TOPICAL REVIEW

Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review

To cite this article: M B Uddin et al 2018 Physiol. Meas. 39 03TR01

View the article online for updates and enhancements.

Related content
- Sleep apnea: a review of diagnostic sensors, algorithms, and therapies
  Mehdi Shokoueinejad, Chris Fernandez, Emily Carroll et al.
- Detection of REM/NREM snores in obstructive sleep apnoea patients using a machine learning technique
  Shahin Akhter and Udanthi R Abeyratne
- Automatic classification of apnea/hypopnea events through sleep/wake states and severity of SDB from a pulse oximeter
  Jong-Uk Park, Hyo-Ki Lee, Junghun Lee et al.
Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review

M B Uddin, C M Chow and S W Su

Faculty of Engineering and Information Technology, School of Biomedical Engineering, University of Technology Sydney, Sydney, Australia
Faculty of Health Sciences, Exercise, Health and Performance Research Group, The University of Sydney, Lidcombe, Australia
E-mail: steven.su@uts.edu.au

Keywords: obstructive sleep apnea, central apnea, polysomnography, machine learning, binary decision-making

Abstract

Objective: Sleep apnea (SA), a common sleep disorder, can significantly decrease the quality of life, and is closely associated with major health risks such as cardiovascular disease, sudden death, depression, and hypertension. The normal diagnostic process of SA using polysomnography is costly and time consuming. In addition, the accuracy of different classification methods to detect SA varies with the use of different physiological signals. If an effective, reliable, and accurate classification method is developed, then the diagnosis of SA and its associated treatment will be time-efficient and economical.

This study aims to systematically review the literature and present an overview of classification methods to detect SA using respiratory and oximetry signals and address the automated detection approach.

Approach: Sixty-two included studies revealed the application of single and multiple signals (respiratory and oximetry) for the diagnosis of SA.

Main results: Both airflow and oxygen saturation signals alone were effective in detecting SA in the case of binary decision-making, whereas multiple signals were good for multi-class detection. In addition, some machine learning methods were superior to the other classification methods for SA detection using respiratory and oximetry signals.

Significance: To deal with the respiratory and oximetry signals, a good choice of classification method as well as the consideration of associated factors would result in high accuracy in the detection of SA. An accurate classification method should provide a high detection rate with an automated (independent of human action) analysis of respiratory and oximetry signals. Future high-quality automated studies using large samples of data from multiple patient groups or record batches are recommended.

Abbreviations

AE Abdominal effort
AF Airflow
AHI Apnea-hypopnea index
ANN Artificial neural network
AUC Area under ROC curve
BHC Binary hierarchical
CA Clustering algorithm
CART Classification and regression tree
CGE Center of gravity engine
CSA Central sleep apnea
ECG Electrocardiography
EEG Electroencephalography
EMG Electromyography
EOG Electrooculography
FCM Fuzzy c-means
FIS Fuzzy inference system
1. Introduction

Breathing disorders, defined by disturbances of the normal breathing process, can cause the development of central nervous, organic, physical, and metabolic disorders (Várady et al 2002). There are various types of breathing disorders, but probably the most common type is sleep apnea (SA). SA affects about 2%–5% of the total human population and over 30% of the elderly male population (Köves 1999). The prevalence of SA is approximately 3% in children (Chang and Chae 2010), 9% in women, and 17% in men ranging in age from 50–70 years (Peppard et al 2013).

SA can be classified into three categories—obstructive, central, and mixed. An episode of obstructive sleep apnea (OSA) occurs when there is complete obstruction of the air passage and cessation of airflow (AF) but with continued respiratory efforts (abdominal and thoracic) against a closed airway (Sezgin and Tagluk 2009). A central sleep apnea (CSA) episode occurs when there is complete cessation of breathing with no respiratory efforts (Sezgin and Tagluk 2009). Both these events must last 10s or more during sleep for them to be scored as such in adults (Guilleminault and Partinen 1990, Köves 1999, Sezgin and Tagluk 2009). Mixed sleep apnea (MSA) is defined by a central respiratory pause followed, in a relatively short duration of time, by obstruction of the airway (Sezgin and Tagluk 2009). There are major differences in sleep and respiratory physiology between children and adults. SA in children is defined as the cessation of AF for at least two respiratory cycles, and it is different from the definition of apnea in adults (Alsubie and BaHammam 2017). Shallow breathing, also known as hypopnea, is often associated with partial obstruction of the upper airway. Although less severe than SA, it can similarly cause desaturation in blood oxygen level.

OSA is the most prevalent sleep disorder and affects 2%–4% of the adult population (Berger et al 1997) with CSA and MSA being relatively less predominant. The severity of apnea and hypopnea is measured by the number
of episodes per hour, using several indexes such as the apnea index (AI), hypopnea index (HI), and apnea-hypopnea index (AHI) or respiratory disturbance index (RDI). In individuals with apnea or hypopnea throughout the night, there can be 5–15 episodes per hour in mild cases, 15–30 episodes per hour in moderate cases, and more than 30 episodes per hour in severe cases in adults (Kryger 2000). The severity of SA is also different in children with 1–5 episodes per hour classified as ‘mild’, 5–10 episodes per hour as ‘moderate’, and more than 10 episodes per hour as ‘severe’ (Alsubie and Bahammam 2017, Kljajic et al 2017).

These respiratory disturbances may cause arousal from sleep (Gleeson et al 1990). Sleep disruption and excessive daytime sleepiness are the most common presenting complaint (Slater and Steier 2012). Other major symptoms include snoring, fatigue, falling asleep, headaches, weight gain, and memory loss. Because of such complications, the quality of life can be significantly decreased as well as increased major risk of associated health problems (Ben-Israel et al 2010). In addition, SA may go undiagnosed for years because of the person’s unawareness (Kryger et al 1996) with supportive statistics of around 100 million people in the world who are suspected to have SA but are undiagnosed (Bousquet et al 2007). In this regard, several challenges regarding SA diagnosis, assessment, and treatment are major concerns in public health.

