IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2019, accepted March 1, 2019, date of publication March 12, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904601

A Novel Technique to Diagnose Sleep Apnea in Suspected Patients Using Their ECG Data

SYEDA QURATULAIN ALI[1], SOHAIL KHALID[2], (Member, IEEE), AND SAMIR BRAHIM BELHAOUARI[3], (Member, IEEE)

[1]Department of Electrical Engineering, Universiti Teknologi Petronas, Seri Iskandar 32610, Malaysia
[2]Department of Electrical Engineering, Riphah International University, Islamabad 44000, Pakistan
[3]Division of Information & Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Corresponding author: Samir Brahim Belhaouari (sbelhaouari@hbku.edu.qa)

This work was supported by the Qatar National Library.

ABSTRACT Sleep Apnea is a breathing disorder that occurs while the patient is sleeping. Traditionally, Polysomnography is used to diagnose it. However, it is quite inconvenient and expensive. Because of the troublesome diagnosis, this ailment often remained undiagnosed. This paper aims at the development of such a method that provides an easy diagnostic solution to the doctors. Electrocardiogram (ECG) is one of the most common tests done at the hospitals. In this paper, we aim to develop a method which deploys ECG data to diagnose the sleep ailment, Apnea. A technique deploying wavelet packet transform on RR interval of ECG has been presented. Probability density functions of data, both when Apnea is present and when it is not, are obtained by constructing histograms of decision variable for each signal segment. From the overlapping PDFs of the normal and abnormal cases, a threshold is then derived. This helped in segregating the Apnea cases from normal cases. The stated method provided a 100% accuracy in diagnosing Sleep Apnea.

INDEX TERMS Sleep Apnea, ECG, signal processing.

I. INTRODUCTION

Sleep Apnea (SA) is an extremely common sleep time disease. An episode of Apnea is considered to occur when there is an absence of breathing for ten seconds or more. A single episode of Apnea can last for more than a minute, in extreme cases. An occasional episode of Apnea is common and treated as normal until there are more than ten episodes per hour. According to the American Academy of Sleep Medicine, a person is considered to have Sleep Apnea if he/she has more than 10 Apnea events per hour. Although, it is roughly as common as asthma [14], however, it remains unrecognized by primary care physicians [10]. Apnea causes constant disruption of sleep and low oxygen levels because of which people with Sleep Apnea face early morning headaches and fatigue, loss of memory and may also lead to brain impairment. Sleep Apnea is sometimes also a factor of high blood pressure, heart disease and stroke [10]. In a few death cases, dying silently in sleep appears to be a merciful way of departure but indeed it may sometimes be unnecessary death because of untreated Sleep Apnea [10]. A comprehensive study for its timely diagnosis is necessary because there are many high-risk health issues associated with Sleep Apnea.

Clinically, SA is diagnosed using nocturnal Polysomnography of the patient. At least one night of polysomnographic recording of patient's usual sleep hours is required for diagnosis. Clinical Polysomnography involves Electrocardiogram (ECG), Electroencephalogram (EEG), Electromyogram (EMG), Electrooculography (EOG) and a number of other electrophysiologic measures [6]. The detection tests are done in sleep laboratories. The number and arrangement of the signals vary from one laboratory to the other but, in general, there is a minimum montage of 12 channels. Based on the recorded data of these signals, a comprehensive study of the nocturnal activity is performed and based on the results the suspected patient is diagnosed [6]. Although polysomnography has a very high accuracy, however, it is very expensive since it involves a number of costly electrophysiologic tests. In addition to the expenses, the associated physical fatigue also demoralize the patients to investigate the sleep disorders they are facing.

Research has been started on signal processing based diagnosis of SA. The disruption in breathing caused by

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

Sleep Apnea affects few physiological signals, in one way or the other. For the extraction of the discriminatory features in the stated signals, many signal processing tools, like Fourier transform, wavelet transform, etc., are in use. Several researchers have deployed signals like EEG, SpO2 and ECG, both individually and in different combinations of each other for developing a signal processing based solution to Sleep Apnea diagnosis [3]–[5], [7], [8], [11]–[13], [15], [16], [18], [19], [22], [23]. Some algorithms have shown 100% accuracy for diagnosing Sleep Apnea [7], [12], [13], [18]. However, there are several issues including complexity, memory inefficiency and the need for human intervention, etc. that are needed to be addressed. There appears a wide scope for improvement in the signal processing diagnostic algorithms especially regarding simplicity, memory efficiency and robustness in diagnosis.

This paper presents a simple method for Sleep Apnea diagnosis deploying ECG data of the suspected patients. For the study purpose, the ECG data of patients has been borrowed from online MIT-BIH Sleep Apnea database [1]. RR intervals of the ECG waveform have been researched for detecting possible changes in case of SA. Based on the physiological aspects associated SA, there is a particular narrow band of frequency in the spectrum of the RR interval, where the changes are expected to occur in case of SA. The presented technique deploys Discrete wavelet packet transform (DWPT) [2] in the stated narrow frequency band. The frequency transformation and statistics generated from it are followed by a comprehensive statistical study to obtain the right threshold for diagnosing Apnea cases. The performance of the developed method is evaluated and validated in terms of probability of detection and probability of error, accuracy, specificity and sensitivity. This paper is organized as follows. The proceeding section discusses the data sources and the technique in detail. Section 3 presents the results with the associated discussion. Section 4 concludes the paper followed by the references in section 5.

II. MATERIALS AND METHODS

A. STUDY GROUP

In this research, MIT-BIH online sleep apnea database [1] has been deployed. The single channel ECG data was extracted from the Polysomnographic recordings by a sampling rate of 100 Hz. The average duration of recording is 8 hours. The sleep recordings originated from subjects between 28 and 63 years of age, with the weight in range of 53 and 135 kg. The duration of the recordings varied between 401 and 578 min (average: 492, standard deviation: 32 min.). Two databases have been provided, one for training and testing (dataset A) and the other for validation (dataset A_v). We have subdivided dataset A in A_l and A_t for the learning and testing phase respectively. By learning the features from A_l, a threshold is achieved, which is later tested using A_t, whereby cases in the dataset A_t are classified as normal or abnormal. Later, the decision of our scheme is compared with the human annotations provided by MIT online database. To observe the behavior of our developed technique in a diverse and richer database, it would be validated using A_v, which contains 35 unique cases of normal and Apnea patients, all shuffled up. Similarly, as in the testing phase, the decisions are compared with the human annotations provided by MIT-BIH online Sleep Apnea database and results are recorded.

B. OUR HYPOTHESIS

Since Sleep Apnea refers to the cessation of breath for 10 seconds or more [6], more dominant low-frequency components are expected to be present in the power spectrum of RRI. Drinnan *et al.* [8] have noticed these low frequency components using Fourier Transform and have developed a technique that relies on the ratio of the spectral power between 0.01 and 0.05 cycles/beat and 0.005 and 0.01 cycles/beat to detect Sleep Apnea. Fig. 1 depict the power spectrum of RRI in Apnea and normal case, respectively, as observed by Drinnan *et al.* [8].

