SleepAp: An Automated Obstructive Sleep Apnoea Screening Application for Smartphones

J Behar[1], A Roebuck[1], M Shahid[1], J Daly[1], A Hallack[1], N Palmius[1], JR Stradling[2], GD Clifford[1]

[1]Dept. of Engineering Science, University of Oxford, Oxford, UK
[2]Sleep Unit, Oxford Centre for Respiratory Medicine, Churchill Hospital, Oxford, United Kingdom

Abstract

Obstructive Sleep Apnoea (OSA) is a sleep disorder with long term consequences. It is often diagnosed with an overnight sleep study or polysomnogram. Monitoring can be costly with long wait times for diagnosis. In this paper we describe a novel OSA screening framework and prototype phone application (app). A database of 856 patients that underwent at-home polysomnography was collected. Features were derived from audio, actigraphy, photoplethysmography (PPG) and demographics, and used as the inputs of a support vector machine (SVM) classifier. The SVM was trained on 735 patients (368 non-OSA and 567 OSA) and tested on 121 patients (44-77 split). Classification on the test set had an accuracy of up to 92.3%. The signal processing and machine learning algorithms were ported to Java and integrated into the phone app. The app records the audio, actigraphy and PPG signals, implements the clinically validated STOP-BANG questionnaire, derives features from the signals, and finally classifies the patient as needing treatment or not using the trained SVM. The resulting software could provide a new, easy-to-use, low-cost and widely available modality for OSA screening.

1. Introduction

Obstructive Sleep Apnoea (OSA) is a disorder characterised by pauses in breathing during sleep which lead to deoxygenation and consequent arousals. Long term consequences include sleep-related issues and cardiovascular diseases. OSA is often diagnosed with an overnight sleep study or polysomnogram which is considered to be the gold standard for OSA diagnosis. Monitoring can be costly (at least $1000 [1]) and wait time for diagnosis can be high. It is estimated that up to 90% of subjects with the disorder are undiagnosed [2].

The recent increase in the adoption of smartphones, along with the inclusion of high-quality internal sensors, has led to the proliferation of sleep screening smartphone applications (apps). Behar *et al.* [3] reviewed more than 40 of the sleep apps (Android market and Apple App Store) that make use of the smartphones' on-board sensors and concluded that none of these applications were based on clear scientific evidence (with the exception of apps that implemented simple sleep questionnaires).

This paper presents a novel OSA screening framework and associated phone application, 'SleepAp', which uses the phones internal sensors and an external pulse oximeter to record audio, activity, body position and oxygen saturation during sleep. This is the first proof of concept of a fully-automated smartphone medical diagnostic system that uses more than simple logic branching and decision support.

2. Methods

2.1. App framework

The ultimate goal of the developed phone application is to give a probability of a user belonging to one of the two following classes: non-OSA (healthy and snorers) and OSA (mild, moderate and severe). Figure 1 illustrates the phone application framework; audio, actigraphy, body position and photoplethysmography (PPG) can be recorded on the phone and sleep-related features are extracted from those signals. Using these features as well as the answers from a clinically validated questionnaire, the subjects are classified into one of the two classes. In addition, body position can eventually be used to advise the patients to change their sleeping position habits.

2.2. Clinical data

To validate the signal processing and machine learning algorithms to be ported on the smartphone, 856 randomly selected patients sent to the Respiratory Medicine Unit (Churchill Hospital, Oxford, UK) were recorded in a home environment using the Grey Flash medical equipment (Stowood Scientic Instruments Ltd, Oxford, UK). Features were derived from the audio, actigraphy and PPG signals and a support vector machine (SVM) classifier was trained on 735 patients (368 non-OSA and 567 OSA) and tested on 121 patients (44-77 split). Two experiments were conducted: EXP-1 where the objective was to classify be-

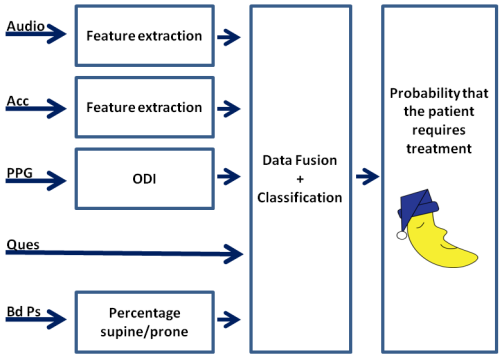

Figure 1. Phone application framework. Acc: Acceleration, PPG: Photoplethysmogram, ODI: Oxygen Desaturation Index, Ques: Questionnaire, Bd Ps: Body Position.

tween {Normal, Snorer} vs {Mild, Moderate and Severe OSA} and EXP-2 where the objective was to classify between {Normal, Snorer} vs {Moderate and Severe OSA} (i.e. dropping the Mild OSA individuals).

