

# Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG

Majdi Bsoul, *Member, IEEE*, Hlaing Minn, *Senior Member, IEEE*, and Lakshman Tamil, *Senior Member, IEEE*

**Abstract**—We have developed a low-cost, real-time sleep apnea monitoring system “Apnea MedAssist” for recognizing obstructive sleep apnea episodes with a high degree of accuracy for both home and clinical care applications. The fully automated system uses patient’s single channel nocturnal ECG to extract feature sets, and uses the support vector classifier (SVC) to detect apnea episodes. “Apnea MedAssist” is implemented on Android operating system (OS) based smartphones, uses either the general adult subject-independent SVC model or subject-dependent SVC model, and achieves a classification  $F$ -measure of 90% and a sensitivity of 96% for the subject-independent SVC. The real-time capability comes from the use of 1-min segments of ECG epochs for feature extraction and classification. The reduced complexity of “Apnea MedAssist” comes from efficient optimization of the ECG processing, and use of techniques to reduce SVC model complexity by reducing the dimension of feature set from ECG and ECG-derived respiration signals and by reducing the number of support vectors.

**Index Terms**—Apnea monitor, ECG, home care, smartphone, support vector machines (SVMs).

## I. INTRODUCTION

**A**PNEA is a sleep related breathing disorder—commonly known as obstructive sleep apnea (OSA) is a common disorder that affects about 4% of the general population. People with sleep apnea literally stop breathing repeatedly during their sleep, often for 10–30 s and as many as hundreds of times during one night. Sleep apnea can be caused by complete “apnea” or partial “hypopnea” obstruction of airway [2], both of which can wake one up. The frequent arousals and the inability to achieve or maintain the deeper stages of sleep can lead to excessive daytime sleepiness, nonrestorative sleep, automobile accidents, personality changes, decreased memory, erectile dysfunction (impotence), and depression. OSA has also been linked to angina, nocturnal cardiac arrhythmias, myocardial infarction, and stroke.

The primary method for diagnosing OSA at present is to have the patient undergo a sleep study, known as polysomnography (PSG). A polysomnogram typically records a minimum of eleven channels of various biosignals requiring a minimum

of 22 wire attachments to the patient in a specialized sleep laboratory with attended personnel. Obstructive sleep apnea is diagnosed, if the patient has an apnea index (AI) (apneic episodes per hour) greater than 5/h, or a respiratory disturbance index, the combination of apneas and hypopneas, greater than 10/h. Several treatment options exist for OSA. These include weight reduction, oral appliances, positional therapy, continuous positive airway pressure (CPAP) therapy, and surgical options. CPAP, the most common of these therapies is usually administered at bedtime through a nasal or facial mask held in place by velcro straps around the patients head [3].

Our research study aims to develop a simpler system that provides a reliable, inexpensive, and faster approach to assessing OSA in patients before, during and after medical treatments. The measurement of sleep quality and/or sleep apnea becomes necessary when certain medical therapies and drugs are administered, and there is a need to assess the side effects of these treatments, which may manifest as sleep or breathing disturbances. The real-time OSA detection is critical in perioperative monitoring, which includes assessing the presence and severity of sleep apnea to avoid complications during and after the surgery [4]. Also, real-time apnea monitoring can provide instantaneous results for any associated medical treatment, such as feedback for CPAP pressure adjustments.

Using only single-channel ECG measurements reduces the complexity of the diagnostic test and enables patients to better monitor physiological changes corresponding to changes in sleep apnea severity. The developed apnea smartmonitor uses a wireless smartphone to realize a cost-effective platform for ECG acquisition, monitoring, and real-time screening and assessment of sleep apnea syndrome. Support vector machine (SVM) [5], [6] is a powerful discriminative method for pattern classification. Its basic idea is to map data into a high-dimensional space and find a separating hyperplane with maximal margin. We developed several SVM classifier (SVC) models to automatically classify apnea episodes according to patient criteria—the existence of prior annotated sleep apnea data for the subject or for another subject matching his physical attributes—and/or device connectivity and processing capabilities. In this study, we also consider a set of features either extracted directly from heart rate variability (HRV) RR-tachogram or from the surrogate ECG-derived respiration signal (EDR) [7]; a total of 111 features (time and spectral domain) from every 1-min epoch. We also apply feature reduction algorithms to find an optimized input set to the SVM classifier models.

This paper is organized as follows. Section II reviews related previous studies. Section III gives an overview of the developed smart monitor (Apnea MedAssist) system architecture.

Manuscript received May 9, 2010; revised August 24, 2010; accepted October 3, 2010. Date of publication October 14, 2010; date of current version May 4, 2011.

M. Bsoul is with the Alcatel-Lucent, Plano, TX 75075 USA, and also with the Quality of Life Technology Laboratory, Richardson, TX 75080 USA (e-mail: majdi@ieee.org).

H. Minn and L. Tamil are with the Department of Electrical Engineering, Quality of Life Technology Laboratory, University of Texas at Dallas, Richardson, TX 75083 USA (e-mail: hlaing.minn@utdallas.edu; laxman@utdallas.edu).

Digital Object Identifier 10.1109/TITB.2010.2087386

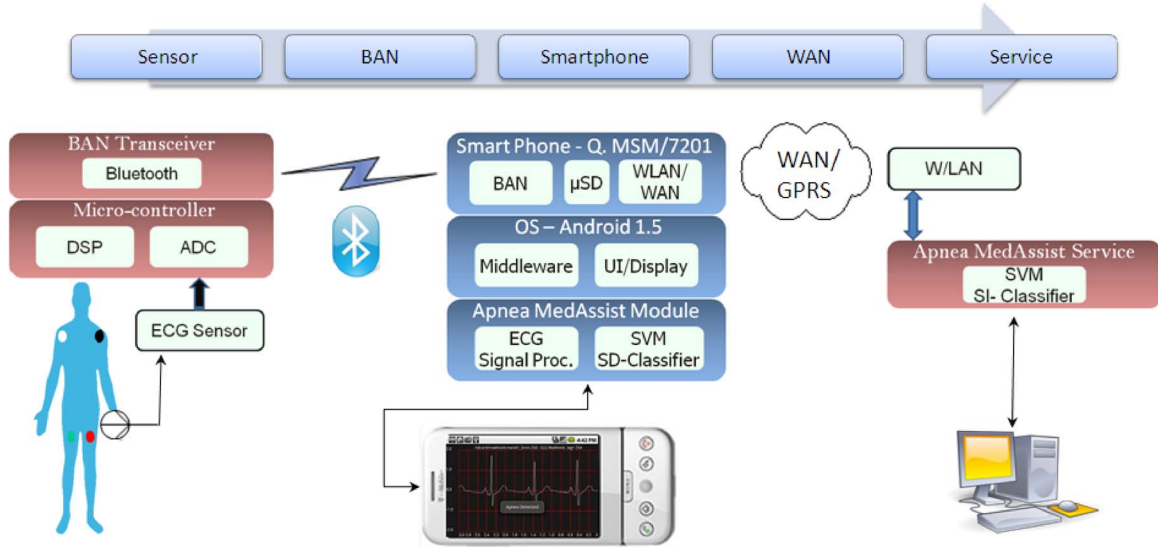


Fig. 1. System architecture of Apnea MedAssist showing components and functionalities.

Section IV details the methodology of this paper including the database used for testing, ECG preprocessing, RR and EDR feature sets, and the SVC models used. In Section V, we provide the results. In Section VI we discuss results of the automated recognition, and in Section VII we give our conclusion.

## II. PREVIOUS WORK

There is a significant body of research literature in the study of simplified sleep apnea monitoring using just one or two physiological signals such as ECG, pulse oximetry, snoring, or nasal airflow [8]–[13]. Oliver and Flores-Mangas [8] used blood oxygen level ( $\text{SpO}_2$ ) for screening OSA. Fu-Chung *et al.* [9] used integrated CPAP airflow signal to identify OSA episodes. Pentagay *et al.* [11] used the heart sound (S1) generated during OSA episodes combined with ECG. The application of SVM classifiers in apnea screening are discussed in [10] and [11]. Other OSA screening studies based on ECG used Gaussian classifiers, linear or quadratic discriminants [12], [13], which depend on the assumption that the feature has Gaussian distribution.

The following summarizes the main contributions of our paper and differentiates it from the aforementioned studies:

- 1) Apnea screening done on 1-min segment of data rather than treating the whole recording as one segment. This measurement segment length is chosen to reduce the hidden number of actual episodes within the selected time interval.
- 2) Real-time screening as opposed to offline screening.
- 3) Apnea severity or apnea/hypopnea index is directly based on the number of apnea episodes detected over the total segments, rather than based on posterior probability of SVM outputs.
- 4) Two SVM classifier models, subject-independent and subject-dependent, are developed for apnea detection.
- 5) SVM models capable of controlling sensitivity and/or specificity are developed. This is accomplished by

designing SVM with unbalanced penalty parameters ( $C^+$ ,  $C^-$ ).