Polysomnography (PSG), considered the gold standard and a reliable method for SA diagnosis, is a multichannel signal recording process throughout the night. The major parameters of a standard diagnostic nocturnal PSG (Standards of Practice Committee of the American Sleep Disorders Association 1997) include recording of electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), nasal airflow (NAF), abdominal and/or thoracic efforts (AE and/or TE), body position, snore sounds, and blood oxygen saturation (SaO₂). In portable devices, a limited-channel PSG is also frequently used for apnea diagnosis. A limited-channel PSG includes only the following signal channels: NAF, abdominal and/or thoracic movements, SaO₂, and heart rate (HR) derived from ECG (Berger et al 1994).

An SA test with PSG involves an overnight diagnostic study in a sleep laboratory, where electrodes are attached to the skin surface and scalp to record physiological signals. Patients may not be able to sleep well due to wires hanging from one’s head and body. Sleep technologists monitor and manually review the overnight study for designating sleep stages and apneic type and length according to the American Academy of Sleep Medicine (AASM) guidelines (Grigg-Damberger et al 2007). This process is time-consuming and costly and requires skilled personnel. Automatic analysis has been developed (Cabrero-Canoza et al 2003, 2004, de Chazal et al 2010, Burgos et al 2010, Alsubie and Bahammam 2017) obviating manual scoring. The limited precision (low correlation coefficient between expert observations and the output of automatic analysis) makes it difficult to validate these systems in terms of detection of apneic events. Because of the aforementioned economic factors associated with apnea detection and limited precision of automatic analysis processes, a need has arisen for developing a system that can detect apneic events with acceptable accuracy with minimal physiological signals (Bucklin et al 2010, Burgos et al 2010, Gao et al 2012, Hamada and Masahiro 2012).

Physiological signals, also referred as biological or biomedical signals, are elemental in SA diagnosis. They are the measurements or recordings that are generated in the physiological process of human beings, e.g. HR, respiratory frequency, skin conductance, electric muscle current, and brain electrical activity. Respiratory signals (NAF/pressure) along with efforts generated by respiratory muscles (thoracic and abdominal effort (AE) manifest as pressure waves), and SaO₂ are fundamental signals employed to detect SA. Obstructive, central, and mixed apnea types can be easily distinguished by AF, thoracic, and AEs. Besides, SA severity (mild, moderate, and severe) can be calculated using the number of events per hour. In addition, every apneic event of 10 s or more results in a reduction of oxygen level. Thus, the level of oxygen desaturation is applied to detect apnea/hypopnea cases. Other physiological signals such as EEG, ECG, and snore sound are also used to detect SA but the reliability of SA detection using these signals is very poor.

Many researchers have employed different classification methods to detect SA using physiological signals. The common signals used include SaO₂ (Zamarrón et al 2001, Oeverland et al 2002), AF (Nazar et al 2001, Morsy and Al-Ashmouny 2005), snore sounds (Ben-Israel et al 2012), ECG (Travieso et al 2014, Song et al 2016), EEG (Tagluk and Sezgin 2011), or a combination of these signals (Várady et al 2002, Kaimakamis et al 2016, Huang et al 2017). The cost of the detection system is proportional to the number of sensors used to collect physiological signals. Many studies have used only one physiological signal to detect SA (Mendez et al 2010, Selvaraj and Narasimhan 2013), whereas others have used multiple physiological signals (Álvarez-Estévez and Moret-Bonillo 2009, Kaimakamis et al 2009, Otero et al 2012, Al-Mardini et al 2014). Reducing the processing cost is not the primary target of SA detection, while the accuracy of the system designed is the first priority.

Detecting SA using the main signals, respiratory and oximetry, would be more realistic and result in better detection accuracy. There are several reasons for selecting these signals. Firstly, the manual scoring of SA events is based on respiratory and oximetry signals, where data segments or epochs are annotated using these signals. Secondly, SA detection that uses the additional signals of EEG, ECG, and snore sound is done using the same annotations based on these signals. Finally, the detection accuracy found in different classification methods is the measure that is based on these annotations.
In addition, adults and children are two distinct entities in SA detection. Sleep architecture, respiratory physiology, apnea definition, and apnea severity in adults differ from children/pediatric subjects (Alsubie and BaHammam 2017, Kljajic et al 2017). Besides, the algorithms for SA detection in pediatric subjects are quite different and usually need special consideration or criteria to obtain better detection results. According to these differences between adults and pediatric subjects, this review focuses only on the adult population.

Detection of SA using respiratory and oximetry signals often includes a four-stage methodology as illustrated in figure 1: acquisition of respiratory and oximetry signals, features extraction, features selection, followed by apnea detection.

Respiratory and oximetry signals are the most effective physiological signals on which a reliable and accurate SA detection system is based. To design such a system, a review of the existing literature on respiratory and oximetry signals has shown greater acceptability in SA detection. However, the benefits, drawbacks, and challenges associated with the use of these signals as well as the existing classification methods in SA detection are unknown. Thus, this systematic review will address these issues to advance the techniques of SA detection to enable the development of a reliable and accurate system. In addition, this paper will address the rationales for and the process of decision-making on the multiple SA scenarios, including their concept, model, performance, plus beneficial and challenging outcomes.

The main research questions were (1) Which respiratory and/or oximetry signals provide the best discriminatory support for decision-making on SA? (2) Which classification methods result in high accuracy in SA detection with respiratory and oximetry signals? (3) What are the beneficial and challenging effects associated with the use of these physiological signals and classification methods?