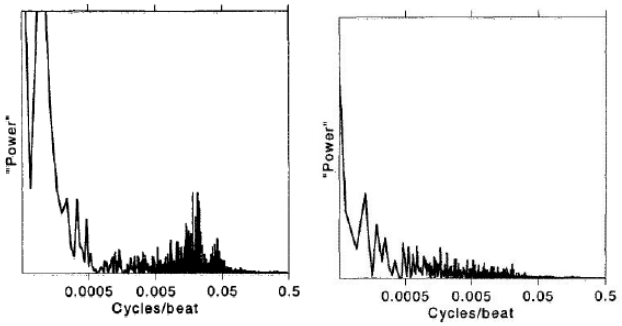

FIGURE 1. Power spectrum Apnea FFT (left) and normal FFT (right).

In Fig. 1, it is observable that the power spectrums of RRI data contain distinctive features when Apnea is present (spectral peak between 0.005-0.05Hz). Our technique also makes use of this distinctiveness, in its own way, in order to make it simpler and time efficient. It takes the ECG data as an input and derives the RRI data from it and then obtains magnitude squared DWPT coefficients over the frequency band of interest (F_bOI)(to be discussed in detail later), where the frequency features specific to Apnea are prominent. A suitable decision variable (DV) is defined in order to classify Apnea or normal cases.

Assume that the underlying statistics of the decision variable (DV) when Apnea is present and also when Apnea is absent are both Gaussian (having different mean values say μ_a and μ_n for Apnea and normal respectively). This assumption would provide a smarter way to derive the probability of miss detection as well as false alarm. Hypothetically, let Fig. 2*a* and 2*b* represent the PDF of the decision variable in normal and abnormal case respectively, while Fig. 2*c* is the combined plot of Fig. 2*a* and 2*b*. Confusion Area (*CA*) in Fig. 2*c* represents the overlap between the PDF of normal and abnormal cases, giving rise to the false alarm and miss detection. *Th* in Fig. 2*c* represents the Threshold, which is

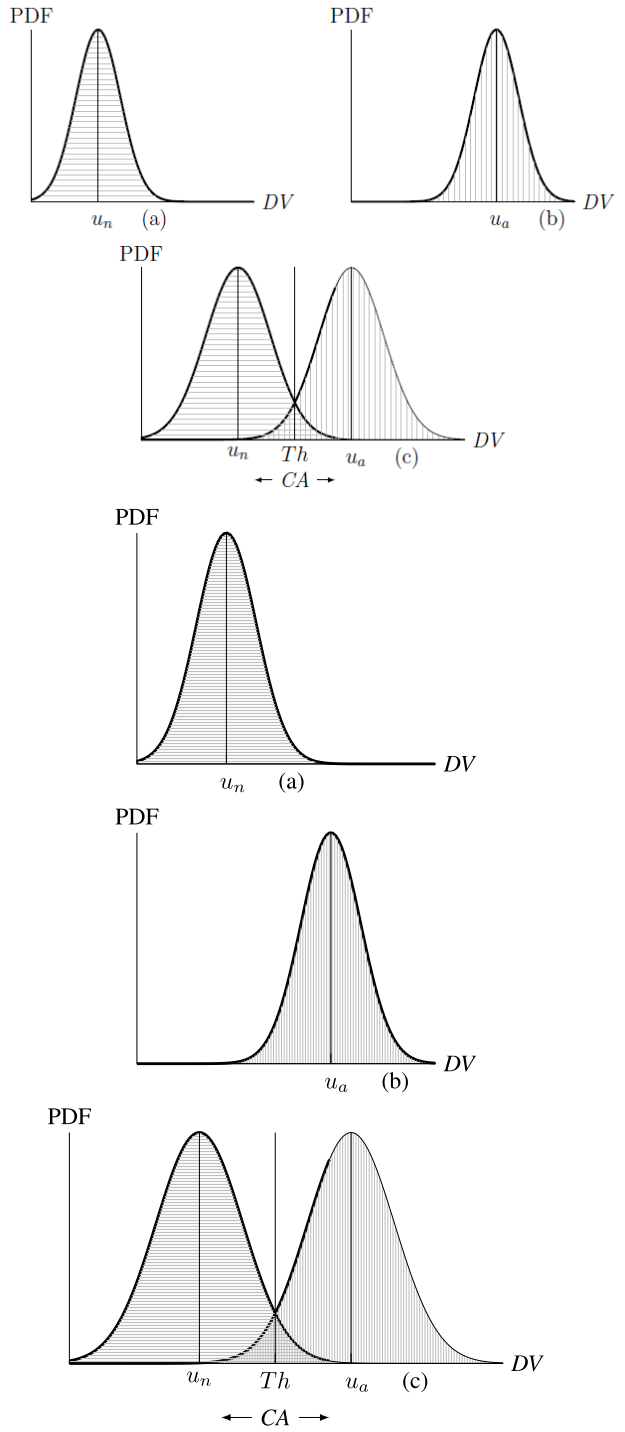

FIGURE 2. Expected PDF in F_bOI for normal and abnormal (Apnea) cases.

defined here as the intersection point of the two overlapping curve.

The RRI data, given by symbol $x(i)$, is segmented into 5 minutes blocks. Let the decision variable (DV) be written as d_k, where k is the block number. d_k is an average number derived from magnitude-squared of the DWPT coefficients of the RRI data over F_bOI. If its statistics (as provided by PDFs $p(d_k|Apnea)$ and $p(d_k|normal)$ are assumed Gaussian, as illustrated in Fig. 2*a* and 2*b* respectively, then probability of miss detection and probability of false alarm are given by:

$$P_{md} = \int_{R_0} P(d_k|Apnea) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \int_{-\infty}^{Th} e^{\frac{-(d_k-\mu_a)}{2(\sigma_a)^2}} \mathrm{d}d_k \quad (1)$$

$$P_{fa} = \int_{R_1} P(d_k|normal) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_{Th}^{\infty} e^{\frac{-(d_k-\mu_n)}{2(\sigma_n)^2}} \mathrm{d}d_k \quad (2)$$

where R_0 is the normal region on the left side of threshold line and R_1 is the Apnea region on the right side of threshold line.

Hence, the probability of error, P_e, can be defined as:

$$P_e = P_{md} + P_{fa} \quad (3)$$

Accordingly, the probability of correct detection, P_d is:

$$P_d = 1 - P_e \quad (4)$$

The assumption of Gaussianity (specifically at the tails of PDFs where confusion area exists) is subject to verification. The first step was to obtain the histogram of the decision variable d_k in both normal and Apnea cases, and, use that to obtain (empirically) the threshold and thereby, semi-analytical values of P_e and P_d. In the light of the above discussion, it is important to derive a suitable decision variable that may support an accurate yet simple diagnosis of Sleep Apnea.

