected to provide better classification performance in comparison to the cases, where only formant magnitudes or formant phases are considered. These observations are further explained with some sample experimentation in the result section. Next our objective is to include more spectral characteristics in the feature dictionary. In view of that spectral entropy computed from the normalized PCG is considered as a potential feature (F^3_{PCG}). Another important phenomenon of our interest is the variation of spectral peaks corresponding to the formant frequency locations. In view of obtaining such information the Shannon's entropy of the magnitude of extracted formants is also considered as features ($F^2_{\text{Mag, Shannon's Entropy}}$). Finally, we propose a 30-dimensional (30-d) feature dictionary as equation (13). One portion of equation (13) covers statistical features (F^{15}_{Stats}) and the other part covers features from PDF fitting (F^{15}_{pdf}).

$$F^{15}_{\text{Stats}} = F^{12}_{\text{Stats}} + F^{2}_{\text{Mag, Shannon's Entropy}} + F^{1}_{\text{PCG, Sp. Entropy}} \quad (12)$$

$$F^{30} = F^{15}_{\text{pdf}} + F^{15}_{\text{Stats}} \quad (13)$$

F. ENSEMBLED BAGGED TREES CLASSIFIER

The proposed feature dictionary is used in the supervised classifier to obtain both 5 class and binary class classification. Among various supervised classifiers, in this paper ensembled bagged trees (EBT) classifier is utilized because

