Identification of Sleep Apnea Events using Discrete Wavelet Transform of Respiration, ECG and Accelerometer Signals

Kevin T. Sweeney, Edmond Mitchell, Jennifer Gaughran, Thomas Kane, Richard Costello, Shirley Coyle, Noel E. O’Connor, Dermot Diamond

Abstract—Sleep apnea is a common sleep disorder in which patient sleep patterns are disrupted due to recurrent pauses in breathing or by instances of abnormally low breathing. Current gold standard tests for the detection of apnea events are costly and have the addition of long waiting times. This paper investigates the use of cheap and easy to use sensors for the identification of sleep apnea events. Combinations of respiration, electrocardiography (ECG) and acceleration signals were analysed. Results show that using features, formed using the discrete wavelet transform (DWT), from the ECG and acceleration signals provided the highest classification accuracy, with an $F_1$ score of 0.914. However, the novel employment of just the accelerometer signal during classification provided a comparable $F_1$ score of 0.879. By employing one or a combination of the analysed sensors a preliminary test for sleep apnea, prior to the requirement for gold standard testing, can be performed.

I. INTRODUCTION

Healthy sleep patterns have long been proven to be essential for maintaining both mental and physical health [20]. Sleep apnea is a common disorder which seriously degrades sleep quality and is characterised by recurrent pauses in breathing (apnea) or by instances of abnormally low breathing during sleep (hypopnoea). Apnea events can be classified into two main groups: obstructive apnea (OA) is the cessation of airflow due to the collapse of the upper airway while central apnea (CA) is due to the lack of neural input from the central nervous system [23].

Patients suffering from sleep apnea have been shown to be more prone to a number of different health complications. The associated reduction in sleep quality has been proven to increase the likelihood of accidents both at home and at work [11][21] whilst patients suffering from sleep apnea have also been shown to be more susceptible to significant health risks, including cardiovascular related deaths [3][14]. In conjunction with this increased health risk is the high cost to national healthcare systems. In 1994, the U.S. National Commission on Sleep Disorders Research [14] estimated that the annual cost to the American taxpayer, for disorders related to sleep apnea, was in excess of $42 million. This figure indicates the potential benefit of employing low cost systems capable of accurately detecting sleep apnea events.

To date there have been a number of independent systems employed to aid in sleep apnea classification [7][16], however the widely accepted gold standard diagnostic method is known as a polysomnograph (PSG). During a PSG the patient attends a specialised sleep clinic and is monitored over the course of a single night using multiple different monitoring systems. These systems commonly include a measurement of the heart (electrocardiogram (ECG)), the skeletal muscles (electromyogram (EMG)), eye movement (electrooculogram (EOG)), respiratory airflow, respiratory effort and oxygen saturation of the blood (PPG) [16]. The position of the patient in the bed and the snoring level is also often recorded to aid in the diagnosis.

This requirement for multiple recording modalities results in a high cost (a private patient test can cost up to €1,000) as well as a large quantity of data which must be examined post-recording by a trained technician for each patient. This has commonly lead to long waiting lists for patients requiring testing. Current research is continuing to examine the use of less complex systems to accurately classify sleep apnea events. Examples of this research includes the use of the ECG to classify between obstructive and central apnea events [8][24] and the use of accelerometers placed on the suprasternal notch to screen for sleep apnea events [13].

This paper examines the ability to accurately identify sleep apnea events by combining data from three different types of sensors. Classification results are obtained using a combination of electrocardiogram (ECG), respiration and acceleration sensors. Results show that the use of one or a combination of these simple sensors can provide accurate results in the classification of sleep apnea events.