C. THE SYSTEM MODEL

The system model is based on three core phases, i.e., RRI processing, designed technique and the developed decision phase, as depicted in Fig. 3. In Fig. 3, $x(i)$ represents RRI data stream, $x_k(m)$ are the processed RRI data blocks; d_k refers to decision variable extracted by the technique phase and A/N

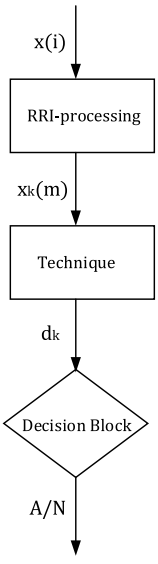

FIGURE 3. The system model.

represents the decision phase output referring an RRI data block as abnormal or normal.

1) RRI PROCESSING PHASE

In RRI processing phase, the RRI data is initially prepared according to the need of the next phase. First of all, $x(i)$ is segmented into N blocks, each of length M. Thus,

$$x_k(m) = x(i - kM), \tag{5}$$

where k is the block number and $0 \leqslant m \leqslant M-1, 0 \leqslant i \leqslant N-1$ and $0 \leqslant k \leqslant N/M-1$.

Afterwards, $x_k(m)$ is corrected by removing outliers and false shoot up in $x_k(m)$. R-R intervals often have missed or impractical magnitudes due to several reasons including equipment sensitivity and programming limitations. In order to eliminate the issue of false shoot up/shrink magnitudes, two thresholds ($th_l = 0.4$ and $th_u = 2$) have been defined for lower and upper cut-off limits. If a sample of $x_k(m)$, crosses the defined threshold, its value is adjusted by taking the mean of two previous and two later samples. The false shoot up/shrink magnitudes in $x_k(m)$, as discussed earlier, are replaced with the corrected values in $x_k(m)$, which is the output of RRI processing block.

2) THE TECHNIQUE PHASE

This phase is the backbone of the methodology as it provides the decision variable as its output while taking in $x_k(m)$ as its input. The schematic diagram of this block along with the sub-blocks is depicted in Fig. 4. The frequency

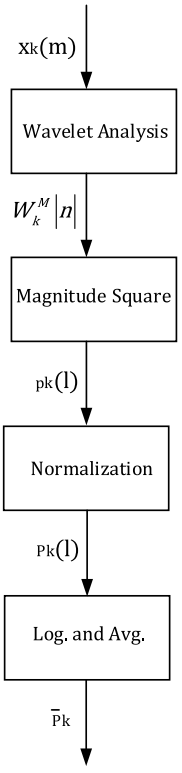

FIGURE 4. The components of the technique phase.

domain analysis of RRI is considered as a tool to discriminate between normal and abnormal cases. Logical determination of frequency band of interest (F_bOI) is of core importance, as the rest of the analysis totally depends on it. F_bOI has been defined from $0-0.125$Hz. This selection has a physiological explanation as Sleep Apnea events consist of respiratory arrests lasting over 10 sec, including the awakening response after Apnea. The minimum limit of respiratory cessation is 10 seconds, that corresponds to the frequency of 0.1Hz. The longest Apnea time usually observed lasts approximately 2 min (0.008 Hz) [21]. Therefore, SA-positive(Sleep Apnea Positive) subjects are expected to have a higher power in $0.008 - 0.1$Hz. In this research, we are performing short time analysis of ECG signal, whereby the whole night of ECG signal is divided in blocks of 5 minutes intervals. Due to programming limitations the upper limit of 0.125Hz has been used. Given the presence of new frequencies in F_bOI during Apnea event, the technique relies on an in-depth study of F_bOI, utilizing wavelet transformation.

In order to achieve F_bOI and decomposition of F_bOI for frequency content analysis, Wavelet multi-resolution analysis would be performed on $x_k(m)$. Till the achievement of F_bOI, detailed coefficients are discarded to ensure minimal memory usage. However, both approximate and detailed components are used while decomposing and analyzing F_bOI. In order to fit into wavelet multi-resolution framework, $x_k(m)$ is considered to be the approximate coefficients at scale $m = 0$ ($S_{0,n}$), defined by

$$S_{0,n} = x_k(m), \tag{6}$$

where n is the translation parameter in the wavelet transformation. In order to achieve maximum simplicity, Haar function [2] has been deployed as base wavelet, which later proved to be sufficient enough to segregate normal and abnormal cases successfully. The general formulas for approximate($S_{m,n}$) and detailed coefficients ($T_{m,n}$) using Haar function are:

$$S_{m,n} = \frac{1}{\sqrt{2}}[S_{m,2n} + S_{m,2n+1}], \tag{7}$$

and

$$T_{m,n} = \frac{1}{\sqrt{2}}[S_{m,2n} - S_{m,2n+1}]. \tag{8}$$

Using Eq. 7 and 8, the first step wavelet decomposition of $S_{0,n}$ leads to the approximate and detailed coefficients as,

$$S_{1,n} = \frac{1}{\sqrt{2}}[S_{0,2n} + S_{0,2n+1}], \tag{9}$$

and

$$T_{1,n} = \frac{1}{\sqrt{2}}[S_{0,2n} - S_{0,2n+1}], \tag{10}$$

where $S_{1,n}$ and $T_{1,n}$ represents the approximate and detailed coefficients respectively. Using the wavelet-multi-resolution concept, two further decompositions are performed on approximate coefficients to achieve the F_bOI while detailed

coefficients are discarded. Third level approximate coefficient $S_{3,n}$ contains the required F_bOI.

$$S_{3,n} = \frac{1}{\sqrt{2}}[S_{2,2n} + S_{2,2n+1}]. \quad (11)$$

At $m = 3$, $S_{4,n}$ and $T_{4,n}$ are achieved as follows:

$$S_{4,n} = \frac{1}{\sqrt{2}}[S_{3,2n} + S_{3,2n+1}], \quad (12)$$

and

$$T_{4,n} = \frac{1}{\sqrt{2}}[S_{3,2n} - S_{3,2n+1}]. \quad (13)$$

For an in-depth analysis of the F_bOI, from this point onward detailed coefficients would also be decomposed. Till $m = 6$, decomposition is performed using Eq. 7 and 8. Fig. 5 highlights the decomposition tree.

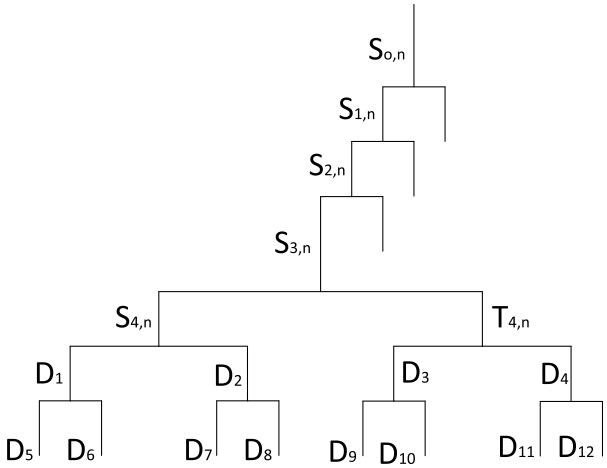

FIGURE 5. DWPT decomposition of RRI.