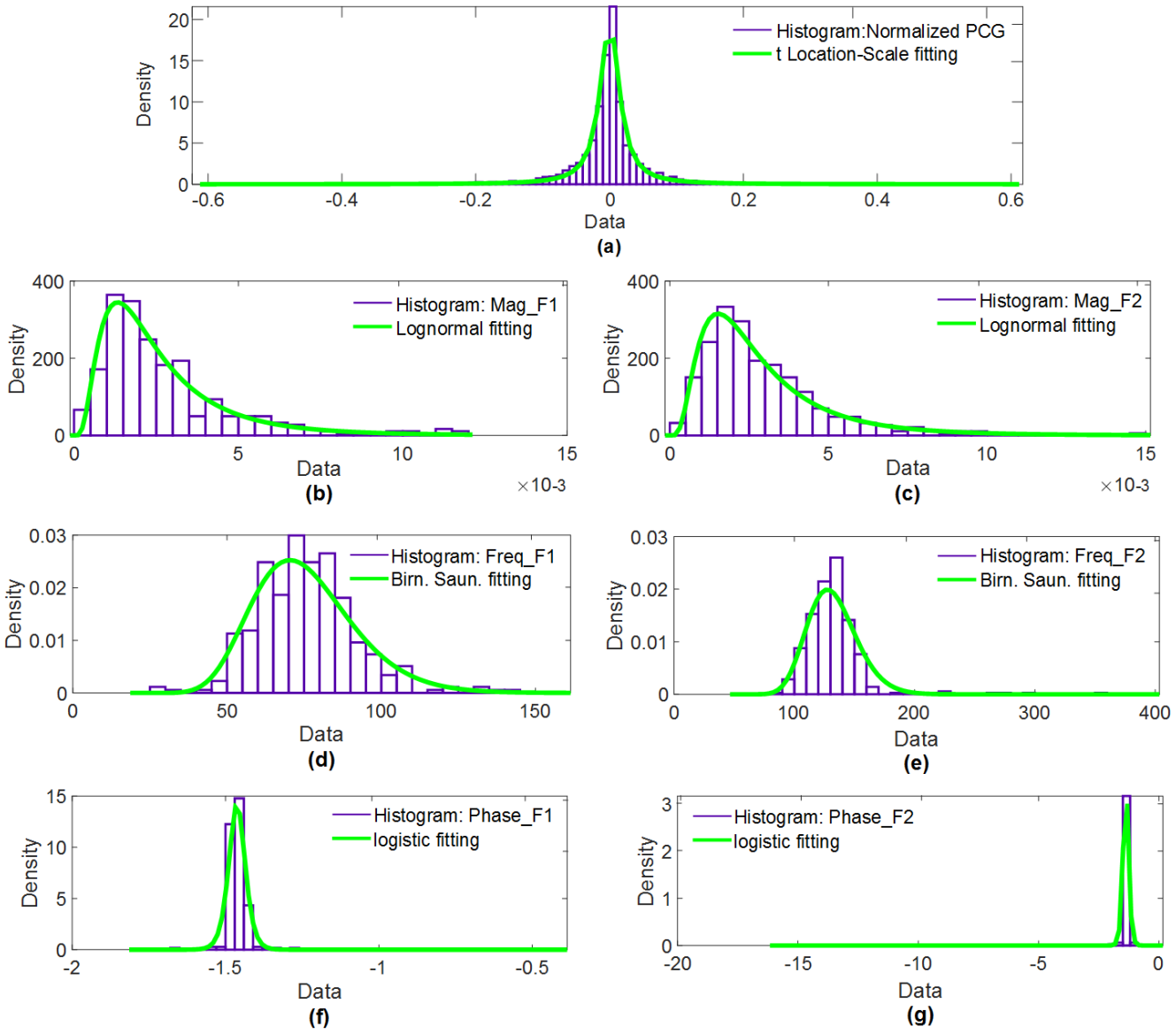

FIGURE 6. Distribution fitting of the feature vectors. (a) Normalized heart sound (t-location-scale distribution). (b) & (c) Magnitudes of the 1st and 2nd formant (Lognormal distribution). (d) & (e) Frequencies of the 1st and 2nd formant (Birnbaum-Saunders distribution). (f) & (g) Phases the 1st and 2nd formant (Logistic distribution).

TABLE 2. Distributions and their corresponding kernels used for model fitting.

Distributions	Kernels	Features
t-location-scale distribution (normalized PCG)	$\frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\eta\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left[\frac{\upsilon+\left(\frac{x-\Sigma}{\eta}\right)^2}{\upsilon}\right]^{-\left(\frac{\upsilon+1}{2}\right)}$, where $\Gamma(\bullet)$ is the gamma function.	Σ (location parameter), η (scale parameter) and ν (shape parameter).
Lognormal distribution (magnitude of the formants)	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{\frac{-(\log x-\mu)^2}{2\sigma^2}\right\}$	μ (mean of the logarithmic values) and σ (standard deviation of logarithmic values)
Birnbaum-Saunders distribution (frequency of the formants)	$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\left(\sqrt{x/\beta}-\sqrt{\beta/x}\right)^2}{2\gamma^2}\right\}\left(\frac{\sqrt{x/\beta}+\sqrt{\beta/x}}{2\gamma x}\right)$	β (location parameter) and γ (shape parameter).
Logistic distribution (phase of the formants)	$\frac{\exp\left(\frac{x-\kappa}{\lambda}\right)}{\lambda\left\{1+\exp\left(\frac{x-\kappa}{\lambda}\right)\right\}^2}$	κ (mean) and λ (scale parameter)

of its satisfactory performance in offering robust prediction against multiple classes. Ensembles help reducing variance and combat overfitting to the training dataset. In EBT a strong learner is formed by taking ensemble of a group of weak learners. In this method, bagging is formed by 'bootstrap' and 'aggregating' each weak learner, where a random subsample of data is sampled with bootstrapping. Since each subsample of data is sampled separately with replacement, bootstrapping secures independence and diversity. At the end, for classification majority of the averaged predictions are taken as the final decision. Because of the average predictions, this classifier is more stable and robust to the final prediction [46]. The classification performance is evaluated for various fold cross validations. Moreover, holdout validation is also performed with 30% held out score as a generalization measure of the classifier.

IV. RESULTS AND DISCUSSION

In this section, the performance of the proposed method is demonstrated considering various experimentation. As a primitive measure, the physicians usually listen to the variation of the intensity, softness and loudness of the heart sound to get an optimistic clue on valve disorders. In case of serious heart problems, which arise due to leaking of blood for imperfect closure of heart valve may listen to as soft swishing or hissing sound. The pathological variation of the sounds of multiclass (N, AS, MR, MS and MVP) and binary class cause significant differences in the statistics of the 1st and 2nd formant. The evaluation procedure of the results of the proposed method is described in the following subsections.

A. PERFORMANCE EVALUATION METRICS

To qualitatively evaluate the performance of the proposed model, well-known evaluation metrics such as, sensitivity (S_e), specificity (S_p), F1 score and accuracy are chosen to evaluate the proposed framework. They are defined as follows:

$$S_e = \frac{TP}{TP + FN}$$

$$S_p = \frac{TN}{TN + FP}$$

$$\text{F1 score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

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

where TP, TN, FP and FN represent the numbers of true positive, true negative, false positive and false negative, respectively..