II. MATERIALS AND METHODS

To allow for an accurate measure of the efficacy of the proposed classification techniques, the ECG, respiration and acceleration data was recorded in conjunction with standard overnight PSG. The three additional signals were recorded using a “Smartex Wearable Wellness System (WWS)” chest strap [19]. This section begins with a brief description detailing the volunteer test patients and, following this, both the PSG and “Smartex” recording systems are described.
A. Participants

Data was collected using the “Smartex WWS” [19] from 5 adult patients (3 female, mean age 52 years, standard deviation 5.89 years) during routine PSG recording. All analysed data was recorded as part of routine sleep apnea diagnosis. Ethical approval was obtained through the Beaumont Research Ethics Committee (Study 12/63, “Use of a wearable sensor system to monitor sleep apnoea”, Shirley Coyle (DCU), Dermot Diamond (DCU), Richard Costello (RCSI). Patient data was analysed post recording and an exclusion criterion was implemented based on a positive indication of the presence of sleep apnea events. One female patient was discovered to not suffer from sleep apnea and was thus excluded from the study.

B. Protocol & Data Acquisition System

Patient data was recorded overnight in St. Joseph’s Clinic, Raheny, Dublin, Ireland. During recording the gold standard PSG monitoring was performed while concurrent measurements were made using the implemented “Smartex” system. The gold standard PSG provided accurate information as to the time points and duration of all apnea events observed overnight, allowing for a detailed measure of the efficacy of post classification results using the “Smartex” system.

1) Polysomnograph (PSG): As stated previously, PSG recordings are regarded as the gold standard method for determining the presence of sleep apnea events. During the routine recording, a number of patent physiological signs were monitored including respiration flow, thorax effort, oxygen saturation, heart rate and breaths per minute. The body position and snoring output of the patient were also monitored. These signals provided a database from which a trained clinician could analyse the data, post recording, and manually tag the epochs relating to apnea events. The detected apnea events were tagged as either obstructive apnea, central apnea, hypopnoea or mixed apnea depending on their nature. Any epochs in which one or more of the recorded signals were observed to be noisy were labeled as an artifact epoch. All remaining epochs were then marked as clean. Data was recorded for a period of between 6-8 hours per subject, during which period the patient slept in a bed situated in an isolated room.

2) Smartex: As described in Section I, three separate signal modalities were measured concurrently with the PSG recording. Figure 1 illustrates the “Smartex Wearable Wellness System (WWS)” chest strap [19] used to house the three recording sensors. This WWS is a wearable system based on textile knitted sensors [15]. The electrocardiography signal (ECG) is used to monitor the electrical activity associated with the pumping of the heart. The ECG signal was recorded using two moistened fabric sensors located at either side of the ribcage (Figure 1 (b)). The use of these fabric electrodes eliminates the requirement for adhesive electrodes which can be cumbersome to apply and have been shown to occasionally cause skin irritation [10], while also allowing for unlimited use. The ECG signal was recorded at a sampling rate of 250 Hz.

Fig. 1. Smartex Wearable Wellness System. (a) Respiration sensor positioned at the front centre of the band. Accelerometer located in the CSEM recording module which is housed in the indicated pouch. (b) Fabric ECG electrodes located on the inside of the chest strap.

The acceleration signal was recorded using a tri-axial accelerometer located in the recording module shown in Figure 1. This recording module was securely stored in a pouch located on the front of the chest strap. This accelerometer was capable of determining patient body position as well as being a proxy for the respiration signal due to the movement of the chest. The sampling rate of the accelerometer was set at 25 Hz.

The respiration signal was also monitored using the chest strap. The respiration signal was recorded using a piezoresistive knitted textile stretch sensor located on the front of the chest strap as can be seen from Figure 1 (a). As the subject both inhales and exhales, the force on the stretch sensor alters, presenting a recordable change in resistance. This resistance change can then be related to a change in lung volume. The respiration sensor was also sampled at the lower frequency rate of 25 Hz. All data recorded was stored on an on-board SD card for post processing using MATLAB®.

The data from the Smartex chest strap was synced against the PSG data post recording using the available information regarding the patient’s body position. Any change in patient body position could be determined using the accelerometer data, the time points relating to positional change could be aligned with the positional data available from the PSG analysis.

The “Smartex WWS” chest strap was secured to the patient’s chest below the pectoral muscles and above the base of the ribcage using the available velcro. This allowed for a similar position for both the male and female patients. It should be noted that the only pre-test requirement was the wetting of the ECG electrodes. This allowed for a very quick and easy application of the recording sensors.