The concatenated array of decomposed wavelet coefficient would be,

$$W_i^6 = (D_{i-7}, D_{i-6}, ..., D_i), \quad (14)$$

where $i = 12$. W^6 depicts that six steps of decomposition have been performed. The achieved set of decomposed values provides an insight to the defined F_bOI. Hence, our focus is to analyze and mark the differences in W_i^6 for normal and abnormal cases. Now in order to obtain a decision variable from W_i^6, few mathematical steps have to be performed. First of all, for avoiding the value cancelation between positive and negative value while averaging, magnitude squared DWPT coefficients are calculated, as shown in Eq. 15

$$p_k(l) = |W_i^6(l)|^2 \quad (15)$$

where l is the length of the array. Later, in order to avoid the block to block variations in peak values, normalization has been done. Normalized $p_k(l)$ is given by $P_k(l)$,

$$P_k(l) = \frac{p_k(l)}{max(p_k(l))} \quad (16)$$

However, it is seen that the Apnea attributes in $P_k(l)$ are a small fraction of unity. In the proposed method, the relative emphasis of these small fraction coefficients is increased by adopting a logarithm approach. We convert $P_k(l)$ to their respective decibel value, $10log_{10}P_k(l)$. By doing so, the peak value of unity becomes $0dB$ while all other smaller fractions are in a range from $-20dB$ to $-80dB$ whose range is far narrower than their linear counterpart in $P_k(l)$. Now an average of this set is taken - an average that is no longer skewed to unity if it was taken without logarithm.

Let this average be given by $\overline{P}_k$, and defined as:

$$\overline{P}_k = \frac{\sum_{n=1}^{l} 10\log_{10}p(n)}{l} \quad (17)$$

where k denotes the kth block. We assume that $\overline{P}_k$ carries the Apnea attributes and can act as the suitable DV. As mentioned in section II-B, if the underlying statistics of this DV are Gaussian under two cases of normal and Apnea, it can be usefully used for detecting Apnea. In order to validate this assumption, we now undertake a histogram analysis of $\overline{P}_k$ in two cases separately - one where $\overline{P}_k$ is for normal cases and another when it is for Apnea cases. $\overline{P}_k$ is the output of the technique block - the decision variable, d_k, of section II-B. In the next section, $\overline{P}_k$ would be used to represent the decision variable, instead of the assumed DV, d_k.

3) THE DECISION PHASE

The role of this phase is to provide robust and reliable detection based on the decision variable ($\overline{P}_k$) provided by the developed technique. PDFs of $\overline{P}_k$ have been determined, in both the Apnea and normal cases. It was assumed in section II-B that if the PDF were Gaussian, the P_{fa} and P_{md} could be easily determined. In this section, we will verify whether the $\overline{P}_k$ is indeed Gaussian or not. If not, whether it can still be used in the same manner. Towards this end, we take the ECG data from MIT online Sleep Apnea database. In order to study the underline statistics and features, we are deploying data set A_l (as discussed in section II-A that contains data from 17 patients with no, mild or severe Apnea).

In order to investigate the statistical trends followed by $\overline{P}_k$, a histogram is constructed using the value of $\overline{P}_k$ from data in A_l. Later the threshold is evaluated and Th, Gaussian curve fitting has been done on the histograms, as shown in Fig. 6. MATLAB 2009a [17], curve fitting tool (cftool) has been deployed to curve fit the histogram.

Individual characteristic equations have been obtained, for both cases. Eq. 18 represents the curve-fit equation for the normal case.

$$f_n(\alpha) = \beta_1 \times e^{(-((\alpha-\gamma_1)/\zeta_1)^2)} + \beta_2 \times e^{(-((\alpha-\gamma_2)/\zeta_2)^2)}, \quad (18)$$

where α, γ, ζ, and β denotes $\overline{P}_k$, mean, variance and relative strength of each curve in the Gaussian mixture.

The available range of the coefficients for curve fitting is obtained whereby the 95% confidence bound values

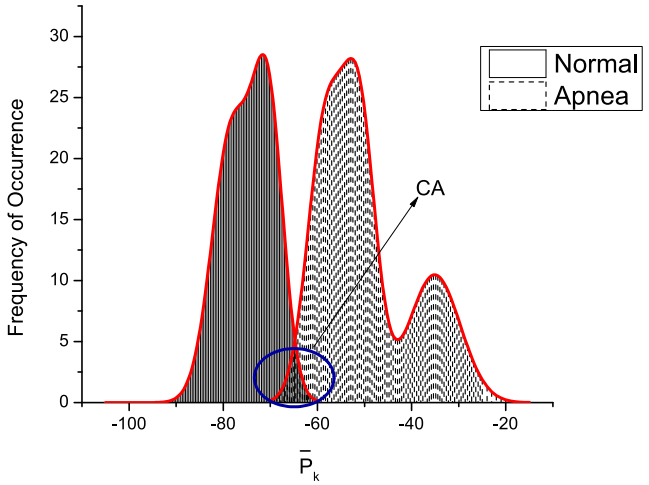

FIGURE 6. Histogram of $\bar{P}_k$ values in Apnea and normal cases for A_l along with Gaussian Fit.

are utilized. For the normal case, a summary of the available and chosen values of coefficients is mentioned in Table 1.

TABLE 1. Variable values for Normal case for 50% learning data.

Variable	Chosen Value	Range
$\beta 1$	23.94	(22.68, 25.2)
$\gamma 1$	-70.53	(-70.73, -70.32)
$\zeta 1$	4.613	(4.451, 4.776)
$\beta 2$	21.89	(21.22, 22.57)
$\gamma 2$	-78.3	(-78.62,77.98)
$\zeta 2$	5.831	(5.557, 6.106)

TABLE 2. Variable values for Apnea (Abnormal) case for 50% learning data.

Variable	Chosen Value	Range
a1	23.47	(20.97, 25.96)
b1	-51.22	(-51.74,-50.7)
c1	4.932	(4.533, 5.33)
a2	21.91	(19.74, 24.08)
b2	-58.44	(-59.08,-57.81)
c2	5.27	(4.832, 5.708)
a3	10.47	(10.13, 10.81)
b3	-35.11	(-35.34,-34.89)
c3	7.715	(7.326, 8.103)

Eq. 19 represents the curve-fit equation for Apnea. For the Apnea case, a summary of the available and chosen values of coefficients, of Eq. 19, is mentioned in Table 2.

$$f_a(\alpha) = a_1 \times e^{(-((\alpha-b_1)/c_1)^2)} + a_2 \times e^{(-((\alpha-b_2)/c_2)^2)} + a_3 \times e^{(-((\alpha-b_3)/c_3)^2)} \quad (19)$$

where α, b, c, and a denotes $\bar{P}_k$, mean, variance and relative strength of each curve in the Gaussian mixture.