- 6) Increased predictivity performance of the classifier is achieved using a comprehensive set of HRV and EDR time and frequency-based features (111 features).
- 7) Fully automated ECG processing, feature extraction, and apnea SVM classifier implemented in a smartphone.

## III. SYSTEM ARCHITECTURE

The increased processing power available in today's smartphones and its capability to connect locally through Bluetooth (IEEE 802.15.1) and to the internet through Wi-Fi (IEEE 802.11) or 3G make it an attractive platform to implement a simplified personal sleep apnea monitor. The architecture and the various functional modules of "Apnea MedAssist" monitor are described in Fig. 1.

The ECG sensor used for recording heart activity is an off-the-shelf one lead ECG sensor with sampling rate of 250 Hz. The lightweight and small form factor of the ECG sensor makes it easy to wear for a prolonged period with the convenience of mobility. Additional advantages for the ECG sensor are the low-power consumption enabling long periods of ECG signal collection and easy connectivity to the smartphone via Bluetooth.

The computing platform used here is an Android operating system (OS) based universal mobile telecommunications system/global system for mobile communications (GSM/UMTS) smartphone [14]. It has Qualcomm MSM 7201A/528-MHz CPU running Android 1.5 OS [15]. It also has 128-MB RAM and 256-MB ROM with microSD extended flash memory storage card. Its supported connectivity interfaces include, Bluetooth for BAN/PAN, Wi-Fi 802.11 b/g for wireless local area network, and GPRS/HSDPA for wide area network. The smartphone provides initialization, configuration, and synchronization of Bluetooth connectivity to ECG sensor. The system also establishes secure communication to "MedAssist Server" using either Wi-Fi or cellular 3G connectivity.

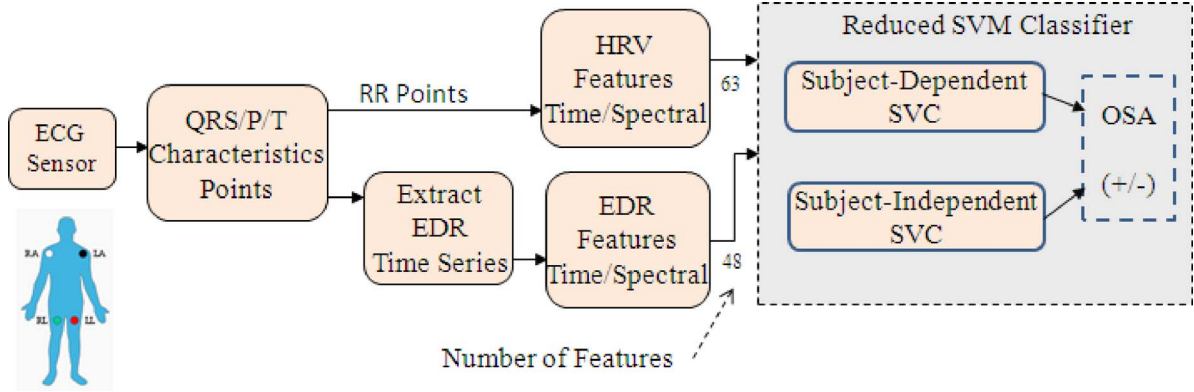


Fig. 2. Functional flow diagram for automated detection of OSA using single-lead ECG measurements.

TABLE I  
PHYSIONET APNEA-ECG DATABASE

	Length (min)	Non-Apnea (min)	Apnea (min)	Apnea (hrs)	AHI	Age	Height (cm)	Weight (kg)
Max	578	535	534	10	82.4	63	184	135
Min	43	20	0	0	0	27	168	56
Avg	489.3	303.2	186.1	5.17	28.6	46.5	177	87.2
Std	$\pm 32$	$\pm 158.6$	$\pm 173.1$	$\pm 3.93$	$\pm 27.3$	$\pm 10.1$	$\pm 4.4$	$\pm 20.2$

Max/min/average/std deviation of physical and apnea attributes of subjects used.

The main functions of the smartphone are the automated ECG signal processing including feature extraction, and OSA assessment using locally customized “*subject-dependent*” classifier or providing gateway connectivity to the server to use the “*subject-independent*” classifier.

#### IV. METHODOLOGY

This paper describes the design of “Apnea MedAssist,” a reliable automated OSA-monitoring device that uses measurements from just one lead ECG sensor. Fig. 2 shows the functional flow diagram for the signal processing and episode classification implemented on an Android-based smartphone.

##### A. Subjects Database

The device and algorithms were tested using Physionet Apnea-ECG Database [16]. The database has a total of 35 subjects’ sleep studies. The recordings were visually scored by an expert for sleep apnea/hypopnea events on the basis of respiration and oxygen on a per minute basis. The subjects’ recordings (30 men, 5 women) were arranged in three groups: Group A recordings (20 subjects) with clear occurrence of sleep apnea (100 min or more,  $AHI \geq 15$ ), Group B (borderline) recordings (five subjects) with some degree of sleep apnea (between 5 and 99 min,  $5 \leq AHI < 15$ ), and Group C (control) recordings (ten subjects) of healthy subjects with no sleep apnea (fewer than 5 min,  $AHI < 5$ ).

For apnea scoring, each record was divided into 1-min nonoverlapping segments [17]. Each minute was classified as either a “nonapnea minute” or an “apnea minute.” Minutes containing either apnea or hypopnea were classified as apnea minutes. The AI is the number of apneas observed per hour, and the HI is the number of hypopneas observed per hour. The

apnea–hypopnea index (AHI) is defined as the sum of AI and HI. Hours containing one to four apnea minutes (not considered to be clinically significant) are counted as hours without apnea. Table I shows the collected subjects’ data. This segment length reduces hidden apnea episodes that actually occur within the segment. Clearly, as we increase the segment length over 1 min, the actual estimate of AHI deviates considerably.

A single channel of ECG was extracted from all polysomnographic recordings, sampled at 100 Hz, with 16-bit resolution and 5- $\mu$ V A–D Converter (ADC) gain per step. The standard sleep laboratory modified lead V2 position ECG is used. The ECG signal was then resampled to an equivalent sample rate of 250 Hz used in “Apnea MedAssist.”

##### B. Automated ECG Processing

The ECG measurements with a sampling period of 4 m/s are segmented into 1-min epochs and then analyzed using “ECG MedAssist” signal-processing module. This is an automated wavelet-based analysis algorithm for denoising and detrending ECG signal, and detecting its characteristic points: QRS complex, P, and T waves. The wavelet transform algorithm used here is based on the undecimated lifting scheme (ULWT) [18], [19]. The ULWT has reduced computational cost compared to the basic finite-impulse response implementation. We use a low-order Daubechies/D4 wavelet to minimize the filter order and consequently reduce computational cost and overall filter delay. Single-decomposition phase with seven stages yields details  $\{D^s\}_{s=1}^7$  and approximations  $C^7$ . Signal denoising is implemented by zeroing out  $D^1$  detail coefficients. Detrending is performed by zeroing out  $C^7$  approximation coefficients.

To separate QRS complex from P–T waves, we extract two signals by reconstructing two groups of ULWT subbands.



$\{D^s\}_{s=2}^5$  are used for QRS-complex signal reconstruction ( $S_{QRS}$ ) and  $\{D^s\}_{s=5}^7$  are used for P and T waves reconstruction ( $S_{PT}$ ). We also apply real-time adaptive thresholding [18] to all details before each reconstruction stage to minimize spectral overlap between the QRS and the PT signals and also to emphasize the core details (those details defining either QRS or PT waves). The novel level-based thresholding that removes an amount proportional to a window centered on the corresponding coefficients of the earlier stage, also amplifies the contributing higher spectral energy detail levels. For  $x_j$ , the  $j$ th coefficient for  $s$ th details  $D^s$ , we define a window of length  $2w + 1$  centered at  $x_j$  of the earlier level  $D^{s-1}$ . Then, the applied threshold for that coefficient is

$$\text{Thr}_j^s = t^s - \frac{1}{2w + 1} \sum_{i=j-w}^{j+w} (|x_i^{s-1}| - V^s) \quad (1)$$

where  $t^s$  is the initial threshold for level  $s$ ,  $x_i^{s-1}$  is the  $i$ th coefficients for level  $s - 1$ .  $V^s$  is the window average for  $D^2$  or  $D^3$ , and 0 for  $D^4$  or  $D^5$  and “ $| \cdot |$ ” is the absolute value operator.

We then use rules-based detection algorithm with no search-back based on Tompkins method [20]. It uses the reconstructed  $S_{QRS}$  and  $S_{PT}$  signals to detect the five characteristic points of QRS-complex including the R fiducial point (PQ, Q, R, S, and J), and the six characteristic points of P and T waves ( $P_{\text{start}}$ ,  $P_{\text{peak}}$ ,  $P_{\text{end}}$ ,  $T_{\text{start}}$ ,  $T_{\text{peak}}$ , and  $T_{\text{end}}$ ).