2. Methods

2.1. Inclusion and exclusion criteria
Studies were included in this review if they met the following criteria: (1) presented a method or systematic approach to detecting SA, (2) written in English, (3) included adult participants only, (4) classification methods based on the use of one or more respiratory and oximetry signals, (5) decisions made on normal or apnea (or different classes of apnea or their severity), and (6) presented definitive overall detection results in the form of accuracy, sensitivity, specificity, and other parameters. These criteria were also applied to studies obtained from cross-reference tracking. Studies that satisfied the above criteria were extracted and included in this review. Articles from conference proceedings were reviewed critically and only extended versions that were published as journal articles were included. Studies were also excluded even though they met the above inclusion criteria: (1) case report of a single subject, and (2) studies where participants have co-morbidities of chronic heart and kidney diseases, diabetes, stroke, etc.

2.2. Search strategy
The phrase ‘sleep apnea detection’ is inter-related to the research fields of health, engineering (biomedical), and information technology. SA is a medical or health complication where knowledge of engineering and information technology is applied to detect or solve this problem. In this regard, the selection of specific databases to extract related articles is a crucial factor. A systematic search was conducted on the following five major electronic databases that are basic sources of articles in the fields of health, engineering, and information technology: Medline (Ovid), Scopus (Elsevier), ACM Digital Library, IEEE Xplore Digital Library, and ProQuest Science and Technology. Studies published in English from January 2001 to July 2017 were included in this study according to the inclusion and exclusion criteria mentioned above.

Databases search was performed using the following words or phrases and all possible combinations: apnea or SA or obstructive SA or SA-hypopnea syndrome or sleep disordered breathing, and the noun or verbal form of classification or detection or identification or prediction or recognition or screening. The following limiting conditions were applied as well during the search: English language, adult human subject, and the stated range of years of publication. All references found in five databases were imported to EndNote for quick manual screening after deleting duplicates. Thus, identified articles were screened for eligible studies. Detailed investigation of eli-
Identification

Relevant studies identified through database searching (n = 4111)

Screening

Studies screened (n = 4111)

Eligibility

Full-text studies assessed for eligibility (n = 148)

Studies excluded (n = 3963)

Cross-reference tracking (n = 7)

Full-text studies excluded (n = 93)

Inclusion

Studies included for qualitative synthesis (N = 62)

Figure 2. Flow diagram of the systematic review process. The combined electronic searches identified 4111 studies. Quick screening of titles and abstracts excluded 3963 studies due to irrelevancy. The remaining 148 full-text articles were eligible for detailed investigation. A quick manual search of the bibliographies of the mentioned full-text articles were performed to extract eligible additional references and new full-text studies. In this way of cross-reference tracking, seven new full-text articles were added. Of 155 full-text articles, 93 failed to satisfy the eligibility criteria. The remaining 62 full-text articles that met the inclusion criteria but did not meet the exclusion criteria were included for qualitative synthesis.

Table 1. Parameters for evaluating metrics used in classification methods.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (Ac)</td>
<td>( \frac{TP + TN}{TP + TN + PP + FN} \times 100% )</td>
</tr>
<tr>
<td>Sensitivity (Se)</td>
<td>( \frac{TP}{TP + FN} \times 100% )</td>
</tr>
<tr>
<td>Specificity (Sp)</td>
<td>( \frac{TN}{TN + FP} \times 100% )</td>
</tr>
<tr>
<td>PPV</td>
<td>( \frac{TP}{TP + FP} \times 100% )</td>
</tr>
<tr>
<td>NPV</td>
<td>( \frac{TN}{TN + FN} \times 100% )</td>
</tr>
<tr>
<td>F-measure</td>
<td>( 2 \times \frac{PPV \times Sensitivity}{PPV + Sensitivity} )</td>
</tr>
<tr>
<td>g-means</td>
<td>( \sqrt{\text{Sensitivity} \times \text{Specificity}} )</td>
</tr>
<tr>
<td>ROC curve</td>
<td>Plot of sensitivity versus 1-specificity</td>
</tr>
<tr>
<td>AUC</td>
<td>Determines which of the used models detect the classes best</td>
</tr>
</tbody>
</table>

a Parameters symbols: AUC = area under ROC curve, NPV = negative predictive value, PPV = positive predictive value, ROC = receiver operating characteristic.

b Definition terms: FN = false negative, FP = false positive, TN = true negative, TP = true positive.

gible studies and their bibliographies retrieved additional pertinent references. Finally, inclusion and exclusion criteria were applied to extract the desired articles for qualitative synthesis. The flow diagram of the systematic review process is presented in figure 2.

2.3 Extraction of study characteristics

The data extracted from the included studies through qualitative synthesis were studies with the year of publication, number of subjects, type of respiratory and/or oximetry signals used, main decision, classification methods, and metrics for classification method evaluation (detection rate). The metrics (based on per-subject or per-recording and per-segment or per-epoch detection) appear in the last four columns of table 2. The parameters and equations used to evaluate the metrics (accuracy, sensitivity, specificity, and other parameters) are set out in table 1. True positive (TP) represents a positive input detected as positive, whereas true negative (TN) represents
Table 2. Included studies to detect SA based on respiratory and oximetry signals.