From Fig. 6, it is clear that although both the curves are not pure Gaussian however, instead superposition of Gaussian functions. However, their tails in the confusion area are quite close to the Gaussian tails of Fig. 2. In the confusion area, an interesting phenomenon is observable whereby, we notice that for normal curve sometimes,

$$|\alpha - \gamma_1| < |\alpha - \gamma_2| \quad (20)$$

Because of this relationship, $e^{-(\alpha-\gamma_2)/\zeta_2}$ in Eq. 18 would be increased incredibly as compared to $e^{-(\alpha-\gamma_1)/\zeta_1}$, hence, in confusion area, Eq. 18 reduces to,

$$f_n(\alpha) = \beta_2 \times e^{(-((\alpha-\gamma_2)/\zeta_2)^2)} \quad (21)$$

While depending upon the values of γ_1 and γ_2 sometimes,

$$|\alpha - \gamma_2| < |\alpha - \gamma_1| \quad (22)$$

Hence, at some α, either of the curve would be negligible reducing the Eq. 18 to only one Gaussian curve equation. Lets assume i denotes the curve with potentially higher value, hence,

$$f_n(\alpha) = \beta_i \times e^{(-((\alpha-\gamma_i)/\zeta_i)^2)} \quad (23)$$

On the similar grounds, the reduced equation in the confusion area for Apnea curve can be written as:

$$f_a(\alpha) = a_i \times e^{(-((\alpha-b_{i1})/c_i)^2)} \quad (24)$$

From Eq. 23 and 24 it is obvious that in the confusion area there exist only two Gaussian curves. Accordingly, the maximum likelihood criteria can be used to maximize the detection of Apnea. The optimum values of P_{fa} and P_{md} are thus uniquely obtained using the threshold, Th, as discussed in the next section. In order to achieve the optimum threshold, a new function $H(\alpha)$ has been defined,

$$H(\alpha) = f_a(\alpha) - f_n(\alpha). \quad (25)$$

The optimum threshold that maximizes the detection of Apnea is the intersection point of $f_a(\alpha)$ and $f_n(\alpha)$ i.e., when $\alpha = Th$, $H(Th) = 0$. Th has been obtained using the well-known Newton Raphson's Method [9] with initial value $\alpha = -70$. By this method $th = -64.82$. Accuracy, specificity, and sensitivity have been calculated for A_l.

D. DATA SIZE OPTIMIZATION TEST

The ratio selection of learning and testing database is crucial. Our aim is to extract the optimum threshold in order to achieve the statistically sufficient size of the learning set. An in-depth study has been done using different ratios of A_l. This test provided an insight to the statistical properties of the dataset A. The change in the value of threshold when the percentage of learning data is changed, is summarized in Fig. 7. The threshold value becomes stable when more than 40% of data is used in the learning set, as depicted in Fig. 7. However, it has been observed that the threshold value fluctuates only a little from the learning data size variation of 10% to 50%. It is because the overlapping area between the two curves is only about 1% of the total area. Hence, our method is dynamic and robust enough to tolerate the small changes in threshold level caused by different ratios of training and testing data.

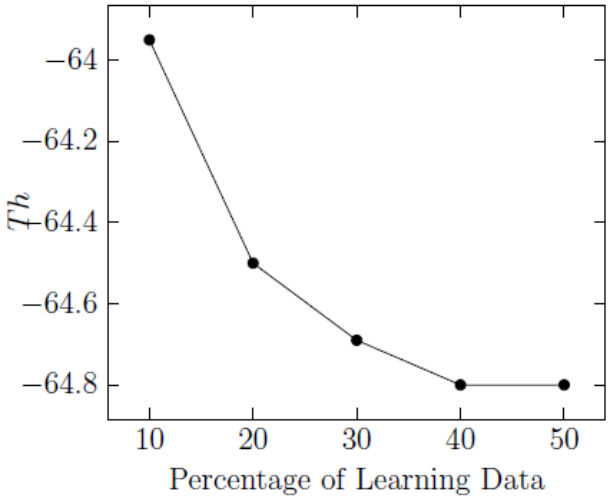

FIGURE 7. Value of threshold vs learning data size.

However, in this particular research we have deployed 50% to 50% data division between A_l and A_t.

Moreover, it is interesting to note that the adjusted R-square (goodness of fit) becomes closer to 1 when the data size is increased from 10% to 40% showing that the fit is a good representation of the statistics of the data. Similar trends can be observed in R-square and RMSE (Root Mean Square Deviation) values from 10% to 40% of data division between learning and testing set when compared with their definitions.

E. CONFIDENCE BOUND TEST

Confidence bound test was performed in order to assure that the developed scheme is not dependent on the patients data being chosen for learning or testing; rather the evaluated threshold has negligible fluctuations when the data sets are changed. For example, in one iteration patient 3 data was chosen for learning while in the second iteration it was dropped and patient 5 was selected. Similarly, about 10 iterations were performed by choosing and dropping different data sequences. The Standard deviation of the thresholds has been evaluated.

F. PERFORMANCE EVALUATION

The performance of the developed technique is evaluated using P_{fa}, P_{md}, P_d, P_e, accuracy, sensitivity and specificity. These measures were evaluated using confusion matrix parameters, as mentioned in Table 3.

TABLE 3. Confusion matrix.

	Predicted Positive (P)	Predicted Negative (N)
Positive	TP	FN
Negative	FP	TN
Positive=SA	Negative=Normal	F=False, T=True

The probability of false alarm can be evaluated using the parameters of confusion matrix as follows.

$$P_{fa} = \lim_{n\to\infty} \frac{FP}{TN}, \tag{26}$$

where n denotes the number of experiments. This equation represents the ratio of false alarm with respect to the correctly determined negative cases.

The probability of misdetection can be evaluated using the parameters of confusion matrix as follows.

$$P_{md} = \lim_{n\to\infty} \frac{FN}{TP}, \tag{27}$$

where n denotes number of experiments. This equation represents the ratio of misdetection with respect to the correctly determined positive cases.

The error introduced in our case is either a false alarm or a misdetection, hence the net probability of error would be the sum the probabilities of these two entities. Mathematically,

$$P_e = P_{fa} + P_{md}. \tag{28}$$

The probability of error-free detection or simply the probability of detection can be determined as:

$$P_d = 1 - P_e. \tag{29}$$

The specificity indicates the ability of a classifier to detect negative cases, i.e. normal cases. It is calculated using Eq. 30.

$$Specificity = \frac{TN}{(TN + FP)} \times 100\% \tag{30}$$

The sensitivity represents the ability of a classifier to detect the positive cases, i.e. *SA* cases. It is calculated using Eq. 31.

$$Sensitivity = \frac{TP}{(TP + FN)} \times 100\% \tag{31}$$

The accuracy represents the overall performance of a classifier. It indicates the percentage of correctly classified positive and negative cases from the total number of cases. It is calculated using Eq. 32.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100\% \tag{32}$$

III. RESULTS AND DISCUSSION

This section presents the results obtained from the developed scheme along with the associated discussion. In order to test the developed method for block-based detection, ECG data stream is divided into 5 minutes data blocks. The status of each minute (as annotated by human experts in the database) is checked and if there is a single Apnea minute in a block the whole block is marked as Apnea. An annotation array, Arr_a, is created for storing the status of all the blocks.