III. CLASSIFICATION

This section will describe the post processing performed on the data. Section III-A describes the initial filtering, tagging and windowing performed on the data. Section III-B next describes the discrete wavelet transform (DWT) which was used to generate the signal features used for classification while Section III-C describes the regression tree classifier used to classify the individual signal epochs as either clean or as containing an apnea event. Finally, Section III-D describes the F1 score efficacy metric used to determine the accuracy of the classification algorithm.
A. Data Set

A number of processing steps were completed prior to the feature selection stage of classification. Initially each signal was filtered to remove any unwanted frequencies. The ECG signal was bandpass filtered between 0.05 Hz and 20 Hz. The DC components were filtered to remove any DC offset from the signal while the ECG frequencies below 20 Hz have been shown to contain the majority of the desired ECG components. Frequencies above this frequency are required if detection of arrhythmias is desired [9]. The respiration signals bandpass cut-off frequencies were 0.05 Hz and 0.8 Hz. This upper limit was chosen as the maximum frequency of human breathing is unlikely to exceed this value [18]. Finally, the accelerometer data was low-pass filtered with a cut-off frequency of 0.8 Hz to again be capable of representing the accelerometer data while the ECG frequencies below 20 Hz were filtered to remove any unwanted frequencies. The ECG signal was bandpass filtered between 0.05 Hz and 20 Hz.

Following the windowing of the data, a total number of 1082 windows containing an apnea event were available. As the number of clean windows (9087) was much higher than the number of apnea contaminated windows, a random selection of 864 clean windows was chosen to ensure the data was balancanced.

B. Feature Extraction using the Discrete Wavelet Transform

Feature extraction was completed using the discrete wavelet transform (DWT) as this has been shown previously to provide accurate classification results [8]. The Wavelet transform operates by decomposing a signal into a number of time shifted and scaled versions of a selected mother wavelet. The wavelet expansion of a signal \( x(t) \) can be written as [1]:

\[
x(t) = \sum_{k} c_{Mk} \varphi_{Mk} + \sum_{j=1}^{M} \sum_{k} d_{jk} \psi_{jk}(t),
\]

where

\[
c_{jk} = \int x(t) \psi_{jk}(t) dt
\]

are called the scaling coefficients. The scaling functions \( \varphi \) are defined as:

\[
\varphi_{jk}(t) = \frac{1}{\sqrt{2^j}} \varphi \left( \frac{t-k2^j}{2^j} \right).
\]

The wavelet coefficients \( d \) are given by:

\[
d_{jk} = \int x(t) \psi_{jk}(t) dt
\]

with the wavelet functions \( \psi \) defined as:

\[
\psi_{jk}(t) = \frac{1}{\sqrt{2^j}} \psi \left( \frac{t-k2^j}{2^j} \right).
\]

The details \( D_j(t) \) and approximations \( A_j(t) \) of the wavelet transform at each level \( j \) can thus be defined as:

\[
A_j(t) = \sum_{k=-\infty}^{\infty} c_{jk} \varphi_{jk}(t)
\]

\[
D_j(t) = \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}(t).
\]

The energy of the signals at each decomposition level were chosen as the features for classification. To ensure independent features, the signals chosen to generate the features were the detail signals at each level and the final approximation signal. The energy of each signal \( s(n) \) was calculated as:

\[
E = \sum_{n} |s(n)|^2.
\]

In this paper the Daubechies 5 mother wavelet was implemented [4] and the signals were decomposed to the 5th level. Therefore for each window of data, only 6 features were calculated for each signal modality. Additional tests were run whilst applying additional signal features, but accuracy results were not observed to improve significantly enough to warrant their inclusion.

C. Choice of Classifier

Classifiers refer to mathematical functions which are employed to map some input data to a specific category or class. A classifier uses a set of features (previously extracted from the input data) to identify from which category a particular new observation belongs. In order to both create and evaluate a classifier, an independent testing and training data-set are required. The training data is first used to create and tune the classifier. This training set generally encompasses a larger proportion of the available data than is used during testing. The testing data is then employed to evaluate the performance of the classifier.