B. FEATURE QUALITY ANALYSIS

In order to demonstrate the quality of the extracted features, a statistical boxplot representation is plotted in Fig. 7, where as an example eight features out of 30-d are chosen. For each feature, boxplot representation of five different

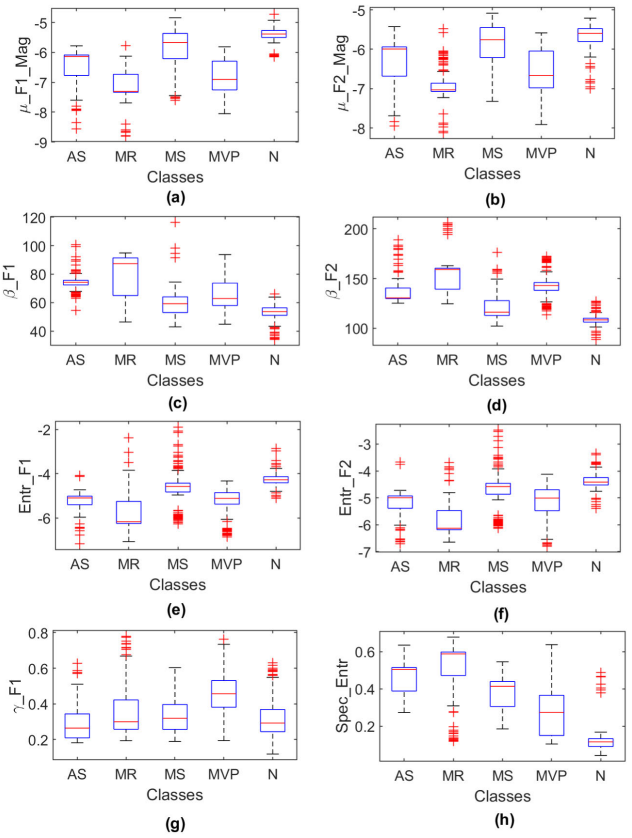

FIGURE 7. Boxplot for some of the features for all classes. (a) Mean of the logarithmic values of the magnitude of the 1st formant. (b) Mean of the logarithmic values of the magnitude of the 2nd formant. (c) Location parameter of the frequency of the 1st formant. (d) Location parameter of the frequency of the second formant. (e) Shannon's entropy of the first formants' magnitude. (f) Shannon's entropy of the 2nd formants' magnitude. (g) Shape parameter of the 1st formant and (h) Spectral entropy of the normalized PCGs.

classes is shown in the same figure. It is observed that for different features class separability varies with a satisfactory label of inter-class differences as well as within class compactness. It can be seen from the Fig. 7 that mean of the logarithmic values of the 1st and 2nd formant (μ_F1_Mag and μ_F2_Mag), location parameter of the 1st and 2nd formant (β_F1 and β_F2), entropy of the 1st and 2nd (Entr_F1 and Entr_F2), location parameter of the 1st (γ_F1) and spectral entropy (Spec_Entr) of the time series PCG have different median values. Based on this intra-class variations in the median values amongst these 30-d features, we have selected them for diseases classification. However, when all 30-d features are combinedly considered a very satisfactory classification performance is achieved. Thus, by combining all 30-d features in the proposed method there is a very high potential to precisely classify the valve defects.

Next, we want to show the affect of avoiding formant estimation in low energy region on overall performance. Based on our discussion in the previous section, it is expected that the elimination of feature extraction in those regions, where energy is very low helps in improving the

TABLE 3. Comparison of the performances for different cross validations with/without low energy regions of the PCG signal (30-d feature dictionary).

Cross validations	Without low energy regions	With low energy regions
	OA (%)	OA (%)
5-fold	99.12	98.72
10-fold	**99.28**	98.72
Holdout	99.20	98.49

TABLE 4. Comparison of the performances for different cross validations after elimination of the low energy regions of the normalized PCG signal (16-d feat- ure dictionary).

Cross validations	F^{16}_{Freq}	F^{16}_{Mag}	F^{16}_{Pha}
	OA (%)	OA (%)	OA (%)
5-fold	94.1	92.4	92
10-fold	**93.5**	91.8	92.2
Holdout	93.3	93	92

classification performance. In Table 3, the OAs obtained for the two cases are shown: with and without using the formants extracted from the low energy regions. It is observed from the Table 3 that OA is improved by avoiding formant features extracted from the low energy regions. In obtaining the classification results three different experimentations are carried out: 5-fold, 10-fold and holdout cross validations (30%). Hence, in what follows in all experiments, we avoid formant extracted from the low energy regions.