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of subjects</th>
<th>Signal</th>
<th>Decision</th>
<th>Classification method</th>
<th>Metrics&lt;sup&gt;ac&lt;/sup&gt; (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nazeran et al (2001)</td>
<td>9</td>
<td>AF</td>
<td>A/H</td>
<td>FIS (fuzzy rules)</td>
<td>83.0</td>
</tr>
<tr>
<td>Zamarrón et al (2001)</td>
<td>197</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(77.0) (58.0) (92.0) PPV (87.0)</td>
</tr>
<tr>
<td>Oeverland et al (2002)</td>
<td>100</td>
<td>SaO₂</td>
<td>Mi/Mo-Se</td>
<td>Threshold-based</td>
<td>(88.6) (76.9)</td>
</tr>
<tr>
<td>Várady et al (2002)</td>
<td>16</td>
<td>AF + TE + AE</td>
<td>A/H/N</td>
<td>ANN</td>
<td>90.0</td>
</tr>
<tr>
<td>Várady et al (2003)</td>
<td>6</td>
<td>TE + AE</td>
<td>C/O</td>
<td>PLA</td>
<td>PPV 90.6</td>
</tr>
<tr>
<td>Lee et al (2004)</td>
<td>7</td>
<td>SaO₂</td>
<td>O/N</td>
<td>Threshold-based</td>
<td>96.6 95.7 97.0</td>
</tr>
<tr>
<td>Fontenla-Romero et al (2005)</td>
<td>6</td>
<td>AF + TE</td>
<td>C/M/O</td>
<td>ANN</td>
<td>83.8</td>
</tr>
<tr>
<td>Morsy and Al-Ashmouny (2005)</td>
<td>10</td>
<td>AF</td>
<td>A/N</td>
<td>FIS (fuzzy rules) + CGE</td>
<td>100 97.0</td>
</tr>
<tr>
<td>Tian and Liu (2005)</td>
<td>30</td>
<td>AF + SaO₂</td>
<td>A/H/N</td>
<td>TDNN</td>
<td>83.7 82.9</td>
</tr>
<tr>
<td>Álvarez et al (2006a)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(87.2) (90.1) (82.9) AUC (96.7)</td>
</tr>
<tr>
<td>Álvarez et al (2006b)</td>
<td>74</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(86.5) (95.5) (73.3) AUC (87.0)</td>
</tr>
<tr>
<td>del Campo et al (2006)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(88.3) (82.9) (92.1) AUC (92.4)</td>
</tr>
<tr>
<td>Álvarez et al (2007a)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(87.2) (90.1) (82.9) AUC (92.4)</td>
</tr>
<tr>
<td>Álvarez et al (2007b)</td>
<td>74</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>CA (KMₚ, FCM, hierarchical)</td>
<td>(90.5) (95.5) (83.3)</td>
</tr>
<tr>
<td>Hornero et al (2007)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(82.1) (87.0)</td>
</tr>
<tr>
<td>Marcos et al (2007)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>NN (MLP, RBF+)</td>
<td>(86.3) (89.9) (81.1) AUC (96.0)</td>
</tr>
<tr>
<td>Ng et al (2007)</td>
<td>26</td>
<td>TE + AE</td>
<td>A/N</td>
<td>Threshold-based</td>
<td>93.3 100</td>
</tr>
<tr>
<td>Salisbury and Sun (2007)</td>
<td>34</td>
<td>AF</td>
<td>O/N</td>
<td>Threshold-based</td>
<td>92.4 88.3</td>
</tr>
<tr>
<td>Han et al (2008)</td>
<td>24</td>
<td>AF</td>
<td>A/N</td>
<td>Threshold-based</td>
<td>90.9 84.0</td>
</tr>
<tr>
<td>Marcos et al (2008a)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>NN (KMₚ, FCM, OLS)</td>
<td>(86.1) (89.4) (81.4) AUC (91.0)</td>
</tr>
<tr>
<td>Marcos et al (2008b)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>NN (MLP)</td>
<td>(85.5) (89.8) (79.4) AUC (90.0)</td>
</tr>
<tr>
<td>Marcos et al (2008c)</td>
<td>157</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>GLM</td>
<td>(88.0) (79.6) (100) AUC (92.0)</td>
</tr>
<tr>
<td>Otero et al (2008)</td>
<td>5</td>
<td>AF</td>
<td>A/H</td>
<td>Fuzzy rules</td>
<td>95.0</td>
</tr>
<tr>
<td>Álvarez-Estévez and Moret-Bonillo (2009)</td>
<td>12</td>
<td>AF + TE + AE</td>
<td>A/H</td>
<td>Fuzzy rules</td>
<td>88.5 88.5 AUC 88.0</td>
</tr>
<tr>
<td>Kaimakamis et al (2009)</td>
<td>86</td>
<td>AF + TE + SaO₂</td>
<td>O/N</td>
<td>Decision tree (C4.5)</td>
<td>(84.9)</td>
</tr>
<tr>
<td>Marcos et al (2009a)</td>
<td>187</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>LDA*, QDA, KNN, LRA</td>
<td>(87.6) (91.1) (82.6) AUC (92.5)</td>
</tr>
<tr>
<td>Marcos et al (2009b)</td>
<td>149</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>SVM</td>
<td>(88.0) (84.4) (93.3) AUC (92.1)</td>
</tr>
<tr>
<td>Morillo et al (2009)</td>
<td>117</td>
<td>SaO₂</td>
<td>Mi/Mo-Se</td>
<td>Threshold-based</td>
<td>(90.9) (84.0) (95.0) AUC (95.0)</td>
</tr>
<tr>
<td>Sezgin and Tagluk (2009)</td>
<td>21</td>
<td>TE + AE</td>
<td>C/M/O</td>
<td>ANN</td>
<td>86.8</td>
</tr>
<tr>
<td>Álvarez et al (2010a)</td>
<td>148</td>
<td>AF + SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(84.5) (84.0) (85.4) AUC (90.4)</td>
</tr>
<tr>
<td>Álvarez et al (2010b)</td>
<td>148</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>LRA</td>
<td>(89.7) (92.0) (85.4) AUC (96.7)</td>
</tr>
<tr>
<td>Burgos et al (2010)</td>
<td>8</td>
<td>SaO₂</td>
<td>A/N</td>
<td>Bagging with ADTree</td>
<td>93.0 92.4 93.5 AUC 98.5</td>
</tr>
<tr>
<td>Caseiro et al (2010)</td>
<td>41</td>
<td>AF</td>
<td>O/N</td>
<td>Threshold-based</td>
<td>(81.0) (95.0)</td>
</tr>
<tr>
<td>Marcos et al (2010)</td>
<td>214</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>LDA</td>
<td>(93.0) (97.0) (79.3) AUC (95.0)</td>
</tr>
<tr>
<td>Tagluk et al (2010)</td>
<td>21</td>
<td>AE</td>
<td>C/M/O</td>
<td>ANN</td>
<td>77.9</td>
</tr>
<tr>
<td>Tagluk and Sezgin (2010)</td>
<td>21</td>
<td>AE</td>
<td>C/M/O</td>
<td>ANN</td>
<td>85.6</td>
</tr>
<tr>
<td>Marcos et al (2011)</td>
<td>96</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>Threshold-based</td>
<td>(81.3) (81.3) (81.3) AUC (87.0)</td>
</tr>
<tr>
<td>Guijarro-Berdiñas et al (2012)</td>
<td>6</td>
<td>AF + TE</td>
<td>C/M/O</td>
<td>ANN</td>
<td>93.5 90.3 95.1</td>
</tr>
<tr>
<td>Gutiérrez-Tobal et al (2012)</td>
<td>148</td>
<td>AF</td>
<td>O⁺/O⁻</td>
<td>LRA</td>
<td>(82.4) (88.0) (70.8) AUC (90.3)</td>
</tr>
<tr>
<td>Ottero et al (2012)</td>
<td>10</td>
<td>AF + SaO₂</td>
<td>A/N</td>
<td>Fuzzy rules</td>
<td>90.0</td>
</tr>
<tr>
<td>Gutiérrez-Tobal et al (2013)</td>
<td>148</td>
<td>AF</td>
<td>O⁺/O⁻</td>
<td>MLR, NN (MLP*, RBF)</td>
<td>(91.5) (92.5) (89.5) PPV (94.9)</td>
</tr>
<tr>
<td>Koley and Dey (2013)</td>
<td>36</td>
<td>AF</td>
<td>A/N</td>
<td>SVM</td>
<td>94.9</td>
</tr>
<tr>
<td>Maali and Al-Jumailly (2013)</td>
<td>5</td>
<td>AF + TE + AE</td>
<td>A/N</td>
<td>ANN</td>
<td>AUC 87.0</td>
</tr>
<tr>
<td>Morillo and Gross (2013)</td>
<td>115</td>
<td>SaO₂</td>
<td>O⁺/O⁻</td>
<td>PNN</td>
<td>(93.9) (92.4) (95.9) AUC (96.1)</td>
</tr>
<tr>
<td>Selvaraj and Narasimhan (2013)</td>
<td>200</td>
<td>AF</td>
<td>A/N</td>
<td>Logical algorithm</td>
<td>(83.6) (100) PPV 72.3</td>
</tr>
<tr>
<td>Thommandram et al (2013)</td>
<td>8</td>
<td>TE</td>
<td>A/N</td>
<td>KNN</td>
<td>95.7 88.1 AUC 96.0</td>
</tr>
</tbody>
</table>