In this work, classification is performed using a decision tree [17]. In a decision tree the class labels (i.e. Apnea/Non-Apnea) act as the leaves and the logical conjunctions (nodes) act as the branches that lead towards class labels. An example of a simple decision tree is presented in Figure 2. When a decision tree is being trained, it analyses the inputted feature set from each individual observation and develops a weighted path to every class label. Therefore, as each observation in
the training set is analysed the tree becomes incrementally refined. Once the decision tree has been trained, any new observation from the test data-set can be assigned to a particular class according to its particular feature set. Classification algorithms are then evaluated by establishing how accurately they can determine the correct class label for each observation in the test data-set. Section III-D describes the efficacy metric used to determine the accuracy of the selected classifier.

The decision tree classification algorithm has previously been implemented successfully on all three of the signals analysed in this paper (acceleration [12], ECG [6] and respiration [22]). This particular classifier was chosen due to its robustness, success in similar work and ease of use.

D. Efficacy Metrics

In order to test the efficacy of the classification procedure the \( F_1 \) score metric was employed. This \( F_1 \) score is a measure of a test’s accuracy; the closer the determined \( F_1 \) score is to unity the more accurate the classification result. Unlike some other biased metrics, the \( F_1 \) score considers both the “precision” and the “recall” of the classification result.

\[
F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}
\]

Precision is calculated as the number of correct results divided by the number of total results while recall is the number of correct results divided by the number of results that should have been returned positive. These metrics are often described in terms of the metrics true positive (\( T_p \)), false positive (\( F_p \)) and false negative (\( F_n \)).

\[
\text{Precision} = \frac{T_p}{T_p + F_p} \tag{9}
\]

\[
\text{Recall} = \frac{T_p}{T_p + F_n}
\]

A true positive is returned when the classifier accurately detects an apnea event, a false positive results from the classifier identifying that an apnea has occurred when in reality the signal window is clean. A false negative is due to the classifier identifying the window as clean when in truth an apnea event is contained in the window.

Ten fold cross-validation was used in this work to assess how accurately our predictive model will perform in practice. In ten fold cross-validation, the original data-set is randomly divided into ten equal sub-sets. One sub-set is selected as the testing data and the remaining sub-sets are used as the training data. This process is repeated ten times with each sub-set acting as the testing data once. The ten results are then averaged to give a single accuracy estimation for the classifier. Cross-validation is a way to predict the fit of a model to a hypothetical validation set when an explicit validation set is not available.

IV. RESULTS & DISCUSSION

The purpose of the analysis performed in this paper was to determine the highest classification results obtainable when classifying between apnea and non-apnea events when using only simple and easily available sensors. The signals from the three independent sensor modalities (ECG, respiration and acceleration) were available to generate the feature set from which the class regression tree classifier was trained. Due to the availability of the three signals, seven individual combinations of feature sets could be obtained, as can be seen from Table I. A decision was made to use an early fusion approach to classification although late fusion was also considered. Investigating the benefits of this (if any) will be the subject of future work.

Employing information from all three signals (i.e. 1 from Table I) resulted in an 18 element feature set, using a combination of any two of the signal modalities (i.e. 2, 3 or 4) produced 12 features while only 6 features were available when using the signals independently (i.e 5, 6, 7). Table I presents the \( F_1 \) score results obtained when employing the seven different feature sets. Figure 3 presents the results from Table I visually and in descending order.