Next, we want to present the overall performance variation due to selecting any one of the three formant parameters, namely magnitude, frequency and phase. For the purpose of demonstration a 16-d feature dictionary is formed for each of these parameters as:

$$F^{16}_{\text{Mag}} = F^{12}_{\text{Mag, Stats}} + F^{2\times2}_{\text{Mag, pdf}(\mu,\sigma)} \tag{14}$$

$$F^{16}_{\text{Freq}} = F^{12}_{\text{Freq, Stats}} + F^{2\times2}_{\text{Freq, pdf}(\beta,\gamma)} \tag{15}$$

$$F^{16}_{\text{Pha}} = F^{12}_{\text{Pha, Stats}} + F^{2\times2}_{\text{Pha, pdf}(\kappa,\lambda)} \tag{16}$$

Here, F^{16}_{Mag} consists of features only related to magnitude, 12 statistical features and 4 PDF features as mentioned before. Similarly, F^{16}_{Freq} and F^{16}_{Pha} are constructed considering frequency and phase features, respectively.

Results obtained for these cases are presented in Table 4. As expected, it is clearly observed a relatively better classification is obtained using (F^{16}_{Freq}). Hence, proposed 30-d feature dictionary contains this 16-d frequency features. The affect of choice of this frequency feature in the proposed 30-d in comparison to selecting magnitude (F^{16}_{Freq}) or phase (F^{16}_{Pha}) features is also investigated. The results obtained for these three cases are shown in Table 5. Here also we can see that the best performance is achieved when F^{16}_{Freq} is used in 30-d feature dictionary. In this table classification performance of each class is also shown. However, to know detail about the class performance a confusion matrix is also shown in Fig. 8.

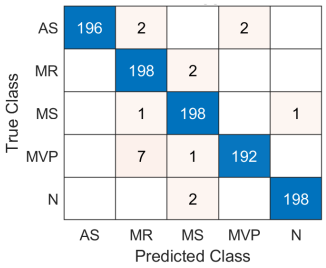

FIGURE 8. Confusion matrix.

TABLE 5. Performance of ensemble bagged trees classifier for different cross validations without low energy regions of the PCG signal (30-d feature diction- ary).

Cross validations	IA (%) N	IA (%) AS	IA (%) MR	IA (%) MS	IA (%) MVP	OA (%)
$F^{15}_{\text{pdf}} + F^{12}_{\text{Freq,Stats}} + F^{3}_{\text{Entropy}}$						
5-fold	99	98	99	97.50	95.50	99.12
10-fold	**99**	**98**	**99**	**99**	**96**	**99.28**
Holdout	100	98.30	98.30	96.70	96.70	99.20
$F^{15}_{\text{pdf}} + F^{12}_{\text{Mag,Stats}} + F^{3}_{\text{Entropy}}$						
5-fold	99.50	98	96	96	98	99
10-fold	99.50	98	96.50	97	95.50	98.75
Holdout	100	100	98.30	100	86.70	98.93
$F^{15}_{\text{pdf}} + F^{12}_{\text{Pha,Stats}} + F^{3}_{\text{Entropy}}$						
5-fold	99.50	97	97.50	96.50	95	98.80
10-fold	99	98	97.50	96.50	95	98.92
Holdout	98.30	96.70	100	98.30	95	98.93

Here in order to obtain the individual accuracy (IA) of a class, number of correctly identified instances of that class is divided by the total instances of that class. The OAs are calculated as per the performance evaluation metrics, that have been reported almost in all literature. It is clearly observed from the confusion matrix, that for each class a very high precision is achieved, even when we analyse the misclassified regions, a very few members of a class is misclassified. It is found that generally a member is misclassified to a class that closely resembles to the true class. For example, some members of the MVP class are found mistakenly classified as MR class. In biomedical perspectives MVP and MR are very closely related [3], and such type of classification may occur.

C. EFFECT OF NOISE ON CLASSIFICATION PERFORMANCE

During the acquisition of the PCG signal, noise can be included from the surrounding environment or from the internal physiological mechanisms. Removing noises from the PCG signal itself is a well-addressed research problem. In the proposed method, no noise-reduction scheme is employed at the front-end. Wavelet based denoising techniques are widely used for PCG denoising. One major concern in using the PCG denoising techniques prior to HVD detection is a chance of introducing distortion in the original PCG, which may deteriorate the original characteristics of the heart sounds [47]. In some cases, the noise reduction

TABLE 6. Effect of noise on the performance of the proposed model.