The RR-interval time series  $\{\text{RR}(m) : \text{rr}_i, i = 1, \dots, m\}$ , where  $\text{rr}_i$  is the  $i$ th RR interval, is analyzed to detect and remove any outliers ( $\geq 2$  s or  $\leq 0.4$  s) due to false detections or missed detections using a sliding window averaging filter (window of 41 samples) with an exclusion threshold of  $\pm 20\%$ .

### C. Respiration Efforts Estimate (EDR and RSA)

The morphology changes in the ECG waves allow deriving a signal proportional to the respiratory movement. In the literature, various methods were proposed to extract the surrogate EDR signal [7], [21] based on R-wave amplitude, R-wave duration, QRS complex area, T-wave amplitude or T-wave area. We use the T-wave method, since it is more suited for ECG recording with low sampling rates ( $< 250$  Hz) as in our case. However, we use the R-wave amplitude method on recordings with undetected T-wave (or inverted T-wave). Fig. 3 shows a 20-min segment of one of the recordings with the computed EDR and corresponding respiratory signals obtained using other respiratory sensor measurements.

The calculated EDR sequence is analyzed to detect and remove any outliers due to false detections or missed beats using a sliding window averaging filter (window of 41 samples) with an empirically selected exclusion threshold ( $\pm 70\%$ ). The extracted respiratory time series is denoted as  $\{\text{EDR}(q) : \text{edr}_i, i = 10, \dots, q\}$ .

Another ECG derived respiratory signal is respiratory sinus arrhythmia (RSA), defined by the heart rate variability in synchrony with the respiration, where R-R interval from the ECG is shortened during inspiration and prolonged during expiration [22]. To estimate respiratory frequency, we define

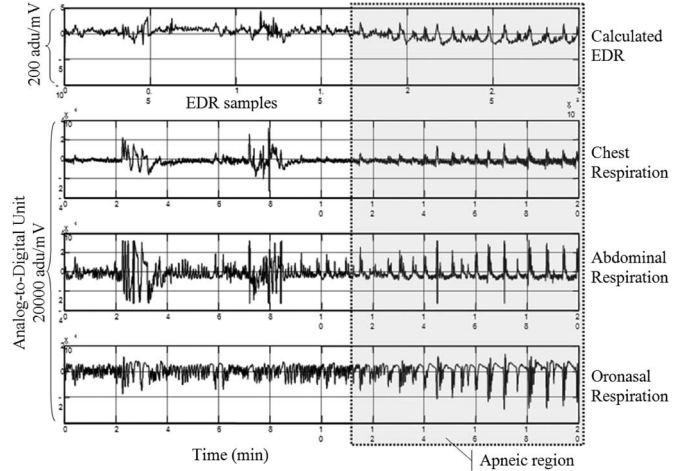


Fig. 3. ECG-derived respiratory (EDR) signal compared to actual readings for the “a01” Apnea-DB subject’s respiratory signals. Shaded region is apneic.

“number of extreme points” (NEP) as the biased average of extreme points [23] as follows:

$$\text{NEP} = \frac{1}{m-2} \sum_{i=2}^{m-1} (1 - \text{unit}[(\text{rr}_i - \text{rr}_{i-1})(\text{rr}_{i+1} - \text{rr}_i)]) \quad (2)$$

where  $\text{unit}[\cdot]$  is the unit step function. This measure is sensitive to the  $\text{RR}(m)$  series filter processing of outliers and missed detections. For each segmented 30 s epoch,  $m$  is the number of heart beats in the segment. The NEP measurements can be used in evaluating parasympathetic nerve activity [22].

### D. Feature Measures

In this research, we considered a set of time-based and spectral-based features for both the training and the test datasets to increase classifier (+/-)OSA predictivity performance. The time-based features in Table II and the spectral-based features in Table III are calculated for both  $\text{RR}(m)$  and  $\text{EDR}(q)$  time series. The majority of these time- and spectral-domain features are described in [24]–[26]. Here, we provide more details on the predominantly used spectral-based features for both signals.

Both of the irregularly sampled time series  $\text{RR}(m)$  and  $\text{EDR}(q)$  are spline-resampled at 2 Hz and then detrended by removing the resampled time-series average. Two spectral analysis methods are performed: wavelet transform (WT) decomposition and fast Fourier transform (FFT). Decimated WT (DWT) using DB4 wavelet is performed with eight levels for  $\text{RR}(m)$  and nine levels for  $\text{EDR}(q)$  with the first-level details and approximations discarded, since they lie outside the *autonomous system* (ANS) frequency ranges [25].

The spectral variances of the DWT decomposition scales are then calculated as follows:

$$\text{Var}[D_k^s] = \sum_{i=1}^{K_{k,s}} (d_{k,i}^s - \mu_s)^2, \quad \text{where } \mu_s = E[d_{k,i}^s] \quad (3)$$

where  $k$  denotes either  $\text{rr}$  or  $\text{edr}$  signal,  $d_{k,i}^s$  is the  $i$ th detail coefficient, and  $K_{k,s}$  is the detail coefficient count at scale  $s$  of signal  $k$ . Three more features are extracted for each sequence using the

TABLE II  
TIME-DOMAIN MEASURES FOR HRV RR( $m$ ) AND EDR( $q$ ) EPOCH SEQUENCE

RR( $m$ )	RR( $m$ ) = $\{rr_i\}_{i=1}^m$	Feature Count
Mean	$\mu_{rr} = E[rr_i]$	1
Standard deviation	$\sigma_{rr} = \sqrt{E[(rr_i - \mu_{rr})^2]}$	1
Sum of beats with interbeat difference over 50ms, variant 1	$NN50v1 = \sum_{i=2}^m \text{unit}[ rr_i - rr_{i+1}  - 50\text{ms}]$	1
Sum of beats with interbeat difference over 50ms, variant 2	$NN50v2 = \sum_{i=1}^{m-1} \text{unit}[ rr_{i+1} - rr_i  - 50\text{ms}]$	1
Ratio of NN50v1 to segment length	$pNN50v1 = \frac{NN50v1}{m}$	1
Ratio of NN50v2 to segment length	$pNN50v2 = \frac{NN50v2}{m}$	1
Standard deviation of interbeat differentials	$\sigma_{rd} = \sqrt{E[(rd_i - \mu_{rd})^2]}$ $rd_i = rr_{i+1} - rr_i$ and $\mu_{rd} = E[rd_i]$	1
Root mean square of interbeat differentials	$RMSSD = \sqrt{E[rd_i^2]}$	1
Serial correlation coefficients ( $k = 1, \dots, 5$ )	$r_k = \frac{\sum_{i=1}^{m-k} (rr_i - \mu_{rr})(rr_{i+k} - \mu_{rr})}{\sum_{i=1}^{m-k} (rr_i - \mu_{rr})^2}$	5
Fractal <i>Alan factors</i> ( $k = 5, 10, 15, 30$ )	$AT_k = \frac{E[(N_{i+1}[k] - N_i[k])^2]}{2E[N_{i+1}[k]]}$ $N_i[k]$ is the number of beats in the $i$ th window of $k$ seconds	4
NEP as defined in Eq. (2)		1
EDR( $q$ )	EDR( $q$ ) = $\{edr_i\}_{i=1}^q$	Feature Count
Mean	$\mu_{edr} = E[edr_i]$	1
Standard deviation	$\sigma_{edr} = \sqrt{E[(edr_i - \mu_{edr})^2]}$	1

WT variances by identifying the frequency bands of WT details corresponding to very low frequency (VLF) range ( $0.003 < f \leq 0.04$  Hz), low-frequency (LF) range ( $0.04 < f \leq 0.15$  Hz), and high-frequency (HF) range ( $0.15 < f \leq 0.4$  Hz). Breathing disorder causes short-term oscillations that reflect changes in the relative balance between the sympathetic and parasympathetic branches of the ANS, the sympathovagal balance. HF range thought to be mainly due to the parasympathetic activity at the sinoatrial node. Since respiration is a parasympathetically mediated activity (through the vagal nerve), a peak corresponding to the rate of respiration can often be observed in the HF band. The LF region is generally thought to reflect sympathetically mediated activity. Similarly, a 256 point FFT power spectral density (PSD) estimate is performed and only the first 32 points related to ANS frequency range (0.003–0.4 Hz) are kept and three additional features are calculated for each VLF, LF, and HF range.

Tables II and III show the extracted features listing and corresponding feature count contribution. Each 1-min ECG segment is now mapped to the full set (FS) of  $n = 111$  total extracted feature measures (63 using RR time series and 48 using EDR time series).