The developed scheme is tested using data set A_t. There are total18 cases in A_t, out of these 18 cases, 10 are abnormal, 5 are normal, and 3 are borderline. One by one, ECG signal from each case is taken that later passed through our model, as described in section II-C and final decision for each block is determined as either normal or abnormal. A decision

array Arr_d stores the status for each block as decided by the developed scheme. At the end, a comparison is made between Arr_a and Arr_d in order to acquire the accuracy status of the developed scheme. The performance has been evaluated using probability measures along with accuracy, specificity, and sensitivity, as discussed in section II-F. The results obtained by these steps are summarized in Table 4.

TABLE 4. Block based detection testing.

Patient	Total Minutes	Normal Blocks	Apnea Blocks	Detected Blocks	TP	FP	TN	FN
a12	578	8	107	107	107	0	8	0
a13	496	50	49	50	49	1	49	1
a14	510	25	77	78	77	0	25	1
a15	511	28	74	74	74	1	27	0
a16	483	32	64	64	64	0	32	0
a17	486	65	32	33	32	2	63	1
a18	490	52	88	90	88	0	52	2
a19	503	59	41	41	41	2	57	0
a20	511	39	63	63	63	0	39	0
b01	488	93	4	6	4	3	90	2
b02	518	85	19	19	19	1	84	0
b05	434	75	12	12	13	0	75	1
c03	455	91	0	0	0	0	91	0
c04	483	95	0	1	0	0	95	0
c06	469	92	1	0	0	0	92	1
c07	454	89	1	1	1	1	89	0
c08	535	107	0	0	0	0	107	0

Afterwards, for the validation of the consistency of the results obtained through this technique, it was tested for a larger database. A similar procedure was performed for the dataset A_v that contains a total of 35 cases, out of which 20 are Apnea, 10 are normal, and 5 are borderline (summarized in Table 5). The performance was also evaluated by comparing the obtained results with the results of the testing data.

TABLE 5. Block based detection (validation).

Case	Net Minutes	Normal Blocks	Sleep Apnea Blocks	Detected Block	TP	FP	TN	FN
x01	524	29	75	76	75	0	29	1
x02	470	52	42	41	41	1	51	1
x03	466	90	2	2	2	1	89	0
x05	506	38	63	62	62	0	38	1
x07	510	54	48	47	47	1	53	1
x08	518	38	65	63	63	0	38	2
x09	509	68	33	33	33	0	68	0
x10	511	83	20	19	19	2	81	1
x11	458	89	2	2	2	1	87	0
x12	528	94	12	12	12	1	93	0
x13	507	43	58	57	57	0	43	1
x14	491	10	88	90	88	0	10	0
x15	499	59	40	39	39	0	59	1
x16	516	90	13	12	12	0	90	1
x19	488	16	81	81	81	1	16	0
x20	514	50	53	53	53	1	49	0
x21	511	81	24	23	23	0	81	1
x23	528	81	24	23	23	0	81	1
x25	511	44	59	57	57	0	44	2
x26	521	35	69	69	69	1	34	0
x27	499	1	98	96	96	1	0	2
x28	496	12	87	85	85	0	12	2
x30	512	37	65	64	64	0	37	1
x31	558	8	103	101	101	0	8	2
x32	539	22	85	83	83	1	21	2
x04	483	96	0	0	0	0	97	0
x06	451	90	0	0	0	0	90	0
x17	401	80	0	0	0	0	80	0
x18	460	92	0	0	0	0	92	0
x22	483	96	0	0	0	0	96	0
x24	430	86	0	0	0	0	86	0
x29	471	94	0	0	0	0	94	0
x33	474	94	0	1	0	1	95	0
x34	476	94	0	1	0	1	94	0
x35	484	96	0	0	0	0	97	0

1) DISCUSSION

It is worth mentioning that the basic idea of dividing the data (in five minutes blocks) is to reduce the number of cycles to determine the nature of data being normal or abnormal. By this technique, instead of 5 iterations to determine the data status of 5 minutes interval, only 1 iteration is required, which makes our technique memory efficient and fast. If there is only a single event of Apnea, the whole set is considered to be abnormal which helps in cases based classification where AASM 1999 [20] criteria is deployed to segregate normal and abnormal cases.

Careful study of the MIT BIH Sleep Apnea database has reveled that in most cases, Apnea events occur in a continuous pattern where there are Apnea (annotated) minutes from 10 contiguous minutes to an hour or more. However, it is also noticed that in some scenarios there is a stand alone minute of annotated Apnea in a whole series of normal minutes. In some cases, where the pattern of previous and following normal minutes is so strong as compared to a sudden and little event of Apnea, this Apnea event remains hidden because of very less activity in F_bOI and the block is classified as normal which in turns reduces the sensitivity of the block-based classification. On the other hand, the situations where the person is in the recovery phase after a severe Apnea episode, the following normal minutes after the long lasting Apnea episode are highly under the influence of Apnea minutes. Although, these minutes have been annotated as normal considering the breathing pattern, but, they do have some traces of the passed Apnea attack that sometimes forces our algorithm for a false alarm of Apnea. In this way, some of the normal blocks that follow intense Apnea episode are falsely regarded as Apnea reducing the specificity of block detection scheme. However, since the goal of this research is to diagnose Apnea cases, it is crucial to note if the accuracy of case diagnosis (next experiment) gets affected by this or not. If it gets affected, it is also important to note the extent of incorrect results because of the slightly reduced accuracy of block detection.

Since the primary goal of this research is to segregate normal and abnormal cases, the block detection results, discussed above, provides the platform for overall diagnosis. According to AASM 1999 criteria [20], a patient is considered to be a Sleep Apnea patient if he/she has 5 or more episodes of Sleep Apnea in an hour. The decision record of each block is kept in history and the number of blocks

flagged as Apnea are counted on per hour basis. If there are more than or equal to 5 Apnea blocks in an hour, the patient is classified as Apnea patient [20]. Moreover, if there are less than 5 blocks in an hour that shows presence of Apnea, the patient is flagged as borderline. In this procedure, flag history of the blocks in A_t and A_v served as the input for testing and validation of case diagnosis respectively. The performance has been evaluated using probability measures along with accuracy, specificity and sensitivity, as discussed in section II-F. Table 6 encapsulates the results obtained for testing scheme as stated above It also highlights the human annotations as a comparison of our achieved results. Similarly, results were achieved for A_v, where in 6/35 cases, a block faced misdetection (as summarized in Table 7). However, the overall diagnosis of the case did not suffer because of this and the case was diagnosed correctly.

TABLE 6. Results of case diagnosis (testing).