(Continued)
a negative input detected as negative. False positive (FP) represents a negative input detected as positive, whereas false negative (FN) represents a positive input detected as negative. The main outcome measurement parameters (accuracy, sensitivity, and specificity) are evaluated based on TP, TN, FP, and FN. Other parameters are positive predictive value (PPV), negative predictive value (NPV), F-measure, g-means, and area under ROC (receiver operating characteristic) curve (AUC). The qualitative analysis of included articles based on respiratory and oximetry signals (single signal or multiple signals) is tabulated in Table 2.

3. Results

A year on year distribution of published articles for SA detection based on respiratory and oximetry signals is shown in figure 3. The highest number of publications (11.3%) was found in the years 2010 and 2013 each. Table 3 displays the number and percent of articles found to detect SA using respiratory and oximetry signals and the number of articles used to make decisions on the types of SA. Of the 62 studies retrieved, 70.97% (44 articles) were categorized based on a single respiratory or oximetry signal, whereas 29.03% (18 articles) were based on multiple respiratory and oximetry signals (table 3). SA detection based on single signals was further

Table 2. (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of subjects</th>
<th>Signal*</th>
<th>Decision*</th>
<th>Classification method*</th>
<th>Metricsd,e (%): Ac</th>
<th>Se</th>
<th>Sp</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al (2013)</td>
<td>40</td>
<td>SaO2</td>
<td>A/N</td>
<td>SVM</td>
<td>90.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morales et al (2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koley and Dey (2014)</td>
<td>34</td>
<td>SaO2</td>
<td>A/N</td>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sánchez-Morillo et al (2014)</td>
<td>115</td>
<td>SaO2</td>
<td>Mi/Mo/Se</td>
<td>BHC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avci and Akbaş (2015)</td>
<td>8</td>
<td>AF + TE + AE</td>
<td>A/N</td>
<td>RFC, AdaBoost, RSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciolek et al (2015)</td>
<td>30</td>
<td>AF</td>
<td>A/H</td>
<td>Envelope detection</td>
<td>95.0</td>
<td>90.0</td>
<td>96.0</td>
<td></td>
</tr>
<tr>
<td>Huang et al (2015)</td>
<td>387</td>
<td>SaO2</td>
<td>O/N</td>
<td>Decision tree</td>
<td>(94.7)</td>
<td>(98.7)</td>
<td>(90.7)</td>
<td>g-means (94.6)</td>
</tr>
<tr>
<td>Jin and Sánchez-Sinencio (2015)</td>
<td>5</td>
<td>AF</td>
<td>A/N</td>
<td>Threshold-based</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gutiérrez-Tobal et al (2016)</td>
<td>317</td>
<td>AF</td>
<td>A/N</td>
<td>AdaBoost (LDA, CART+)</td>
<td>(86.5)</td>
<td>(89.0)</td>
<td>(80.0)</td>
<td></td>
</tr>
<tr>
<td>Kagawa et al (2016)</td>
<td>35</td>
<td>TE + AE</td>
<td>Mi/Mo-Se</td>
<td>Threshold-based</td>
<td>96.4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaimakamis et al (2016)</td>
<td>100</td>
<td>AF + TE + SaO2</td>
<td>O/N</td>
<td>LRM, Decision tree</td>
<td>(88.6)</td>
<td>(92.9)</td>
<td>(71.4)</td>
<td></td>
</tr>
<tr>
<td>Lee et al (2016)</td>
<td>50</td>
<td>AF</td>
<td>A/H</td>
<td>Rule-based algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huang et al (2017)</td>
<td>30</td>
<td>AF + SaO2</td>
<td>A/H</td>
<td>Respiratory events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morales et al (2017)</td>
<td>79</td>
<td>SaO2</td>
<td>A/N</td>
<td>KNN*, LS-SVM</td>
<td>(93.7)</td>
<td>(96.9)</td>
<td>(78.6)</td>
<td></td>
</tr>
<tr>
<td>Rolón et al (2017)</td>
<td>954</td>
<td>SaO2</td>
<td>Mi/Mo-Se</td>
<td>NN (MLP)</td>
<td>(85.8)</td>
<td>(85.6)</td>
<td>(85.9)</td>
<td>AUC (93.7)</td>
</tr>
</tbody>
</table>