#### E. Support Vector (SV) Classifier Models

The SVM classifier simply performs classification by constructing an  $n$ -dimensional hyperplane that optimally separates the data into two classes. The SVM has been developed by Vapnik [5], [6] and its formulation uses the structural risk min-

TABLE III  
SPECTRAL-DOMAIN MEASURES FOR RR( $m$ ) AND EDR( $q$ ) EPOCH SEQUENCE

RR( $m$ ) = $\{rr_i\}_{i=1}^m$	Feature Count
DWT Variance $\{Var[D_{rr}^s]\}_{s=2}^8$	7
DWT Variance (over VLF, LF and HF)	3
PSD/FFT {0.003–0.4 Hz} Points	32
PSD/FFT (over VLF, LF and HF)	3
EDR( $q$ ) = $\{edr_i\}_{i=1}^q$	Feature Count
DWT Variance $\{Var[D_{edr}^s]\}_{s=2}^9$	8
DWT Variance (over VLF, LF and HF)	3
PSD/FFT {0.003–0.4 Hz} Points	32
PSD/FFT (over VLF, LF and HF)	3

imization (SRM) principle, which proved superior to the traditional empirical risk minimization (ERM) principle used by traditional neural networks [27]; SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. SVM finds the Optimum-separating hyperplane with maximum distance from nearest training patterns. Those nearest training vectors are called support vectors (SVs). Fig. 4(a) shows the case where the training data have linearly separable classes (+)OSA and (–)OSA.

For training vectors  $\mathbf{x}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, l$ , of two classes, and a label vector  $\mathbf{y} \in \mathbb{R}^l$  such that  $y_i \in \{+1, -1\}$ , there exist a vector  $\mathbf{w} \in \mathbb{R}^n$  normal to the hyperplane and margin bias  $b \in \mathbb{R}$ , and the points  $\mathbf{x}$  that lie on the hyperplane satisfy  $\mathbf{w} \cdot \mathbf{x} + b = 0$ ,

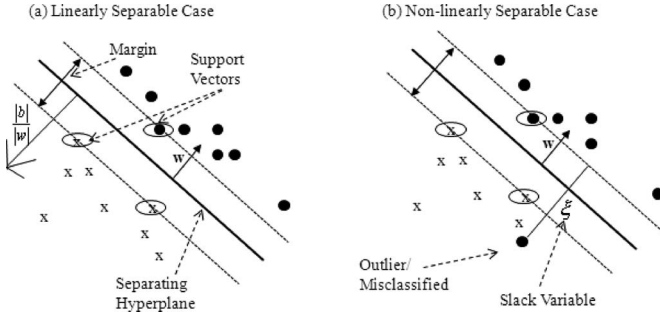


Fig. 4. SVM optimal separating hyperplane for two classes (+1, -1). (a) Linearly separable datasets. (b) Nonlinearly separable datasets.

( $\mathbf{w} \cdot \mathbf{x} = \mathbf{w}^T \mathbf{x}$ , where  $(\cdot)^T$  is the transpose operator) and the training data satisfy the constraint  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \forall i$ , with a maximum distance in this case  $2 \|\mathbf{w}\|^{-2}$  [27]. This is a convex optimization problem that can be solved by quadratic programming (QP) [27], [28], with the use of Lagrange multipliers  $\alpha_i$ ,  $\forall i$ , which is equivalent to solving the dual problem

$$L_D \equiv \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (4)$$

such that  $\alpha_i \geq 0 \forall i$  and  $\sum_{i=1}^l \alpha_i y_i = 0$ .

Fig. 4(b) shows the general case, where the training data are not linearly separable due to noise and outliers or just training data distribution. No linear separating hyperplane can be constructed now. In this case, two methods are introduced. First we introduce nonnegative “slack” variables ( $\xi_i \geq 0 \forall i$ ), where  $l$  is the number of training vectors, and minimize the penalty function  $F_\sigma(\xi) = \sum_i \xi_i^\sigma$ , for small  $\sigma > 0$ , subject to the modified constraints

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \quad (5)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l. \quad (6)$$

The problem now corresponds to minimizing the following objective function:

$$\frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^l \xi_i^\sigma \quad (7)$$

where  $C$  is a constant defined as the “penalty” parameter or “misclassification tolerance” parameter [27]. This SVM is defined as soft-margin SVM using L1-norm penalty ( $C$ ) or C-SVM. A larger  $C$  assigns a higher penalty to errors.  $F_\sigma$  describes the training errors and is a monotonic convex function. As shown in [5],  $F_\sigma$  is a nondeterministic polynomial (NP) complete problem. To avoid the NP completeness,  $\sigma = 1$  is considered as the smallest value of  $\sigma$  for which the optimization problem has a unique solution.

Second, we introduce a mapping  $\Phi: \mathbb{R}^n \rightarrow \mathcal{H}$  by which the dataset is mapped to some other (possibly infinite dimensional) “feature space”  $\mathcal{H}$  that has a linear hyperplane separating the two classes. The dot product in  $\mathcal{H}$ -space is computed using a “kernel” function  $K$  such that  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$ . The

TABLE IV  
KERNEL-MAPPING FUNCTIONS USED IN SVCs

Kernel Parameters	Mapping Function $K(\mathbf{x}_i, \mathbf{x}_j)$
Linear ( $C$ )	$\mathbf{x}_i \cdot \mathbf{x}_j$
Polynomial ( $C, \gamma, r, d$ )	$(\gamma \mathbf{x}_i \cdot \mathbf{x}_j + r)^d, \gamma > 0$
Gaussian/RBF ( $C, \gamma$ )	$e^{-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2}, \gamma > 0$
Sigmoid/MLP ( $C, \gamma, r$ )	$\tanh(\gamma \mathbf{x}_i \cdot \mathbf{x}_j + r)$

QP dual problem solution is modified to use the kernel mapping  $\mathbf{x}_i \cdot \mathbf{x}_j \Rightarrow K(\mathbf{x}_i, \mathbf{x}_j)$ , and modified  $\alpha$ -constraints.

To classify a test vector  $\mathbf{x}$  as either class (+1, -1), we compute the decision function  $f(\mathbf{x})$  as follows:

$$f(\mathbf{x}) = \text{sgn} \left[ \sum_{i=1}^{N_s} (\alpha_i y_i K(\mathbf{x}_i, \mathbf{x})) + b \right] \quad (8)$$

where  $\text{sgn}[\cdot]$  is the sign function and  $N_s$  is the number of SVs. Some commonly used kernels with required parameters are shown in Table IV.

The Gaussian radial-basis function (RBF) kernel is a popular kernel since RBF kernels were shown to perform similar or better than linear or polynomial kernels for appropriately selected ( $C, \gamma$ ) values [5]. The sigmoidal kernel is a representation of a two-layer neural network multilayer perceptron (MLP). In this study, we developed SVM models based on these kernels and investigated the classification performance. A library for SVMs [28] is used for SVM training and classification, and MATLAB is used for SVs and feature reduction.

#### F. Learning and Classification Phases

Even though SVMs can generalize well, a careful choice of kernel function is necessary to produce a classification boundary that effectively separates (+/-)OSA classes. We use the automated system, aforementioned in two phases; training and classification (see in Fig. 5). Features with most predictive value are selected during feature search. During the learning phase we find the best performing classifier kernel type and parameters, and solve the SVM QP problem [5] to generate the set of SVs for the optimum separating hyperplane for the OSA’s two classes.

In classification phase the SVM classifies test epoch feature vector  $\mathbf{x}$  into (+/-)OSA classes. We use here  $K$ -fold cross-validation method to evaluate the performance of the classifier [29]. The cross-validation method can prevent the overfitting problem. Here,  $K$  depends on the classifier input training dataset size for both subject-dependent (SVC-SD) and subject-independent (SVC-SI) classifiers. For  $K$ -fold method, we divide the training set into  $K$  subsets of equal size. Sequentially one subset is tested using the SVM classifier trained on the remaining  $K-1$  subsets. The cross-validation accuracy is the percentage of data that are correctly classified.

The  $F$ -measure is used to compare the performance of SVM types and kernels using test sensitivity and specificity measures [30]

$$F = 2 \times \frac{\text{Sens} \times \text{Spec}}{\text{Sens} + \text{Spec}}. \quad (9)$$



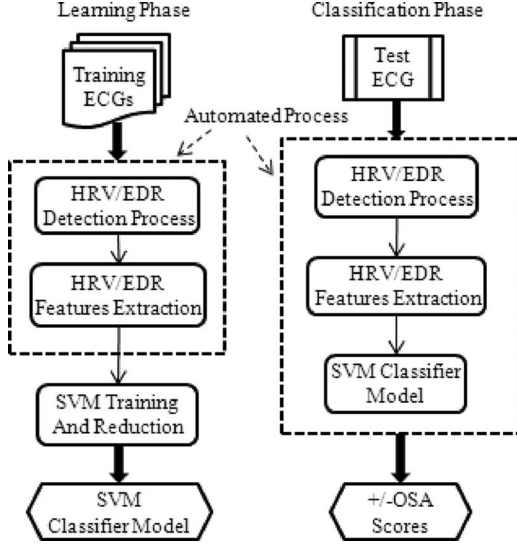


Fig. 5. Automated learning and classification of (+/-)OSA classes.