2.3. Operating system and general set-up

There are a number of suitable mobile phone operating systems available, with the most common including Android, iOS, Windows Mobile and BlackBerry. This application was developed for Android (versions 2.3.3 to 4.2.2) because of its large (and growing, especially in developing countries [4]) market share and open source licensing.

2.4. Signals

2.4.1. Actigraphy and body position

Mobile phone accelerometers vary in their precision and sensitivity. Acceleration range is measured in g which is equal to the Earth's gravitational strength at sea level ($9.8m.s^{-2}$). For the purpose of this study a frequency of $4Hz$ is sufficiently rapid and is in the range of some devices used in medical practice (e.g. $1Hz$ and $4Hz$ for VISI and Grey Flash (both from Stowood Scientic Instruments Ltd, Oxford, UK)). The accelerometer measures the acceleration applied to the device ($\overrightarrow{a}$). Conceptually, this is achieved by measuring forces applied to the sensor ($\overrightarrow{F_i}$) with $\sum_{i=1}^{N} \overrightarrow{F_i} = m\overrightarrow{a}$, where N is the number of external forces applied to the system. Gravity always influences the measured acceleration; splitting the left hand side in two: $\sum_{i=1 \neq gravity}^{N} \overrightarrow{F_i} + m\overrightarrow{g} = m\overrightarrow{a}$. This phenomenon can be observed by leaving the phone on a flat surface ($foralli \in N \neq gravity, \overrightarrow{F_i} = 0$) where the accelerometer gives a value of magnitude $g = 9.8ms^{-2}$. To access the actual acceleration of the device, the contribution of the force of gravity needs to be removed. This is achieved by applying a high-pass filter to the raw time series with cut-off frequency of $0.75Hz$.

The space was mapped into the different sleeping positions using the Tait angles provided by the Android API (pitch and roll angles were used but not azimuth). The space was mapped under the assumption that the phone is worn on the upper arm and is oriented parallel to the coronal plane. Actigraphy and body position recording thus require the user to wear an armband containing the phone.

2.4.2. Audio

The audio file recorded by the mobile phone has to be of the best possible quality in order to preserve all the features of the signal with their potentially associated diagnostic information. There are multiple parameters in the audio acquisition workflow that impact the audio quality [3]. The audio signal is recorded at $8kHz$ using uncompressed one-channel (mono) 16-bit pulse-code modulation. Recording the audio signal requires a microphone (which is usually part of the hands-free kit); the location 'next to but not under the nose' was identified as the best qualitative compromise between sound amplitude and noise. If the microphone is located too far from the nose (e.g. on the clothing, next to the ear) quiet breathing sounds are inaudible. Conversely, if the microphone is attached under the nose, the noise induced by the airflow is overly dominant and although a respiration pattern is apparent, thorough analysis of sound frequency content will be limited. The microphone can be attached on the face with surgical tape, plasters or any other medical adhesive.

2.4.3. Photoplethysmogram

The PPG was recorded using the WristOx2 3150 pulse oximeter (Nonin Medical Inc, MN, USA). The device was connected to the app via Bluetooth and records PPG at a sampling frequency of $75Hz$. The WristOx2 was attached to the subject's wrist and one finger (index, middle or ring) is inserted into the finger cuff. The pulse oximeter is powered by two 1.5V AAA batteries, and operating life is given as "24 hours minimum" in the model specification.

2.5. Feature extraction and classification

Many alternative algorithms for deriving features from sleep time series exist (see [5] for a review). Based on the work of Roebuck *et al.* [6], Higgins *et al.* [7], and Behar *et al.* [8], multi scale entropy (MSE), which corresponds to sample entropy calculated over varying time scales, was chosen as the features to quantify irregularity in the audio and actigraphy recordings, with extremes of entropy indicating pathological physiology. MSE was implemented

Table 1. EXP-1, SVM classification on clinical data: {Normal, Snorer} vs {Mild, Moderate and Severe OSA}