<table>
<thead>
<tr>
<th>#</th>
<th>Signals Employed</th>
<th>( F_1 ) Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Respiration ECG Acceleration</td>
<td>0.912</td>
</tr>
<tr>
<td>2</td>
<td>Respiration ECG &quot;</td>
<td>0.831</td>
</tr>
<tr>
<td>3</td>
<td>Respiration &quot; Acceleration</td>
<td>0.890</td>
</tr>
<tr>
<td>4</td>
<td>&quot; ECG Acceleration</td>
<td>0.914</td>
</tr>
<tr>
<td>5</td>
<td>Respiration &quot; &quot;</td>
<td>0.750</td>
</tr>
<tr>
<td>6</td>
<td>&quot; ECG &quot;</td>
<td>0.830</td>
</tr>
<tr>
<td>7</td>
<td>&quot; &quot; Acceleration</td>
<td>0.879</td>
</tr>
</tbody>
</table>

A number of interesting conclusions can be inferred from the information presented in Figure 3. Primarily, the results obtained provide a high classification accuracy, similar to that achieved by deChazal et al. (89 %) [5] and Yilmaz et al. (80-90 %) [24]. This result strongly suggests that the employment
of the simple wearable sensors with a low number of features is a viable option for sleep apnea classification.

![Graph](image)

Fig. 3. Impact of changing the input signals on the determined $F_1$ score. It can be observed that the removal of the acceleration data significantly degrades the calculated $F_1$ score.

Of the three signals, the respiration signal can be seen to have the lowest individual classification accuracy whilst also adding little in terms of classification improvement when added to other signal modalities. The inclusion of the respiration signal features with the acceleration signal features sees a rise in classification accuracy of only 0.011, whilst its inclusion with the ECG signal features results in a lower accuracy improvement of just 0.001. Interestingly the highest classification results are obtained when the respiration signal is omitted during feature selection. This result may be due to the accelerometer signal being capable of more accurately representing the subject’s respiration and thus the respiration signal obtained using the stretch sensor does not provide any additional useful information. Further, a change in the patient’s position can cause a significant variation in the force placed on the respiration stretch sensor, reducing its functionality.

Of the three sensors, the accelerometer signal provided the best performance when employing the sensors independently. By using only a single sensor, the accuracy dropped by a mere 0.035 compared to the highest accuracy obtained. This result demonstrates the realistic opportunity of solely using the accelerometer signal to attempt to classify sleep apnea events. Accelerometer sensors are very cheap to produce and can easily be attached to the subject using either a strap or an adhesive. Also, as the sensors do not require direct contact with the skin (as ECG does), it is less likely to output inaccurate or false results over a full night of testing due to the motion of the subject causing movement of the sensor with respect to the skin.

The combination of the accelerometer signal with the ECG signal provides the best results when employing the full “Smartex” system. This combination results in an accuracy of 0.914 allowing for a high confidence rate when applied over a large number of apnea events.

V. CONCLUSION AND FUTURE WORK

In this paper, an analysis into the use of cheap and easy to use respiration, ECG and accelerometer sensors for the classification of sleep apnea events has been investigated. Results show that the accelerometer signal provides the best results when a single sensor is used to classify the data. This result has, to the authors knowledge, never previously been highlighted. Many authors have discussed the sole use of an ECG signal to classify the data [5][8][24] however results shown within this paper instead propose the use of a simple accelerometer signal for classification purposes. However more accurate results are available if features from the ECG data is also included in the analysis. The respiration signal was determined to not improve the classification results and thus should not be included in analysis.

The ability to accurately classify the sleep apnea data using low cost and easy to use sensors can allow for both cheap pre-PSG testing and for post-diagnostic monitoring of the efficacy of interventions on an individual’s condition. As PSG testing can be expensive (approx. €1,000), the ability to pre-test for sleep apnea using a cheap and accurate system could allow the tests to be preliminarily carried out outside of the clinic environment. This pre-test would reduce the number of patients who do not suffer from sleep apnea applying for PSG testing, thus saving them money. The availability of this pre-test would also reduce the waiting time for PSG tests which currently can be up to 6 months between initial referral and testing. As these systems continue to become more accurate, the requirement for final analysis using the full PSG system may eventually become redundant.

Although the system proposed provided high accuracy results when classifying apnea events there is still a requirement to be able to classify between the different apnea events. This will be the focus of future work as additional data is required to ensure that a sufficient number of the separate apnea events are available for analysis.

REFERENCES