SNR (dB)	-10	-5	0	5	10	15
Accuracy (%)	93.76	96	97.60	98.40	98.52	99.12

techniques require secondary recording of the surrounding artifacts by using an extra microphone, which is unavailable in most of the cases. In [48], considering the noise in PCG signal as the additive white Gaussian noise (AWGN), an adaptive denoising algorithm is proposed based on the overlapping-group sparsity of the first-order difference of the PCG signal. Similar to various methods available in literature that deal with PCG denoising, we have considered a noisy environment by adding white Gaussian noise with the original PCG signal. The objective here is not to carry out denoising operations, rather to analyze the performance of the proposed method under noisy conditions. To check the robustness of our proposed model in case of noise, AWGN with zero mean is added to the normalized signal. An appropriate noise level is determined based on the value of signal-to-noise ratio (SNR). From -10 dB to 15 dB of SNR is varied to evaluate the accuracy level as per the performance metric described in the previous subsection. Results obtained by using the proposed method on noise-corrupted PCG signal at various levels of SNRs are presented in Table 6. It can be seen in Table 6 that at moderate to high SNR levels (SNR $>$ 0 dB), consistently very satisfactory performance is observed. As expected, a decrease in performance is found at a very low SNR of -5 dB and -10 dB. Further investigation on employing a noise reduction scheme and investigating various types of real-life noises could be a potential future work.

D. PERFORMANCE EVALUATION AND COMPARISON

Finally, we have demonstrated the performance of our proposed method with that of some recently reported results in Table 7 and Table 8, obtained by using variety of features along with different types of classifiers. Two widely available reliable datasets [33] and [27] are used in the proposed method for classification of HVDs. Different cross validations (5-fold/10-fold/holdout) are applied in implementing the proposed method for both noisy and without noisy data. Finally, binary and multiclass classification have been addressed with/without discarding low energy regions of the PCG signal. Usually, MFCCs are considered to be the baseline features for audio detection. With SVM classifier, an OA of 91.6% is reported with MFCCs features in [28]. At the same time the authors have tried with DWT coefficients for the same classifier, and have managed slightly higher accuracy. They use MFCCs and DWT features together, and found that this combination can ''diagnose heart disorders in patients up to 97% accuracy'' [28]. Here 5-fold cross validations are used to achieve a maximum accuracy of 97.90%.

TABLE 7. Comparison of the performances of different approaches on github dataset [27].

Methods	No. of Classes	Classifier	OA (%)
WSST [25]	4	RF	95.13
CT [26]	4	MCC	98.33
MFCC [28]	5	SVM	91.60
DWT [28]	5	SVM	92.30
MFCC + DWT [28]	5	SVM	97.90
TQWT + EMD [6]	5	DLT	98.48
CNN + RNN [38]	5	CNN	98.32
Proposed method	5	EBT	**99.28**

TABLE 8. Comparison of the performances of different approaches on Physionet/CinC 2016 [33].

Methods	No. of data	S_e (%)	S_p (%)	F1 score (%)	Accuracy (%)
CNN [41]	3240	84.67	98.33	-	91.50
ECNN [42]	3240	82.33	86.47	-	86.05
DNN [8]	1081	85.74	86.73	84.58	85.65
CNN [39]	301	78.10	96.40	-	87.30
LSTMs+GA [40]	405	84.50	-	87.40	87.30
SVM [6]	744	78.59	93.30	-	86
Proposed method	2766	93.45	93.46	93.45	**93.46**

In [6], nonlinear dynamics of the PCG signal is used. The signal's first derivative is decomposed into a set of frequency subbands with TQWT method. Followed by FA-MVEMD, the first two IMFs are extracted according to predominant energy, where Shannon energy is used to extract the characteristic envelope of IMFs. Finally, neural networks are then used to model, identify and classify HVDs for 5-classes of PCG signals based on deterministic learning theory, achieving an OA of 98.48%. In [25], the authors utilize WSST and CT to find time-frequency matrix for feature extraction to identify 4-classes of HVDs [25], [26]. The OA obtained from statistical features of magnitude and phase of the TFM of WSST is 95.13%. Altogether 13 features were fed to RF classifier to have this OA. On the other hand, 300 features from time-frequency matrix of CT are used to classify the same four classes of diseases. By sacrificing considerable computational burden, an OA of 98.33% is found from this huge set of features with MCC classifier. For both the cases 10-fold cross validations were used. It is to be noted that the dataset had 5 classes of heart sounds for classification. In both the cases the authors dropped MVP from the dataset while evaluating the OA. Apart from these feature based approach, authors of [38] utilized deep learning, where CNN networks are used for the identification of various valvular heart diseases through PCG recordings of [27]. On an average 0.88 ($\pm$0.21) s is required to get the classification result for a single trial, which is found reasonable considering the duration of the input data (2.44 $\pm$ 0.37 s). In this case, an Intel® Core i7 5500U CPU @ 2.4 GHz along with 12 GB RAM is used.