The SVM training is very sensitive to the dynamic ranges of feature matrix and differences in magnitude scale of these dynamic ranges. Any significant difference can lead to loss in precision during the computationally intensive SVM QP algorithm. All features are normalized to be within  $(-1, +1)$ . The normalizing factors generated during training are applied on test vectors prior to classification.

#### G. Reducing Classifier Complexity: SVs and Feature Counts

The  $F1$ -score measures the relative discrimination level of two sets of real numbers. Given training vectors  $\{x_i\}_{i=1}^l$ , if the number of positive and negative instances are  $l^+$  and  $l^-$ , respectively, then the  $F1$ -score of the  $k$ th feature is as follows [31]:

$$F1(k) = \frac{(\bar{x}_{k+} - \bar{x}_k)^2 + (\bar{x}_{k-} - \bar{x}_k)^2}{X_k^+ + X_k^-}$$

$$X_k^+ = \frac{1}{l^+ - 1} \sum_{i=1}^n (x_{i,k+} - \bar{x}_{k+})^2 \quad (10)$$

$$X_k^- = \frac{1}{l^- - 1} \sum_{i=1}^n (x_{i,k-} - \bar{x}_{k-})^2$$

where  $\bar{x}_k$ ,  $\bar{x}_{k+}$ , and  $\bar{x}_{k-}$  are the averages of the total, positive, and negative datasets, respectively, of the  $k$ th feature. The  $F1$ -score reaches its best value at 1 and worst value at 0.

Fig. 6(a) shows the  $F1$ -score discrimination measure for all features considered. These  $F1$ -scores were obtained after applying appropriate transformation over feature distribution. A transformation was applied where appropriate so that the histogram of the transformed feature approximates more closely a Gaussian distribution [13]. This transformation enhances the  $F1$ -scores of those features. A logarithmic transform was applied to  $\sigma_{rr}$ ,  $\sigma_{rd}$ ,  $RMSSD$ ,  $\sigma_{edr}$ , and spectral-domain features. A square-root transform was applied to  $AT_k$ .

One disadvantage of the  $F1$ -score is that it does not reveal mutual information among features. Despite this disadvantage, the  $F1$ -score is simple and generally quite effective. Another feature discrimination score used in isolating best performing features is receiver operating characteristics (ROC) graphs [30], [32]. Fig. 6(b) summarizes the calculated ROC analysis scores for the various features. We select high  $F1$ -scoring features from those listed in Tables II and III according to features type or group. We then intersect these with features having ROC scores  $\geq 0.5$  [31].

A local type of the SVM classifier uses only the training set of each subject to generate a customized classifier for the subject. This subject-dependent SVM classifier (SVC-SD) significantly decreases the complexity of the SVM model and enhances the classifier prediction and accuracy.

The  $K$ -fold cross-validation method is also used for evaluating the performance of SVC-SD. Here we use  $K = 30$ , based on the number of subjects in the database and available subject's data. The final classification performance was calculated based on the total true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Each input features' data vector is classified only once.

#### H. Controlling Classification Sensitivity and Specificity

In this classification problem, the numbers of datasets in the two classes (+/-)OSA are unbalanced. Depending on the subject, the data can be either heavily (-)OSA or heavily (+)OSA. In this study, the overall 1-min segments distribution is heavily (-)OSA, where  $\Pr[(-)OSA] = 0.63$ . In clinical applications of apnea detection, the sensitivity optimization is preferred over specificity optimization within an appropriate accuracy cost. In this case, we should locate our optimal separating hyperplane with an increasing bias toward the (+)OSA class. With a modified C-SVM formulation [5] that uses different penalty parameters for the two classes  $y^+ \equiv (y_i = +1)$  and  $y^- \equiv (y_i = -1)$ , (7) becomes

$$\frac{\|\mathbf{w}\|^2}{2} + C^+ \sum_{i \in y^+} \xi_i + C^- \sum_{i \in y^-} \xi_i \quad (11)$$

subject to  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$  and  $\xi_i \geq 0 \forall i$ .

By assigning a higher penalty for the (+)OSA class than the (-)OSA class, we allow higher predictivity at the expense of specificity within an appropriate accuracy cost.

## V. RESULTS

Fig. 6(a) identifies the groups of features with highest  $F1$  discrimination scores, which are also summarized in Table V. Our reduced feature set (RS) comprises feature groups A, B, C, E, and G. Groups D and F were not selected because of the higher computation cost using PSD/FFT. The ANS ranges (LF and VLF) were excluded from group E since they are covered by using FS of EDR spectral variances  $\{\text{Var}[D_{edr}^s]\}_{s=2}^9$ . Each 1-min ECG segment is now mapped to the RS of  $n' = 19$  total extracted feature measures (11 using RR time series and 8 using EDR time series).

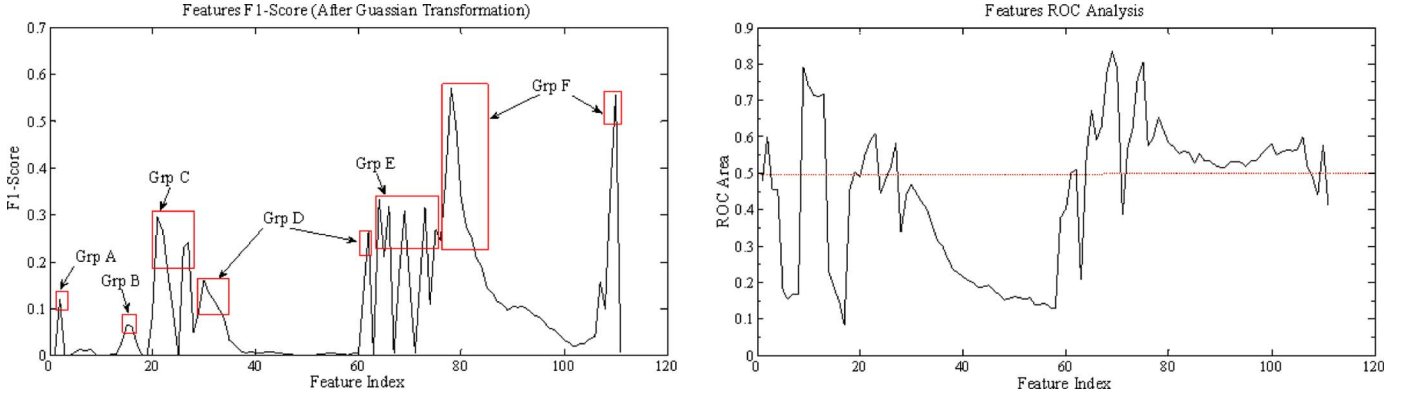


Fig. 6. Feature selection using  $F1$ -measure and ROC. (a) Using  $F1$ -score discrimination measure for features considered after normal distribution transformation. (Grp\* denotes Group\*, see Table V). (b) Features' ROC scores (Only features with ROC scores  $\geq 0.5$  are selected).

TABLE V

BEST DISCRIMINATION FEATURES BASED ON FEATURE'S  $F1$ -SCORE.  $D_{RR}^s$  IS  $s$ TH SCALE COEFFICIENTS OF  $RR(m)$  WAVELET DECOMPOSITION AND  $D_{EDR}^s$  IS  $s$ TH SCALE COEFFICIENTS OF  $\{EDR(q)\}$  WAVELET DECOMPOSITION

Group	Indices	Features	RS
A	2	$\sigma_{rr}$	1
B	15–16	$AT_{10}, AT_{15}$	2
C	20–24	$Var[D_{rr}^4] \cdots Var[D_{rr}^8]$	5
	31–32	WT ANS ranges: LF, VLF	2
D	29–33	$[S_{rr}(i)]^2, i = 3, \dots, 7$	
	61–62	FFT ANS ranges: LF, VLF	
E	65–72	$Var[D_{edr}^2] \cdots Var[D_{edr}^9]$	8
	73–75	WT ANS ranges: LF, VLF	
F	77–107	$[S_{edr}(i)]^2, i = 3, \dots, 31, 32 - 53$	
	108–110	FFT ANS ranges: LF, VLF	
G	111	NEP	1

Using our automated process described earlier, we extracted a large number ( $l = 14\,700$ ) of training samples (of 1-min segment feature vectors) from the subjects Apnea-ECG database. We then used the  $K$ -fold cross-validation method for assessing the performance of the various classifier models developed in our study and for finding the optimum kernel and SVM parameters. Different  $K$  values ( $K = 5, 10$ , and  $35$ ) are used for each cross-validation run. The classification performances accuracy, sensitivity, and specificity, are calculated using the overall classification results over the  $K$ -fold iterations of the same  $K$ -value. In this way, every training vector  $\mathbf{x}_i$  is classified only once.  $K$ -fold cross-validation methods are used to prevent the classifier overfitting issues [29].