* Signal symbols: AF = airflow, AE = abdominal effort, TE = thoracic effort, SaO2 = oxygen saturation or oximetry signal.

* Decision symbols: A/N = apnea or normal, A/H = apnea or hypopnea, A/H/N = apnea or hypopnea or normal, O/N = OSA or normal, O'/O" = OSA positive or OSA negative, C/O = CSA or OSA, C/M/O = CSA or OSA or MSA, Mi/Mo-Se = mild OSA or moderate-to-severe OSA, Mi/Mo/Se/N = mild OSA or moderate OSA or severe OSA or normal.

* Classification methods symbols: ANN = artificial neural network, BHC = binary hierarchical, CA = clustering algorithm, CART = classification and regression trees, CGE = center of gravity engine, FCM = fuzzy c-means, FS = fuzzy inference system, GLM = generalized linear models, KM = k-means, KNN = k-nearest neighbors, LDA = linear discriminant analysis, LRA = logistic regression analysis, LRM = linear regression model, LS-SVM = least squares support vector machine, MLP = multi-layer perceptron, MLR = multiple linear regression, NN = neural network, OLS = orthogonal least squares, PLA = piecewise linear approximation, PNN = probabilistic neural network, QDA = quadratic discriminant analysis, RBF = radial basis function, RFC = random forest classifier, RSS = random subspace, SVM = support vector machine, TDNN = time-delay neural network.

* Metrics: results based on per-recording detection are enclosed by brackets, whereas results without brackets indicate per-epoch detection.

* Metrics symbols: Ac = accuracy, AUC = area under ROC curve, PPV = positive predictive value, Se = sensitivity, Sp = specificity.
clustered as AF signal-based detection (22.58%, 14 articles), TF signal-based detection (1.61%, one article), AE signal-based detection (3.23%, two articles), and SaO₂ or oximetry signal-based detection (43.55%, 27 articles).

In addition, as presented in table 3, the decision-making process is substantiated by the following scenarios: applying a binary decision (in 54 articles, 87.09%) such as apnea or normal, apnea or hypopnea, OSA or normal, OSA positive or OSA negative, CSA or OSA; a three-option decision (in seven articles, 11.29%) such as apnea or hypopnea or normal and CSA or MSA or OSA; and a four-option decision (in one article, 1.62%) such as mild OSA or moderate OSA or severe OSA or normal.

3.1. Decision-making combined with classification methods

Decision-making on the SA types according to different classification methods is tabulated in table 4. Sixty-two studies revealed single or a combination of classification methods that were clustered as follows: machine learning (ML) methods (64.52%), threshold-based methods (27.42%), and other methods (8.06%). A conceptual mind map of different classification methods used in this review is depicted in figure 4.
3.1.1. ML-based classification methods

ML has revolutionized the possibility of dealing with large and complex data sets. Different ML approaches were applied to detect SA using respiratory and oximetry signals. Out of 62, four studies applied multiple approaches of ML for detection purpose. In total, 40 out of 62 studies applied different ML approaches that included 64.52% of the total classification methods. ML approaches were further segmented as follows: neural networks (22.58%), linear methods (14.52%), regularization (1.61%), instance-based (11.29%), clustering (1.61%), dimensionality reduction (3.23%), ensemble learning (6.45%), and decision trees (3.23%).

Neural network (NN) is a powerful tool for data analytics. The aim of artificial neural networks (ANNs) is to perform tasks analogous to biological brains based on the connections among many simple processing elements, known as neurons. These neurons are organized into layers, where outputs from one layer are used as inputs into the following layer. Other neural networks techniques reported in this review are time-delay neural network (TDNN), radial basis function (RBF) neural network (Haykin 1994), multilayer perceptron (MLP) neural network, and the probabilistic neural network (PNN). A TDNN is able to recognize features independent of time-shift (Waibel et al 1989). The RBF neural network is commonly used for modeling nonlinear problems through a fixed nonlinear transformation (Pombo et al 2014). A MLP is a class of feedforward ANN that utilizes a supervised learning technique and can distinguish data that are not linearly separable (Cybenko 1989). Finally, the PNN introduced by Specht (1990) is also a feedforward NN that uses probability distributions.