The performance of the Physionet/CinC 2016 [33] on various algorithms are presented in Table 8, where

normal/abnormal classes are targeted for classification. In most cases, different types of neural networks are used on various number of data. Rather than considering all recordings, we put all male and female subjects for feature extraction. This helps using diversity of data from different perspective, like age, length of the recordings, places etc. A total of 2766 data are used, by discarding 474. An OA of 93.46% is achieved with proposed method, which is very satisfactory compared to other recently reported methods. In our method, the feature dimension is 30-d which is extremely low in comparison with some methods. The proposed method offers very satisfactory performance using only a very low feature dimension.

V. CONCLUSION

The unique idea proposed in this paper is to utilize the temporal variation of formants of the PCG signal for HVD classification. Moreover, by introducing the PDF models to represent the formant variation, a drastic reduction in feature dimension is achieved along with a satisfactory feature quality. It is shown that the classification performance can be improved by avoiding formant estimation in the low energy regions, which is expected as per the acoustic nature of the formants. In the proposed features, the PDF model parameters obtained for all three cases of the formant variation, namely magnitude, frequency and phase are utilized. However, statistical measures for the formant frequency variation is found more effective than that of the magnitude and phase variation. Classification performance on each class is found very satisfactory in case of various cross validations setups. Comparative performance analysis is carried out with some recent methods and it is found that the proposed method offers better classification accuracy with the overall accuracy values of 93.46% and 99.28% in 10-fold cross validations for the two datasets. However, this automated method hopefully will play an efficient role with enhanced diagnostic accuracy for the clinicians and development of noninvasive biomedical instrumentation.

REFERENCES

[1] G. Roth, C. Johnson, A. Abajobir, and F. Abd-Allah, "Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015," *J. Amer. College Cardiol.*, vol. 70, no. 1, pp. 1–25, Jul. 2017.

[2] L. H. Schwamm, N. Chumbler, E. Brown, G. C. Fonarow, D. Berube, K. Nystrom, R. Suter, M. Zavala, D. Polsky, K. Radhakrishnan, N. Lacktman, K. Horton, M.-B. Malcarney, J. Halamka, and A. C. Tiner, "Recommendations for the implementation of telehealth in cardiovascular and stroke care: A policy statement from the American heart association," *Circulation*, vol. 135, no. 7, p. e24–e44, Feb. 2017.

[3] A. K. Abbas and R. Bassam, "Phonocardiography signal processing," *Synth. Lectures Biomed. Eng.*, vol. 4, no. 1, pp. 1–194, Jan. 2009.

[4] C. Liu, 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, Dec. 2016.

[5] J. Zamorano, P. Lancellotti, L. Pierard, and P. Pibarot, *Heart Valve Disease: State of the Art*. Cham, Switzerland: Springer, 2019.

[6] H. Hazeri, P. Zarjam, and G. Azemi, "Classification of normal/abnormal PCG recordings using a time-frequency approach," in *Proc. Anal. Integr. Circuits Signal Process.* Cham, Switzerland: Springer, 2021, pp. 1–7.

[7] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, and A. K. Sangaiah, "Diagnosis of heart diseases by a secure Internet of Health Things system based on autoencoder deep neural network," *Comput. Commun.*, vol. 162, pp. 31–50, Oct. 2020.

[8] 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, pp. 1–11, Jun. 2020.

[9] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9393–9400, Sep. 2018.

[10] Q. Que, Z. Tang, R. Wang, Z. Zeng, and J. Wang, "CardioXNet: Automated detection for cardiomegaly based on deep learning," in *Proc. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, 2018, pp. 612–615.

[11] 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.

[12] T. T. K. Munia, "Heart sound classification from wavelet decomposed signal using morphological and statistical features," in *Proc. Comput. Cardiol. Conf. (CinC)*, Vancouver, BC, Canada, 2016, pp. 597–600.