#### A. SVM Classifier—Subject-Independent (SVC-SI)

Table VI shows the results for the  $K$ -fold cross-validation performance of SVC-SI classifier with different kernels and parameters ( $C, \gamma$ ) using the FS; comprising  $n = 111$  temporal and spectral domain (RR + EDR) features. Table VII shows the results for SVC-SI using the reduced set (RS); comprising of  $n' = 19$  features. The kernels considered are linear, polynomial, Gaussian/RBF, and Sigmoid/MLP. Each kernel type has different number of kernel parameters in addition to the SVM model penalty parameter  $C$  that determine the training strategy to find the optimal kernel values and also affects the compu-

tational complexity during the training and classification. The linear kernel represents effectively “no kernel mapping” and hence has no parameters. The polynomial kernel parameters are  $(\gamma, r, d)$ . The sigmoid/MLP kernel parameters are  $(\gamma, r)$ . The Gaussian/RBF kernel's only parameter is  $(\gamma)$ . The RBF is considered very efficient numerically and geometrically suited for nonlinear data distributions.

The best performing classifier for both sets (FS and RS) was the one using Gaussian-kernel (Tables VI and VII). Fig. 7 shows the performance of the Gaussian-kernel classifiers for both features sets.

#### B. SVM Classifier—Subject-Dependent (SVC-SD)

Table VIII using the FS and Table IX using the reduced set (RS) show the best performing  $(C, \gamma)$  for the SVC-SD classifier with several kernel types: linear, polynomial ( $d = 2$ ), RBF, and MLP. We used a balanced penalty parameter  $C$  for both classes (+)OSA and (−)OSA ( $C^+ = C^- = C$ ).

When biasing the penalty parameter  $C$  towards (+)OSA to enhance sensitivity within an accepted accuracy, the  $C^+$  was chosen as  $5 \times C$ , and  $C^-$  as  $1 \times C$ . Table X shows the results of SVC-SD with sensitivity bias. We selected only the linear and Gaussian/RBF kernels since they exhibited better performance and provided efficient implementation in real-time monitors.

## VI. DISCUSSION

Although SVM can build classifiers with high-classification accuracy, the response time of SVM classifiers still need to be improved when applied in real-time systems. In this study our goal is to design an automated processing system to screen 1-min sleep epochs for apnea using SVM classifiers and identify the parameters and methods that allow efficient implementation in real-time monitors. Three elements affecting the response time of SVM classifiers [33], [34] are the number of feature input variables, the mapping kernel type, and the number of the SVs.

#### A. Feature Set Selection

The applied feature selection methods are independent of SVM model and kernel type. Using both  $F1$  score and ROC areas [30], we selected the best performing set (RS) that is a reduction



TABLE VI  
K-FOLD CROSS-VALIDATION PERFORMANCE OF SVC-SI USING FS

LINEAR		POLY ( $d=2$ )			RBF			MLP		
$C$	Acc (%)	$C$	$\gamma$	Acc (%)	$C$	$\gamma$	Acc (%)	$C$	$\gamma$	Acc (%)
1	85.65	1	0.15	86.57	1	1	89.50	1	0.5	51.05
2	85.63	1	0.25	86.74	1	0.6598	89.54	1	1	56.66
8	85.64	2	0.25	86.50	0.6598	1	89.07	2	1	56.37
15	85.71	4	0.25	86.14	0.6598	0.6598	89.12	1	2	57.60
16	85.70	8	0.25	85.66	2	1	89.48	2	2	57.30
64	85.61	1	0.50	86.14	2	2	88.01	4	2	57.04
128	85.63	2	0.50	85.66	2	0.5	89.88	8	2	56.86
512	85.62	1	0.75	85.54	1.7411	0.8706	89.61	16	2	56.74

Accuracy (Acc) performance of subject-independent  $C$ -SVM classifier (SVC-SI) with different kernels and parameters ( $C, \gamma$ ) for the full feature set (FS).

of the original feature set by 83%. This set had two components: RR-intervals based features representing HRV variability that captured vagal and sympathetic variability and EDR and NEP features, which represent the approximate respiratory variability efforts. The spectral-domain features corresponding to VLF and LF frequency bands provide the best discriminant features.

The other features of the FS have increased the classification performance to 91%, but it significantly increases the memory size requirement and test phase computations. One advantage in using a large number of features is it enhances the performance for a simple linear kernel without the need to map data to a higher dimensional space. In this case, the nonlinear mapping does not improve the performance significantly (see in Tables VIII and IX).

### B. SVM Classifier Kernels

The performance of various kernel types was evaluated with constrained and unconstrained SVM penalty parameter  $C$  in the training phase using grid-search methods with a coarse dyadic scale. Then a more detailed search was performed in the kernel parameters region with a high accuracy. The best performing combination of SVM-kernel parameters are reported in Tables VI–X.

The linear kernel results are shown for various  $C$  values where  $C \geq 512$  approximates unconstrained SVM. The linear kernel performs very well in a high-dimension feature set with  $n = 111$  as shown in Table VI and its performance degrades as we reduce the feature set (see Table VII). The polynomial kernel is a common method to work with nonlinear separation of data. We selected quadratic polynomial of ( $d = 2$ ). It shows very good performance for both FS and RS feature sets. The Poly/ $d = 2$  also exhibits very good resilience to the feature set reduction from  $n$  to  $n' = 19$ . The disadvantages of the polynomial kernel are the number of kernel parameters to consider and high computation cost for real-time monitors.

The sigmoid kernel was quite popular for SVMs due to its origin from neural networks, where it models a multilayer perceptron (MLP) classifier with a single hidden layer; the SVs correspond to the first layer and the Lagrange multipliers ( $\alpha_i$ ) to the weights [35]. The performance results using the MLP reported are very poor, ( $<65\%$ ). This can be due to our SVM and kernel parameter search that was not able to find high perform-

ing ( $C, \gamma, r$ ) values. The sigmoid kernel matrix was shown to be conditionally positive definite (CPD) in certain parameters and thus are valid kernels only there [35] and hence, a kernel matrix CPD analysis is needed to identify the valid search boundaries.

The Gaussian RBF kernel nonlinearly maps the input dataset into a higher dimensional space. The RBF has less kernel parameters than polynomials and MLP kernels, hence reducing the complexity of the SVM model. The RBF is also numerically efficient and should be the first choice for general users [35].

### C. Subject Dependent and Independent SVMs

OSA screening and monitoring use can range from offline analysis, side effect studies, CPAP verification and feedback control, and also in ambulatory systems and perioperative monitoring. Although, SVM provides high generalization accuracy [5], the response time of SVM classifiers is still a concern when applied in real-time OSA monitors. By reducing the number of SVs (Tables VIII–X), and also by optimizing feature space using feature set reduction and selection of appropriate kernel (Tables VII and IX), we can speed up the response of SVM classifiers.

Table VII shows that the polynomial or Gaussian/RBF performs better with a RS. The advantage of collecting more efficient features for OSA screening (extracted from ECG or other physiological signals) is to enable the use of more efficient linear kernel in SVC-SI, as well as to improve the performance of the SVC. In this case, the number of SVs is high ranging from 2500 to 4100, which is 15%–30% of input dataset. We deployed the subject-independent SVC on “MedAssist Server” due to its high computation cost requirement.

A reduced subject-dependent SVC can be generated either based on prior subject’s sleep study or based on subject’s physiological attributes. This SVC can be deployed efficiently on real-time monitors that are used as home monitoring devices to allow OSA recognitions for applications like CPAP treatment assessment and feedback control. Tables VIII and IX show the performance of various kernels for the SVC-SD. The range of SVs is 15–160.

### D. SVM $C^+/C^-$ Penalty Parameters

Clinically, for OSA screening, sensitivity performance optimization is preferred over specificity performance within an