Support vector machine (SVM) and least squares support vector machine (LS-SVM) (Suykens and Vandewalle 1999) are the most popular linear methods for data analytics. Other linear methods found in this review are the logistic regression analysis (LRA) (Harrell et al 1984), the multiple linear regression (MLR) (Draper and Smith 2014), the linear regression model (LRM) (Efron et al 2004), the piecewise linear approximation (PLA) (Hamann and Chen 1994), and the generalized linear model (GLM) (Nelder and Baker 1972). Another ML approach included in this study is the instance-based that includes the $k$-nearest neighbors (KNN) (Dudani 1976) and the fuzzy rules (Zadeh 1965). Less information is used in dimensionality reduction models to summa-

### Table 4. Decision-making according to classification methods.

<table>
<thead>
<tr>
<th>Decision</th>
<th>ML</th>
<th>Threshold-based</th>
<th>Other methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/N</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>A/H</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>A/H/N</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O/N</td>
<td>5</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>O+/O−</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/O</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/M/O</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mi/Mo-Se</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mi/Mo/Se/N</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>22.58</td>
<td>14.52</td>
<td>1.61</td>
</tr>
</tbody>
</table>

* Decision symbols: A/N = apnea or normal, A/H = apnea or hypopnea, A/H/N = apnea or hypopnea or normal, O/N = OSA or normal, O+/O− = OSA positive or OSA negative, C/O = CSA or OSA, C/M/O = CSA or OSA or MSA, Mi/Mo-Se = mild OSA or moderate-to-severe OSA, Mi/Mo/Se/N = mild OSA or moderate OSA or severe OSA or normal.
rize or describe data. Two such techniques are the linear discriminant analysis (LDA) (Belhumeur et al 1997) and the quadratic discriminant analysis (QDA).

Ensemble learning obtains the overall detection by combining multiple independent models. Several methods observed are the AdaBoost (Bishop 2006), the Bagging (Breiman 1996), the binary hierarchical (BHC) (Casasent and Wang 2005), the random forest classifier (RFC), and the random subspace (RSS). Decision tree (C4.5), clustering, and regularization (elastic net) techniques are also reported in this review.

3.1.2. Threshold-based classification methods

The threshold-based classification methods are dependent on the selection of appropriate values (Coenen and Leng 2007) or different limits (typically support and confidence thresholds). Seventeen articles out of 62 included in this study were based on threshold-based classification methods.

3.1.3. Other classification methods

Five articles employed different algorithms for SA detection. The algorithms included in this study are logical algorithm, event detection algorithm, envelope detection algorithm, rule-based algorithm, and respiratory events algorithm.

3.2. SA detection based on single signals

In this section, SA detection based on the single respiratory signals of AF, TE, and AE, and SaO2 (an oximetry signal) is discussed in the following four sub-sections in light of the published articles.

3.2.1. AF-based detection

The AF signal is the most important respiratory signal used to detect SA because the impact of airway obstruction is reflected in this signal. The authors of this review found 14 articles that used only an AF signal to detect SA as indicated in tables 2 and 3. Most of those articles distinguished between apneic and normal events using binary classification methods. Nazeran et al (2001) reported poor performance of fuzzy inference systems (FIS) to detect apneic events using a respiratory AF signal from nine adults. Another approach to detect SA applied an adaptive fuzzy logic to the AF signal of ten subjects (Morsy and Al-Ashmouny 2005). Two classification engines (fuzzy logic-based and center of gravity) were used in series to distinguish normal and abnormal (apneic) events. This two-step, adaptive approach allowed high accuracy and permitted testing with a large clinical dataset.

A daytime AF recording of short duration (5 min), as opposed to occurrence of night-time apneic events, was proposed to separate obstructive and normal events (Salisbury and Sun 2007). A nonlinear and nonstationary signal analysis technique (Hilbert–Huang transformation) was applied to extract features from the AF signal and a threshold-based technique resulted in significant advantages over the previous methods. Han et al (2008) introduced a new algorithm to detect apneic events based on the mean magnitude of the second derivatives of the NAF signal. The suggested algorithm was found to be robust and useful due to a good overall agreement rate between the algorithm and manual scoring. An automated method was proposed by Otero et al (2008) to distinguish apnea from hypopnea events by applying fuzzy set theory. Further evaluation of this proposed approach was needed due to the use of a very small sample size (five subjects).

Caseiro et al (2010) screened and separated OSA patients from normal using 5 min oronasal airway pressure signal during waking. Though the approach did not require a whole night of recording, the performance of the separation technique was poor. Two studies published by Gutiérrez-Tobal et al (2012, 2013) used the same large dataset (148 patients) to distinguish between OSA positive and OSA negative. In the first study, an LRA model was used but performance improved using an MLP model in the second study. A real-time adaptive apnea event detection method was proposed by Koley and Dey (2013) using a two-stage classifier model. An SVM classifier was used to distinguish normal and abnormal (apneic) episodes that resulted in good accuracy. A per-second basis logical algorithm separated apneic and normal patients but further improvement in the classification approach was needed to get an acceptable performance (Selvaraj and Narasimhan 2013).

3.2.2. TE-based detection
SA detection using only a TE signal appears almost impossible and only one article was found that used a TE signal for the mentioned purpose (tables 2 and 3). Thommandram et al (2013) used chest movement waveform (also known as respiratory impedance signal) for SA detection. Four clinical features were extracted from each 1 min epoch of respiratory impedance waveform and a KNN classifier was used to separate apneic from normal epochs. The results of this study were promising but the size of the sample was small (only eight records) and demands further testing using a much larger dataset.

3.2.3. AE-based detection
Two studies reported the application of an AE signal to distinguish central, mixed, and obstructive SA events (tables 2 and 3). In one study, AE signals were separated into spectral components using multi-resolution wavelet transform (Tagluk et al 2010). The coefficients of discrete wavelet transform were fed to the input of the ANN to separate the apnea types. Tagluk and Sezgin (2010) used sub-band spectral energy instead of wavelet coefficients to minimize the size of the input vector and reported improved performance. It should be mentioned here that the above two studies were not solely dependent on the AE signal but also used the AF signal in the initial stage. AF signals were used to select the corresponding sections of the AE signals related to SA events, and then the selected AE sections were used to extract features as well as to detect apneic events. It is almost impossible to distinguish SA types using an AE signal only.