[13] A. Moukadem, A. Dieterlen, and C. Brandt, "Shannon entropy based on the S-Transform spectrogram applied on the classification of heart sounds," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 704–708.

[14] A. Thalmayer, S. Zeising, G. Fischer, and J. Kirchner, "A robust and real-time capable envelope-based algorithm for heart sound classification: Validation under different physiological conditions," *Sensors*, vol. 20, no. 4, p. 972, Feb. 2020.

[15] Z. Yan, Z. Jiang, A. Miyamoto, and Y. Wei, "The moment segmentation analysis of heart sound pattern," *Comput. Methods Programs Biomed.*, vol. 98, no. 2, pp. 140–150, May 2010.

[16] S. Ismail, I. Siddiqi, and U. Akram, "Localization and classification of heart beats in phonocardiography signals—A comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–27, Dec. 2018.

[17] 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, 2012.

[18] D. Kumar, P. Carvalho, M. Antunes, J. Henriques, L. Eugenio, R. Schmidt, and J. Habetha, "Detection of S1 and S2 heart sounds by high frequency signatures," in *Proc. Int. Conf. Eng. Med. Biol. Soc.*, 2006, pp. 1410–1416.

[19] V. N. Varghees and K. I. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomed. Signal Process. Control*, vol. 13, pp. 174–188, Sep. 2014.

[20] M. Rosół and L. Więckowski, "Embedded heart rate analysis based on sound sensing," in *Proc. Int. Conf. Methods Models Automat. Robot. (MMAR)*, 2019, pp. 629–633.

[21] V. Nigam and R. Priemer, "Accessing heart dynamics to estimate durations of heart sounds," *Physiol. Meas.*, vol. 26, p. 1005, Oct. 2005.

[22] 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.

[23] 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 Inform.*, vol. 18, no. 4, pp. 1138–1152, Feb. 2013.

[24] A. Gharehbaghi, T. Dutoit, A. Sepehri, P. Hult, and P. Ask, "An automatic tool for pediatric heart sounds segmentation," in *Proc. Conf. Comput. Cardiol.*, 2011, pp. 37–40.

[25] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using timefrequency magnitude and phase features," *IEEE Sensors Lett.*, vol. 3, no. 12, Dec. 2019, Art. no. 7002604.

[26] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals," *Comput. Biol. Med.*, vol. 118, May 2020, Art. no. 103632.

[27] *Classification of Heart Sound Dataset Available in GitHub*. Accessed: Aug. 30, 2021. [Online]. Available: https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features

[28] T. Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, 2018.

[29] S. Patidar and R. B. Pachori, "Classification of cardiac sound signals using constrained tunable-Q wavelet transform," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7161–7170, 2014.
[30] A. D. Ricke, R. J. Povinelli, and M. T. Johnson, "Automatic segmentation of heart sound signals using hidden Markov models," in *Proc. Conf. Comput. Cardiol.*, 2005, pp. 953–956.
[31] P. Sedighian, A. W. Subudhi, F. Scalzo, and S. Asgari, "Pediatric heart sound segmentation using hidden Markov model," in *Proc. 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 5490–5493.
[32] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Apr. 2016.
[33] *PhysioNet*. Accessed: Aug. 30, 2021. [Online]. Available: https://physionet.org/content/challenge-2016/1.0.0/
[34] T. Hen, K. Kuan, L. Celi, and G. Clifford, "Intelligent heart sound diagnostics on a cell phone using a hands-free kit," in *Proc. AAAI Spring Symp. Artif. Intell. Develop.*, 2010, pp. 26–31.
[35] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021.
[36] T. Li, C. Qing, and X. Tian, "Classification of heart sounds based on convolutional neural network," in *Proc. Int. Conf. Internet Multimedia Comput. Service*, 2017, pp. 252–259.
[37] Y. Tsao, T.-H. Lin, F. Chen, Y.-F. Chang, C.-H. Cheng, and K.-H. Tsai, "Robust S1 and S2 heart sound recognition based on spectral restoration and multi-style training," *Biomed. Signal Process. Control*, vol. 49, pp. 173–180, Mar. 2019.
[38] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105940.
[39] Y. He, W. Li, W. Zhang, S. Zhang, X. Pi, and H. Liu, "Research on segmentation and classification of heart sound signals based on deep learning," *Appl. Sci.*, vol. 11, no. 2, p. 651, Jan. 2021.
[40] P. Li, Y. Hu, and Z.-P. Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102474.
[41] 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.
[42] J. M.-T. Wu, "Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model," *Appl. Soft Comput.*, vol. 78, pp. 29–40, May 2019.
[43] A. S. M. M. Jameel, S. A. Fattah, R. Goswami, W.-P. Zhu, and M. O. Ahmad, "Noise robust formant frequency estimation method based on spectral model of repeated autocorrelation of speech," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 6, pp. 1357–1370, May 2016.
[44] K. Kajihara, S. Izumi, S. Yoshida, Y. Yano, H. Kawaguchi, and M. Yoshimoto, "Hardware implementation of autoregressive model estimation using Burg's method for low-energy spectral analysis," in *Proc. Int. Workshop Signal Process. Syst. (SiPS)*, 2018, pp. 199–204.
[45] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. London, U.K.: Pearson, 1988.
[46] S. B. Kotsiantis, G. E. Tsekouras, and P. E. Pintelas, "Bagging model trees for classification problems," in *Panhellenic Conf. Informatics*. Volos, Greece: Springer, 2005, pp. 328–337.
[47] D. Gradolewski and G. Redlarski, "Wavelet-based denoising method for real phonocardiography signal recorded by mobile devices in noisy environment," *Biomed. Signal Process. Control," Elsevier*, vol. 52, pp. 119–129, 2014.
[48] S.-W. Deng and J.-Q. Han, "Adaptive overlapping-group sparse denoising for heart sound signals," *Biomed. Signal Process. Control*, vol. 40, pp. 49–57, Feb. 2018.