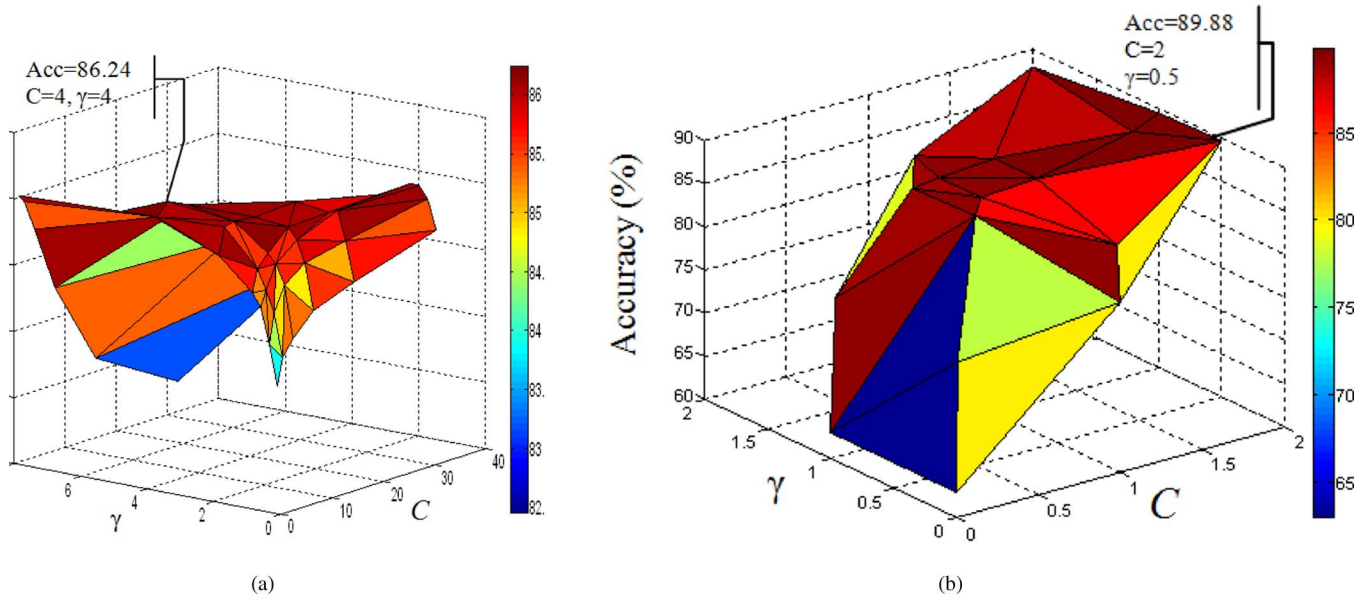


Fig. 7. Gaussian/RBF kernel cross-validation accuracy performance of SVC-SI for different SVM and kernel parameters ( $C, \gamma$ ) Using: (a) reduced feature set (RS). (b) complete feature set (FS).

TABLE VII  
K-FOLD CROSS-VALIDATION PERFORMANCE OF SVC-SI USING RS

LINEAR		POLY ( $d=2$ )			RBF			MLP		
$C$	Acc (%)	$C$	$\gamma$	Acc (%)	$C$	$\gamma$	Acc (%)	$C$	$\gamma$	Acc (%)
1	82.14	1	0.5	83.64	8	0.25	84.85	16	0.25	62.20
4	82.30	32	0.5	84.03	32	0.50	85.92	32	0.25	62.19
8	82.32	8	1	84.03	16	0.75	86.11	32	0.50	61.59
16	82.37	16	1	84.23	16	0.75	86.11	1	0.75	63.23
32	82.32	32	1	84.06	32	1	86.10	4	1	65.20
64	82.33	2	2	84.03	4	2	86.13	8	1	62.25
128	82.35	4	2	84.23	4	4	86.24	4	4	62.98
512	82.34	8	2	84.06	1	8	86.03	8	4	62.98

Accuracy (Acc) performance of subject-independent  $C$ -SVM Classifier (SVC-SI) with different SVM and kernels parameters ( $C, \gamma$ ) for the RS.

accepted accuracy (or  $F$ -measure). SVM model can be designed with unbalanced penalty parameters ( $C^+, C^-$ ) to optimize either sensitivity and/or specificity.

Table X shows the increased sensitivity performance (96%) while keeping the same  $F$ -measure around (89.90%).

#### E. Comparison With Previous Studies

As reported earlier, other studies [10], [11] addressed SVM classifiers in OSA screening and also other researchers have applied different other methods [12], [13].

The SVM studies in OSA are still new in published studies. These studies either [10] used the complete nocturnal recording as one segment or used posterior probability as OSA severity indicator. In addition, they do not consider model complexity or method of controlling predictivity. In [11], authors use two different biomedical sensor signals without using EDR.

The authors of [12], [13] use different classifier models, or different feature extraction methods and coarser time resolution

for OSA recognition because they apply temporal averaging over multiple epoch segments to enhance performance.

In contrast, in this study we considered per minute segment based OSA detection and presented several new methods to increase the predictivity of the classifier performance and also reduce the complexity of the classifier for efficient implementation in real-time monitors. We presented two models of soft-margin support vector classifiers ( $C$ -SVC); subject-independent and subject-dependent SVC models combined with unbalanced “penalty parameters” ( $C^+, C^-$ ). Also, we presented methods for efficient automated processing of ECG recording and feature extraction for the OSA recognition in real-time monitors. We have implemented our SVM classifier successfully on a smartphone with Android OS and this is first of such implementation to the best of our knowledge.

The other studies have presented accuracies (90%–91%) similar to or lower than ours for the subject independent classifier. For the subject-dependent classifier, our methods achieved a higher predictivity of 96%.

TABLE VIII  
K-FOLD CROSS-VALIDATION PERFORMANCE OF SVC-SD USING FS

<i>C</i>	LINEAR				<i>C</i>	POLY ( <i>d</i> =2)				<i>C</i>	RBF				<i>C</i>	MLP			
	Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)
0.5	90.02	91.09	90.86	91.73	0.5	85.42	88.79	88.01	90.78	0.5	89.62	90.76	90.52	91.43	0.5	72.71	80.27	78.25	84.71
1	89.80	91.18	90.88	92.00	0.5	83.47	87.74	86.73	90.26	0.5	88.74	90.36	90.01	91.31	0.5	72.75	80.28	78.28	84.71
2	89.62	91.11	90.78	91.98	0.5	83.17	87.52	86.48	90.08	1.0	89.21	90.84	90.49	91.81	1.0	72.39	80.15	78.07	84.71
8	89.02	90.99	90.55	92.14	1.0	84.44	88.25	87.36	90.49	2.0	89.02	90.86	90.46	91.94	2.0	72.39	80.15	78.07	84.71
32	89.12	91.16	90.70	92.35	1.0	81.74	88.79	85.24	89.06	4.0	88.62	90.73	90.26	91.96	4.0	72.39	80.15	78.07	84.71
					1.0	89.06				0.5					2.0				

Accuracy (Acc) and F-measure performance of subject-dependent *C*-SVM Classifier (SVC-SD) with different SVM and kernels parameters (*C*,  $\gamma$ ) for the full feature set (FS).

TABLE IX  
K-FOLD CROSS-VALIDATION PERFORMANCE OF SVC-SD USING RS

<i>C</i>	LINEAR				<i>C</i>	POLY ( <i>d</i> =2)				<i>C</i>	RBF				<i>C</i>	MLP			
	Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)		Sens (%)	Acc (%)	F (%)	Spec (%)
0.5	88.35	89.83	89.52	90.72	0.5	88.25	89.85	89.52	90.82	1.0	88.35	89.71	89.43	90.51	0.5	74.66	80.45	79.04	83.96
1	88.03	89.86	89.48	90.97	1.0	87.81	89.65	89.26	90.76	2.0	87.97	89.75	89.38	90.8	1.0	74.38	79.83	78.52	83.13
2	87.99	89.91	89.50	91.06	2.0	87.51	89.59	89.15	90.84	4.0	87.23	89.40	88.94	90.71	0.5	74.50	80.03	78.69	83.38
8	87.69	89.87	89.40	91.18	0.5	87.51	89.59	89.15	90.84	1.0	88.15	89.54	89.25	90.71	1.0	74.36	79.98	78.62	83.39
32	87.35	89.72	89.21	91.14	1.0	86.63	89.23	88.67	90.81	2.0	87.55	89.46	89.06	90.61	2.0	75.21	78.98	78.12	81.26
					1.0	90.81				1.0					2.0				

Accuracy (Acc) and F-measure performance of subject-dependent *C*-SVM Classifier (SVC-SD) with different SVM and kernels parameters (*C*,  $\gamma$ ) for the RS.

TABLE X  
K-FOLD CROSS-VALIDATION PERFORMANCE OF SVC-SD USING FS, ( $C^+ > C^-$ )

<i>C</i>	LINEAR				<i>C</i>	RBF			
	Sens (%)	Spec (%)	Acc (%)	F (%)		$\gamma$	Sens (%)	Spec (%)	Acc (%)
0.45	96.88	83.42	88.40	89.65	0.8	0.4	96.05	84.98	89.08
0.50	96.85	83.54	88.47	89.70	0.9	0.3	96.17	84.76	88.98
0.55	96.83	83.57	88.48	89.71	1.0	0.3	96.05	84.88	89.02
0.60	96.77	83.62	88.49	89.71	1.0	0.5	95.10	85.90	89.30

Accuracy (Acc) and F-measure performance of subject-dependent *C*-SVM Classifier (SVC-SD) with unbalanced penalty parameters ( $C^+ > C^-$ ) biased towards sensitivity.

## VII. CONCLUSION

Several *C*-SVM classifiers have been proposed for obstructive sleeping apnea screening and recognition based on ECG readings on a per 1-min epoch basis. An *F*-measure of 90% and sensitivity of 96% were achieved for the SVC-SD classifier. The subject independent *C*-SVM classifier can be used for initial screening or when no previous sleep study (PSG) data are available. The subject-dependent *C*-SVM classifier is based on subject's PSG that physicians can utilize in a custom designed classifier to achieve better performance. An efficient automated processing system on a smartphone has been implemented and evaluated to capture and process ECG, generate input features for the SVM classifier and classify OSA events. Several methods are proposed and evaluated to reduce the numerical computational complexity of the SVC model: feature selection and reduction strategies, kernel mapping, reduced SV subject-dependent classifier and use of unbalanced "penalty parameters" *C*. The classification accuracies were evaluated for these various methods.