	Training set (735)			Test set (121)		
	Ac	Se	Sp	Ac	Se	Sp
ODI	84.4	80.1	88.6	85.1	80.5	93.2
AU	76.6	69.5	83.7	74.4	70.1	81.8
AC	63.1	60.5	65.8	72.7	74.0	70.5
DE	65.6	64.6	66.6	67.8	63.6	75.0
AU+AC	74.6	65.9	83.2	76.9	66.2	95.5
AU+AC+ODI	85.6	81.5	89.7	88.4	81.8	100

ODI: oxygen desaturation index, AU: audio, AC: actigraphy, DE: demographic, Ac: accuracy, Se: sensitivity, Sp: specificity. Maximum accuracy is indicated by an underline.

using a new Java version of the open-source C-code found on PhysioNet [9]. The oxygen desaturation index (ODI) was derived from the recorded oxygen saturation and corresponds to the average number of desaturation events per hour. A desaturation event was defined as a decrease of at least 4% from the average oxygen saturation (calculated from the previous 120sec) and lasting for at least 10sec.

Once the features have been derived – from audio, actigraphy, body position, PPG and the questionnaire – the user's data are classified using a trained SVM. The implementation draws upon the multi-platform open-source software library LIBSVM [10]. Development of the classifier took place in MATLAB to generate a structure containing the SVM type, kernel type (a Gaussian kernel was used), number of classes, offset, class labels and the support vectors (see [7] for details). This structure was then saved as a text file to allow the Android application to import it into usable form.

Since recording PPG requires the user to have a pulse oximeter which might be expensive (and thus not realistic for mass screening) and recording actigraphy requires an armband - which the user might not have- multiple SVM structures were trained on all possible combinations of signals. This means that a prediction can be obtained with any combination of recorded signals (audio, actigraphy, PPG and demographics).

3. Results

3.1. Classifier performance

Table 1 presents the result obtained for EXP-1 and Table 2 obtained for EXP-2. Each of the SVM classifiers use a different sub-combination of signals. Best results on the test set for EXP-1 and EXP-2 were obtained using a combination of audio, actigraphy and ODI, achieving an accuracy of up to 88.4% and 92.3% respectively.

3.2. App framework

SleepAp is composed of five modules: 1: "About apnoea" includes information about the disease, its manifestation and treatments. This module also allows the patient to listen to healthy, snoring and OSA breathing sounds; 2: "Questionnaire" is an adapted version of the clinically-validated STOP-BANG questionnaire [11] which provides individuals with an initial screening tool for assessing their chance of having OSA; 3: "Record" guides the user through the setup of the microphone, armband, and pulse oximeter and records the signals of interest (audio, body movement, body position and PPG). Signals are recorded for four continuous hours starting 30 minutes after the user initiates recording (to allow time to fall asleep). The recorded signals and the answers from the questionnaire are stored on the phone; 4: "Analyse" applies signal processing algorithms to extract features from the signals and outputs a probability of the patient having OSA or not, as well two independent scores (ODI and the STOP-BANG questionnaire score); 5: "View previous" displays the output of previous recordings, allowing for treatment monitoring. Figure 2 shows sample screenshots of the app. Computational time for deriving the features and classifying a patient based on a 4 hour recording using a Google Nexus 7 was on average 13.2 seconds $\pm$ 1.2 s (20 runs).

Table 2. EXP-2, SVM classification on clinical data: {Normal, Snorer} vs {Moderate and Severe OSA}

	Training set (646)			Test set (104)		
	Ac	Se	Sp	Ac	Se	Sp
ODI	90.9	87.8	93.2	90.4	83.3	100.0
AU	81.7	68.0	92.1	77.9	66.7	93.2
AC	68.0	44.6	85.6	72.1	53.3	97.7
DE	70.7	63.7	76.1	73.1	61.7	88.6
AU+AC	81.3	71.9	88.3	79.8	70.0	93.2
AU+AC+ODI	91.3	88.1	93.8	92.3	86.7	100.0

Maximum accuracy is indicated by an underline.

4. Discussion and conclusion

The system is designed to satisfy the needs and requirements of different users. If used as a simple home prescreening tool by individuals then it may not be realistic to make use of a pulse oximeter; the system should be low cost and use the sensors that are already available in smartphones. Moreover, if a pulse oximeter is not connected accurately, the resulting noisy signal can actually degrade the performance of the screening system and thus the design should be focused on users with relatively little to no training in physiological monitoring. If used as a screening system to replace home kits where the devices are given to the patients with some direction about how to use them, then additional medical devices (such as the pulse oximeter) can be used to improve the accuracy of the diagnosis.