MONJUR MORSHED (Graduate Student Member, IEEE) received the B.Sc. degree from the Chittagong University of Engineering and Technology and the M.Sc. degree from the University of Ulm, Germany. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology. He is also working as an Associate Professor with the Department of EEE, Ahsanullah University of Science and Technology, Bangladesh. His research interests include biomedical signal processing, image processing, and machine learning.

SHAIKH ANOWARUL FATTAH (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, and the Ph.D. degree in electronics and communication engineering (ECE) from Concordia University, Canada. He held a visiting postdoctoral position at Princeton University, Princeton, NJ, USA. He is currently working as a Professor with the Department of Electrical and Electronics Engineering (EEE), BUET. He has published more than 215 international journal articles/conference papers with some best paper awards and delivered more than 80 keynote/invited talks in many countries. His research interests include signal processing, machine learning, and biomedical engineering. He is a fellow of the IEB. He is a member of the IEEE PES LRP, the IEEE Public Visibility Committee, and the IEEE Smart Village Education Committee. He received several awards, namely the Concordia University's Distinguished Doctoral Dissertation Prize in ENS in 2009, the 2007 URSI Canadian Young Scientist Award, the Dr. Rashid Gold Medal (in M.Sc.), the BAS-TWAS Young Scientists Prize in 2014, the 2016 IEEE MGA Achievement Award, the 2017 IEEE R10 HTA Outstanding Volunteer Award, and the 2018 IEEE R10 Outstanding Volunteer Award. He served as the Chair/the Founding Chair for different IEEE Society Chapters in Bangladesh, such as IEEE SPS, EMBS, RAS, and SSIT. He was the IEEE Bangladesh Section Chair for the period 2015–2016. He is the Chair of the IEEE PES-HAC and the IEEE SSIT Chapters Committee. He served on IEEE HAC and various committees for IEEE R10, IEEE EAB, and IEEE SIGHT. He served key positions in many international conferences, such as the General Chair for the IEEE R10-HTC2017 and the TPC Chair for the IEEE TENSYMP2020. He is the Editorial Board Member of IEEE Access, *IEEE Potentials*, *BioMed Research International*, and *JEE* (IEB), and the Editor-in-Chief of IEEE PES Enews.

MOHAMMAD SAQUIB (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Bangladesh University of Engineering and Technology, Bangladesh, in 1991, and the M.S. and Ph.D. degrees in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1995 and 1998, respectively. He worked as a member of the Technical Staff with the Massachusetts Institute of Technology Lincoln Laboratory, and as an Assistant Professor with Louisiana State University, Baton Rouge, LA, USA. He is currently a Professor with the Electrical Engineering Department, The University of Texas at Dallas, Richardson, TX, USA. His current research interests include the various aspects of wireless data transmission, radio resource management, and signal processing techniques for low-cost radar applications.