Other SVM simplification strategies like SVM clustering or *K*-means clustering can be added to this study to simplify further the SVC model [36]. Addition of other input features (either extracted from ECG or other biomedical sensor like SpO<sub>2</sub>) can be added to the ones considered here to increase the accuracy.

## ACKNOWLEDGMENT

The authors would like to thank Prof. M. Nourani and Prof. G. Gupta of University of Texas and Dr. W. Brock of Presbyterian Hospital, Dallas, TX for helpful discussions.

## REFERENCES

- [1] W. T. McNicholas and P. Levy, "Sleep-related breathing disorders: Definitions and measurements," *Eur. Respir. J.*, vol. 15, no. 6, pp. 988–989, 2000.
- [2] American Academy of Sleep Medicine (AASM) Task Force, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, pp. 667–689, 1999.



- [3] R. K. Kakkar and R. B. Berry, "Positive airway pressure treatment for obstructive sleep Apnea," *Chest*, vol. 132, pp. 1057–1072, 2007.
- [4] C. den Herder, J. Schmech, D. Appelboom, and N. de Vries, "Risks of general anaesthesia in people with obstructive sleep Apnea," *Br. Med. J.*, (BMJ), vol. 329, pp. 955–959, 2004, 2011.
- [5] C. Cortes and V. Vapnik, "Support-vector machines," *Mach. Learning*, vol. 20, pp. 273–297, 1995.
- [6] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [7] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multilead ECGs," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.
- [8] N. Oliver and F. Flores-Mangas, "HealthGear: A real-time wearable system for monitoring and analyzing physiological signals," in *Proc. IEEE Int. Workshop Wearable Implantable Body Sens. Netw. (BSN)*, Apr. 2006, p. 4.
- [9] Y. Fu-Chung, K. Behbehani, E. Lucas, J. Burk, and J. Axe, "A noninvasive technique for detecting obstructive and central sleep Apnea," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 12, pp. 1262–1268, Dec. 1997.
- [10] A. H. Khandoker, M. Palaniswami, and C. Karmakar, "Support vector machines for automated recognition of obstructive sleep Apnea syndrome from ECG recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 37–48, Jan. 2009.
- [11] A. Patangay, P. Vemuri, and A. Tewfik, "Monitoring of obstructive sleep Apnea in heart failure patients," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS) Citi 1/2 Int.*, Lyon, France, Aug. 2007, pp. 1043–1046.
- [12] B. Raymond, R. M. Cayton, R. A. Bates, and M. J. Chappell, "Screening for obstructive sleep Apnea based on the electrocardiogram—The computers in cardiology challenge," *Comput. Cardiol.*, vol. 27, pp. 267–270, 2000.
- [13] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep Apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 686–696, Jun. 2003.
- [14] HTC Corporation, G1 Specifications. (2009). [Online]. Available: <http://www.htc.com/>.
- [15] Android Open Source Project. (2009). [Online]. Available: <http://www.android.com/>.
- [16] T. Penzel, "The Apnea-ECG database," *Comput. Cardiol.*, vol. 27, pp. 255–258, 2000.
- [17] G. B. Moody, R. G. Mark, A. L. Goldberger, and T. Penzel, "Stimulating rapid research advances via focused competition: The computers in cardiology challenge," *Comput. Cardiol.*, vol. 27, pp. 207–210, 2000.
- [18] M. Jansen and P. Oonincx, *Second Generation Wavelets and Applications*. New York: Springer-Verlag, 2005.
- [19] C. S. Lee, C. K. Lee, and K. Y. Yoo, "New lifting based structure for undecimated wavelet transform," *IEEE Electron. Lett.*, vol. 36, no. 22, pp. 1894–1895, Oct. 2000.
- [20] P. Hamilton and W. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986.
- [21] G. D. Furman, Z. Shinar, A. Baharav, and S. Akselrod, "Electrocardiogram derived respiration during sleep," *Comput. Cardiol.*, vol. 32, pp. 351–354, 2005.
- [22] F. Yasuma and J. Hayano, "Respiratory sinus arrhythmia," *Amer. College Chest Physicians, Chest*, vol. 125, pp. 683–690, 2004.
- [23] Y. Yoshida, K. Yokoyama, and N. Ishii, "Real-Time continuous estimation of respiratory frequency during sleep based on heart rate time series," in *Proc. 29th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 648–651.
- [24] M. Teich, S. B. Lowen, B. M. Jost, K. Vibe-Rheymmer, and C. Heneghan, "Heart-rate variability: Measures models," in *Nonlinear Biomedical Signal Processing*, M. Akay, Ed. Piscataway, NJ: IEEE Press, 2001, ch. 6, pp. 159–213.
- [25] G. D. Clifford, *Advanced Methods & Tools for ECG Data Analysis*. London, U.K.: Artech House, 2006.
- [26] A. Aldroubi and M. Unser, *Wavelets in Medicine and Biology*. Boca Raton, FL: CRC-Press, 1996.
- [27] S. R. Gunn, "Support vector machines for classification and regression," University of Southampton, Southampton, U.K., Tech. Rep., 1998.
- [28] C. Chang and C. Lin. (2001). "LIBSVM: A library for support vector machines," [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [29] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 1995, pp. 1137–1143.
- [30] C. J. van Rijsbergen, *Information Retrieval*, 2nd ed. London, U.K.: Butterworth-Heinemann Press, 1979.
- [31] Y. Chen and C. Lin, *Combining SVMs with Various Feature Selection Strategies*. New York: Springer-Verlag, 2006.
- [32] J. Hanley and B. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, pp. 839–843, 1983.
- [33] A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik, "Support vector clustering," *J. Mach. Learning Res.*, vol. 2, pp. 125–137, 2001.
- [34] Q. Tran, Q. Zhang, and X. Li, "Reduce the number of support vectors by using clustering techniques," in *Proc. 2nd Int. Conf. Mach. Learning Cybern.*, Xi'an, Nov. 2003, pp. 1245–1248.
- [35] S. Keerthi and C. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.
- [36] J. Wang, X. Wu, and C. Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems," *Int. J. Bus. Intell. Data Mining*, vol. 1, no. 1, pp. 54–64, 2005.



**Majdi Bsoul** (S'95–M'98) received the B.S. degree in electrical engineering from Jordan University of Science and Technology, Irbid, the M.S. degree in electrical and computer engineering from the University of Missouri, Columbia, in 1998, and Ph.D. degree in electrical engineering from the University of Texas at Dallas, Richardson, in 2010.

He was involved in R&D of next generation networks (NGN) in the wireless and IP networks for several telecommunication companies, and is currently engaged at Alcatel-Lucent, Plano, TX. His research interests include wireless sensor networks and applications in body area networks, sleep-related staging analysis and breathing disorders, and wavelet signal processing techniques for biomedical systems.



**Hlaing Minn** (S'99–M'01–SM'07) received the B.E. degree in electronics from Yangon Institute of Technology Yangon, Myanmar, in 1995, the M.Eng. degree in telecommunications from Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1997, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2001.

During 1998, he was a Laboratory Supervisor at AIT, where he was involved with the telecommunications program. From 1999 to 2001, he was a Research Assistant and during 2002, a Postdoctoral Research Fellow in the Department of Electrical and Computer Engineering, University of Victoria. In September 2002, he joined the Erik Jonsson School of Engineering and Computer Science, the University of Texas at Dallas, Richardson, where he is currently an Associate Professor. His research interests include wireless communications, statistical signal processing, error control, detection, estimation, synchronization, signal design, cross-layer design, cooperative/relay systems, cognitive radios, and biomedical and healthcare-related signal processing and wireless systems.

Dr. Minn is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.



**Lakshman Tamil** (S'81–M'88–SM'05) received the B.E. degree in electronics and communication engineering from Madurai Kamaraj University, India, in 1981, the M.Tech. degree in microwave and optical communication from the Indian Institute of Technology, Kharagpur, India, in 1983, and the M.S. degree in mathematics and the Ph.D. degree in electrical engineering from the University of Rhode Island in 1989.

He is a Professor in electrical engineering at the University of Texas at Dallas (UTD), Richardson, where he is also the Director of the Quality of Life Technology Laboratory. He has directed more than a dozen doctoral dissertations. He was also the Founder, CEO, and CTO of Yotta Networks Inc., which designed and marketed terabit-switching platforms. He has also directed research on advanced optical networks at Alcatel's Corporate Research Center, and he was a leader in creating both the early optical IP router and a multi-channel, multipoint distribution service that was a precursor to WiMAX. He has been an optical and wireless communication consultant to the Naval Research Laboratories, Raytheon Co., Electrospace Systems Inc., and Spike Technology. He has authored or coauthored more than 100 research publications and holds 18 U.S. patents.