Record	Total Hours	Apnea Hours	Detected Hours	Actual Result	Our Result	Accuracy (hours)	Accuracy (case)
a12	9.6333	10	9	A	A	90	100
a13	8.2667	9	8	A	A	88.8889	100
a14	8.5	8	8	A	A	100	100
a15	8.5167	9	9	A	A	100	100
a16	8.05	7	7	A	A	100	100
a17	8.1	5	5	A	A	100	100
a18	8.1667	9	8	A	A	88.8889	100
a19	8.3833	9	9	A	A	100	100
a20	8.5167	9	9	A	A	100	100
b01	8.1333	2	2	B	B	100	100
b02	8.6333	5	4	B	B	80	100
b05	7.2333	3	3	B	B	100	100
c03	7.5833	0	0	C	C	100	100
c04	8.05	0	0	C	C	100	100
c06	7.8167	0	0	C	C	100	100
c07	7.5667	0	0	C	C	100	100
c08	8.9167	0	0	C	C	100	100

All of the decisions of our classifier are in 100% agreement with the results provided by human experts for each case. Although our scheme does not yield 100% accuracy for block based detection experiment, yet it does so for the cases diagnosis. The reason lies in the fact that we have divided data into 5 minutes blocks, which means there are 12 blocks in an hour. For instance, if one block gives a miss detection or a false alarm yet there are 11 more blocks to cover the mistake. Moreover, we are using whole night data averaged 8 hours of recording, hence in a very complex situation (normally the borderline case) where because of a single miss detection our count falls to 4 Apnea blocks in one hour where it should have been 5 Apnea blocks per hour, we still have the chance to detect Sleep Apnea in the rest of 8 hours recording. This leads to correct diagnosis of normal and abnormal cases yielding 100% accuracy.

However, these are only the worst case scenarios, or more precisely it happens only in borderline cases where the patient is only suspected of Apnea. An Apnea patient, as mentioned earlier, usually has continuous disruption of respiration yielding a lot of Apnea events. Moreover, the 98% classification accuracy in block based detection scheme is quite high to help in the correct classification at the overall case level. The tables in this section highlight the values of performance evaluation measures evaluated from the tables presented in section III (with the help of the definitions explained in section II-F). The value of P_{fa} is decreased while calculated for A_v. It is because of the fact that A_v is a larger database and so owns more number of normal blocks than that of A_t. Hence TN cases increase, subsequently decreasing the P_{fa}.

TABLE 7. Results of case diagnosis (testing).

Record	Total Hours	Actual Apnea Hour	Detected Apnea	Actual Result	Our Result	Accuracy for Hours	Accuracy for Case Detection
x01	8.7333	9	8	A	A	88.8889	100
x02	7.8333	7	7	A	A	100	100
x03	7.7667	1	1	B	B	100	100
x05	8.4333	9	8	A	A	88.8889	100
x07	8.5	8	7	A	A	87.5	100
x08	8.6333	8	8	A	A	100	100
x09	8.4833	5	5	A	A	100	100
x10	8.5167	6	6	A	A	100	100
x11	7.6333	1	1	B	B	100	100
x12	8.8	4	4	B	B	100	100
x13	8.45	8	8	A	A	100	100
x14	8.1833	9	9	A	A	100	100
x15	8.3167	8	8	A	A	100	100
x16	8.6	4	4	A	A	100	100
x19	8.1333	9	8	A	A	88.8889	100
x20	8.5667	8	8	A	A	100	100
x21	8.5167	4	4	B	B	100	100
x23	8.8	3	3	B	B	100	100
x25	8.5167	9	8	A	A	88.8889	100
x26	8.6833	9	9	A	A	100	100
x27	8.3167	9	8	A	A	88.8889	100
x28	8.2667	9	9	A	A	100	100
x30	8.5333	9	9	A	A	100	100
x31	9.3	10	9	A	A	90	100
x32	8.9833	9	9	A	A	100	100
x04	8.05	0	0	N	N	100	100
x06	7.5167	0	0	N	N	100	100
x17	6.6833	0	0	N	N	100	100
x18	7.6667	0	0	N	N	100	100
x22	8.05	0	0	N	N	100	100
x24	7.1667	0	0	N	N	100	100
x29	7.85	0	0	N	N	100	100
x33	7.9	0	0	N	N	100	100
x34	7.9333	0	0	N	N	100	100
x35	8.0667	0	0	N	N	100	100

TABLE 8. Cases based performance evaluation using A_t.

Evaluation Measure	Value
Specificity	5/5
Sensitivity	12/12
Accuracy	18/18

TABLE 9. Cases based performance evaluation using A_v.

Evaluation Measure	Value
Specificity	10/10
Sensitivity	25/25
Accuracy	35/35

TABLE 10. Block based performance evaluation using A_v.

Evaluation Measure	Value
Specificity	99.08%
Sensitivity	98.59%
Accuracy	98.89%
P_{fa}	0.006
P_{md}	0.017
P_e	0.023
P_d	0.977

TABLE 11. Block based performance evaluation using A_t.

Evaluation Measure	Value
Specificity	99.34%
Sensitivity	98.24%
Accuracy	98.93%
P_{fa}	0.01
P_{md}	0.014
P_e	0.024
P_d	0.976

IV. CONCLUSIONS

The main target of this study was the development of an ECG based Sleep Apnea detection algorithm that is fast and simple; which may contribute towards easy and early diagnosis of sleep apnea. Most of the available schemes mentioned in the literature are complicated either because of the use of more discriminatory features or deployment of more than one technique.

The scheme proposed in this research focuses on the least number of features to be incorporated. Only the RRI time-series, derived from the ECG signal, has been used and a single decision variable $\overline{P}_k$ is selected. This unitary signal and discriminatory variable selection have made the algorithm quite simple.

Moreover, during the wavelet analysis, only F_bOI is focused and till the achievement of F_bOI, detailed components have been discarded, hence conserving the memory. This technique not only helped in reducing the memory consumption, but it also increased the accuracy as precisely the activity during Apnea event is focused and redundant detail is discarded. Hence, all the analysis iteration process is performed on a small frequency band while neglecting the rest of the spectrum.

As summarized in section III-.1, the proposed scheme segregate Apnea patients from normal people with 100% accuracy. The scheme has first been tested on 18 different data sets followed by blind validation of 35 more cases.

Chazel *et al.* [7], have deployed multiple techniques while using both the RRI and EDR (ECG Derived Respiration) signals, in order to achieve 100% accurate classification of Sleep Apnea patients. In contrast, our method used only the RRI feature of ECG and used single DWPT based technique to achieve the same accuracy. In 2008, Khandoker *et al.*, [13] had discovered a method for 100% classification accuracy of Sleep Apnea patients using Wavelet transform. But this technique also deploys two features of ECG signal, i.e., RRI and EDR. Study of two features needs more computation as compared to our technique where the analysis of only one feature is required. Using Spectrogram, McNames and Fraser [18], have achieved 100% accuracy but they have also used two features, i.e., RRI and S-amplitude of ECG wave. Jarvis and Mitra [12], have presented a spectrogram based solution, but the threshold of the technique is not standard. Every time, the entire data has to be analyzed, a new threshold is produced and in the end a decision can only be made after the analysis of whole data once again. On the other hand, we have trained our classifier based on the statistics of learning data and a robust standard threshold is achieved. Our proposed scheme makes decisions based on five minutes data blocks and after the analysis of 12 blocks (1 hour) a tentative decision can be made that may further be strengthened by studying more blocks.