SleepAp is the first step towards a clinically-validated automated sleep screening tool which is available on a mass scale and at negligible cost to smartphone users. The general approach of selecting features which can be recorded from smartphones (which are ubiqutious and continuing to grow in popularity) and training a machine

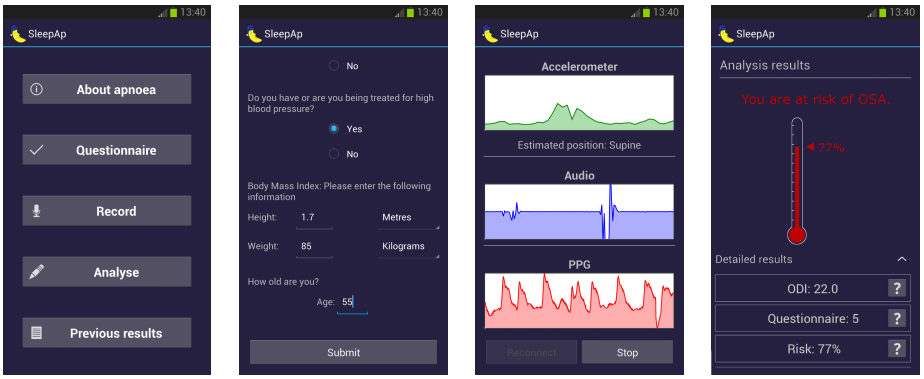

Figure 2. SleepAp screenshots. From left to right: main menu, STOP-BANG questionnaire, recording and analysis modules.

learning algorithm to screen populations is likely to be of tremendous help in low resource countries with little access to expensive screening procedures.

Nevertheless, all individuals included were referred for screening by their GP and, most importantly, data were not recorded using smartphones. As such, they do not represent the typical user and recording setting for the mobile phone use-case, so retraining is likely to be required on mobile phone data and a more general population before any related system is used in practice.

To our knowledge, this is the first fully-automated smartphone medical diagnostic system which uses more than simple logic branching and decision support. Novel signal processing techniques together with a machine learning approach provide a simple home system which can output a probability of a user requiring hospital screening for OSA. Such software provides a new, easy-to-use, low-cost and widely available modality for sleep apnoea screening at negligible cost to smartphone users. Moreover, it provides a generic framework for health screening in low resource environments. We are releasing the code under an open source license so as to encourage other users to test it on sub-populations and continue to develop the platform (see **http://sleepap.herokuapp.com/**).

Acknowledgements

JB is supported by the UK Engineering and Physical Sciences Research Council and the Balliol French Anderson scholarship fund. AR, JD, NP, AH acknowledge the support of the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). AH acknowledges Capes Foundation, Ministry of Education of Brazil, Brasilia - DF 7004-020, Brazil (Proc. No BEX 0725/12-9).

References

[1] Epstein LJ, Dorlac GR. Cost-effectiveness analysis of nocturnal oximetry as a method of screening for sleep apnea-hypopnea syndrome. CHEST Journal 1998;113(1):97–103.

[2] Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997;20(9):705–706.

[3] Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiological Measurement 2013;34(7):R29.

[4] Canalys. Google's Android becomes the world's leading smart phone platform. Available from: http://www.canalys.com, January 2011.

[5] Roebuck A, Monasterio V, Gederi E, Osipov M, Behar J, Malhotra A, Penzel T, Clifford GD. Signal processing of data recorded during sleep. Submitted to Phys Meas 2013;.

[6] Roebuck A, Clifford GD. Multiscale Entropy Applied To Audio Data For Classifying Obstructive Sleep Apnoea Patients. In Am Thor Soc 2012 Conference Abstracts, San Francisco, USA. May 18-23, 2012; .

[7] Higgins N. supervised by Clifford GD The detection of obstructive sleep apnoea using a mobile phone. MSc Thesis University of Oxford Trinity term 2012;.

[8] Behar J. supervied by Clifford GD. Analysis of accelerometer data for apnea screening. MSc Thesis University of Oxford August 2011;.

[9] Costa M. Multiscale entropy analysis. http://physionet.org/physiotools/mse/, August 2004.

[10] Chang C, Lin C. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2011;2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[11] Chung F, Yegneswaran B, Liao P, Chung S, Vairavanathan S, Islam S, Khajehdehi A, Shapiro C. STOP Questionnaire: A Tool to Screen Patients for Obstructive Sleep Apnea. Anesthesiol 2008;108:812–21.

Address for correspondence:

Behar Joachim - joachim.behar@eng.ox.ac.uk