In a nutshell, a comprehensive statistical study has been performed using the learning database. The best fit applied to the probability density function validates that a simple binary maximum likelihood detection can be used for the purpose. Accordingly, a unique threshold is obtained that minimizes the decision error. Based on the threshold, Sleep Apnea is diagnosed in each block of every patient's record in the testing database. Overall classification of the patients is performed by using the AASM 1999 criteria on the diagnosed blocks-AASM 1999 criteria says that a person is said to suffer from Sleep Apnea if he\she has 5 or more episodes of Sleep Apnea per hour in whole night recording. Validation has been performed by using a richer database containing 35 records. The decisions of our scheme are compared with the human annotations provided at the MIT-BIH online database. Validation process revealed that this scheme has the capability to segregate the Apnea and normal cases accurately. Hence, the technique serves to minimize the chance of undiagnosed Sleep Apnea with 100% accuracy while being very simple and time efficient as compared to the traditional methods.

The developed scheme may be employed to make a commercial device for Sleep Apnea detection that will take digitized ECG signal and will display the status of the patient as suffering from apnea or normal. The hassles of Sleep Apnea detections will lessen along with associated costs. Moreover, this disease will no more remain a silent killer after the availability of a small and cheap handheld Sleep Apnea detector.

REFERENCES

[1] *MIT-BIH online database Sleep Apnea Challenge 2000*, 2000.
[2] P. S. Addisson, "*The Illustrated Wavelet Transform Handbook*. New York, NY, USA: Taylor & Francis, 2002.
[3] L. Almazaydeh, M. Faezipour, and K. Elleithy, "A neural network system for detection of obstructive sleep apnea through SpO2 signal features," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 5, pp. 7–11, May 2012.
[4] M. R. Azim, S. A. Haque, M. S. Amin, and T. Latif, "Analysis of EEG and EMG signals for detection of Sleep Disordered Breathing events," in *Proc. Int. Conf. Elect. Comput. Eng. (ICECE)*, Dhaka, Bangladesh, Dec. 2010, pp. 646–649.
[5] M. Ballora, B. Pennycook, P. C. Ivanov, A. Goldberger, and L. Glass, "Detection of obstructive sleep apnea through auditory display of heart rate variability," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 739–740.

[6] R. D. Cartwright, *Treating Sleep Disorders: Principles and Practice of Behavioral Sleep Medicine*. Hoboken, NJ, USA: Wiley, 2003.
[7] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley "Automatic classification of sleep apnea epochs using the electrocardiogram," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 745–748.
[8] M. J. Drinnan, J. Allen, P. Langley, and A. Murray, "Detection of sleep apnoea from frequency analysis of heart rate variability," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 259–262.
[9] H. F. Begnaud, *Introduction to Numerical Analysis*. Chelmsford, MA, USA: Courier Corporation, 1987.
[10] M. D. V. Hoffstein and S. Linde, *No More Snoring*. Hoboken, NJ, USA: Wiley, 1999.
[11] A. Hossen, "Identification of obstructive sleep apnea from normal subjects: Fft approaches wavelets," *Int. J. Biometrics Bioinf.*, vol. 4, no. 2, p. 22, Jun. 2010.
[12] M. R. Jarvis and P. P. Mitra, "Apnea patients characterized by 0.02 Hz peak in the multitaper spectrogram of electrocardiogram signals," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 769–772.
[13] A. H. Khandoker, C. K. Karmakar, and M. Palaniswami, "Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings, "*Comput. Biol. Med.*, vol. 39, no. 1, pp. 88–96, Jan. 2009.
[14] M. H. Kryger, *A Women's Guide to Sleep Disorders*. New York, NY, USA: McGraw-Hill, 2004.
[15] D. Liu, Z. Pang, and S. R. Lloyd, "A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG," *IEEE Trans. Neural Netw.*, vol. 19, no. 2, pp. 308–318, Feb. 2008.
[16] C. Maier, M. Bauch, and H. Dickhaus, "Recognition and quantification of sleep apnea by analysis of heart rate variability parameters," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 741–744.
[17] *MATLAB, version R2009a*, Natick, MA, USA, MathWorks, 2009.
[18] J. N. McNames and A. Fraser, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram," in *Proc. Comput. Cardiol.*, vol. 2, pp. 749–752, Sep. 2000.
[19] J. E. Mietus, C. K. Peng, P. C. Ivanov, and A. L. Goldberger, "Detection of obstructive sleep apnea from cardiac interbeat interval time series," in *Proc. Comput. Cardiol.*, vol. 27, pp. 753–756, Sep. 2000.
[20] American Academy of Sleep Medicine Task Force, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research, "*Sleep*, vol. 2, no. 5, pp. 667–689, Aug. 1999.
[21] N. Oliver and F. Flores-Mangas, "Automatic sleep apnea detection and monitoring with a mobile phone," *J. Commun.*, vol. 2, no. 2, pp. 1–9, Mar. 2007.
[22] P. K. Stein and P. P. Domitrovich, "Detecting OSAHS from patterns seen on heart-rate tachograms," in *Proc. Comput. Cardiol.*, vol. 27, Sep. 2000, pp. 271–274.
[23] C. W. Zywietz, V. V. Einem, B. Widiger, and G. Joseph, "Sleep apnea detection in single channel ecgs by analyzing heart rate dynamics," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Istanbul, Turkey, Oct. 2001, pp. 1816–1819.

SYEDA QURATULAIN ALI was born in Karachi, Pakistan, in 1987. She received the bachelor's degree in electrical engineering (telecommunications) from CIIT, Islamabad, and the master's degree in electrical engineering (by research) from Universiti Teknologi Petronas, Malaysia. She is ambitious to do her Ph.D. in the field of biomedical signal processing. She is currently a home-based freelance researcher. Her research interests include exploring the biomedical signals and images by using signal processing techniques.

SOHAIL KHALID (M) was born in Kharian, Pakistan, in 1984. He received the B.Eng. degree (Hons.) from CIIT, Islamabad, in 2008, the M.Sc. degree in wireless networks from the Queen Mary University of London, in 2009, and the Ph.D. degree from Universiti Teknologi Petronas, Malaysia, in 2014. He is currently serving as an Associate Professor and the Head of the Electrical Engineering Department, Riphah International University, Islamabad, Pakistan, where he is teaching various under graduate and postgraduate engineering courses. His research interests include synthesis and design of passive microwave devices, millimeter wave applications for biomedical engineering, and biomedical signal processing.

SAMIR BRAHIM BELHAOUARI received the master's degree in telecommunications from the Institute National Polytechnic (INP), Toulouse, France, in 2000, and the Ph.D. degree from the Federal Polytechnic School of Lausanne (EPFL), Switzerland, in 2006. He has been an Assistant Professor since 2008. His research interest includes applied and pure mathematics related to engineer field, specifically on image/signal processing and stochastic processes.