3.2.4. SaO2 signal-based detection

An alternative proposal that promoted not only a transmission of oximetry data but also a real-time analysis of those data locally with a mobile device was presented by Burgos et al (2010) and reported the best performance when using the Bagging classifier with an ADTree classifier. A novel multivariate system using PNN was proposed by Morillo and Gross (2013) that overperformed the existing univariate and multivariate approaches. Zhang et al (2013) presented a real-time auto-adjustable smart pillow system for apnea detection and treatment but the approach needs further improvement in event detection. Koley and Dey (2014) used an SVM classifier to detect apneic from normal events, whereas Sánchez-Morillo et al (2014) used a BHC classifier to detect four classes (mild, moderate, severe, and normal). A decision tree (C4.5) was applied to detect a large set of obstructive and normal population and reported good detection accuracy (Huang et al 2015). A recent study (Rolón et al 2017) reported a discriminative method to detect mild and moderate-to-severe patients using MLP NN. Another recent study by Morales et al (2017) performed better with KNN than LS-SVM to detect apnea and normal subjects.

3.3. SA detection based on multi-signals
A combination of different respiratory signals (AF, TE, and AE) and an oximetry or SaO2 signal has been also used to detect SA. A combined application of AF, TE, and AE signals was used to detect apnea, hypopnea, and normal events (Várady et al 2002), whereas only TE and AE signals were applied to distinguish central and obstructive events (Várady et al 2003). Fontenla-Romero et al (2005) detected different apnea types (central, mixed, and obstructive) using an ANN applying AF and TE signals, whereas Tian and Liu (2005) applied AF and SaO2 signals to detect apnea and hypopnea events using TDNN. Ng et al (2007) used a threshold-based technique to separate apnea and normal events using AE and TE signals. Álvarez-Estévez and Moret-Bonillo (2009) applied three respiratory signals (AF, TE, and AE), and an SaO2 signal to detect apnea and hypopnea events using fuzzy logic, whereas Kaimakamis et al (2009) applied two respiratory signals (AF and TE) and an SaO2 signal to separate obstructive and normal patients using a decision tree (C4.5) algorithm. Energy-based features were applied to an ANN to detect central, mixed, and obstructive events using TE and AE signals (Sezgin and Tagluk 2009). Álvarez et al (2010a) applied a nonparametric threshold-based method to distinguish OSA positive and OSA negative subjects using AF and SaO2 signals.
Guijarro-Berdiñas et al (2012) reported a good detection accuracy with an ANN. They used AF and TE signals with a small sample size to separate apnea types. Apnea and normal events were separated by a multivariable fuzzy temporal profile model (Otero et al 2012) and an ANN (Maali and Al-Jumaily 2013) using different combination of respiratory and oximetry signals. Combined signals were used to separate apnea and normal patients using event detection algorithm (Bianchi et al 2014) and elastic net classifier (Carmes et al 2014). Avci and Akbaş (2015) reported an outstanding detection accuracy where RFC was applied to detect apnea and normal events. Though the detection accuracy was very high, the sample size was small (eight subjects). Kagawa et al (2016) using Doppler radar proposed non-contact diagnosis of mild and moderate-to-severe OSA from TE and AE signals. LRM resulted in better than decision trees to detect obstructive and normal subjects (Kaimakamis et al 2016). A very recent study by Huang et al (2017) reported good precision in detecting apnea and hypopnea events using a respiratory events detection algorithm.

3.4. SA detection based on per epoch and per recording
Per-epoch- and per-recording-based detection is another distinguishing point of interest of this review. This review found 27 articles based on per-epoch detection (last four columns of table 2, numbers not in brackets), whereas the remaining 35 articles were based on per-recording detection (indicated with brackets, table 2). Per-recording detection was used to detect patients as OSA positive or OSA negative without addressing SA severity. On the other hand, per-epoch detection was used to detect each epoch as apnea or normal, and OSA positive or OSA negative. In the case of per-epoch detection, varying epoch length (5s, 15s, 30s, 1 min, and so on) was applied. The selection of epoch length is critical. However, it is challenging to determine an epoch length that is suitable for good reliability and accuracy since it also depends on classification methods. It is currently not possible to provide guidance on how best to select epoch length or the entire recording that is equally applicable to all situations.

4. Discussion
In this study, the authors presented an overview of the respiratory and oximetry signals used to detect SA. The authors also presented the metrics (accuracy, sensitivity, specificity, and other parameters) of corresponding classification methods by conducting a systematic review study on articles published from 2001–2017. From a high-level overview, the observed detection of SA fell into two categories, single-signal (only one respiratory or oximetry signal) based and multi-signal (i.e. a combination of more than one respiratory and oximetry signal) based. Detection based on a single-signal was further subdivided into the four signals of AF, TE, AE, and SaO2. This review reveals that respiratory and oximetry signals have increasingly been used in SA detection (figure 3). On the other hand, as presented in table 3, decision-making was based on both respiratory and oximetry signals as well as classification methods (ML, threshold-based, and other techniques) (table 4). The most common scenario observed was based on the use of SaO2 signal (43.55%) to detect SA and most of the decisions were made on binary classes.

Single respiratory- or oximetry signal-based SA detection was done more in the case of binary classes (Lee et al 2004, Marcos et al 2010, Gutiérrez-Tobal et al 2013, Cioczek et al 2015), whereas multi-signal application was done to detect multi-classes (Várady et al 2002, Sezgin and Tagluk 2009, Guijarro-Berdiñas et al 2012, Sánchez-Morillo et al 2014). AF and SaO2 signals used separately to detect SA were effective, whereas TE and AE signals alone were almost unable to detect SA. A combined application of TE and/or AE with AF was effective for multi-class detection (Várady et al 2002, Avci and Akbaş 2015), which was challenging when using a single respiratory or oximetry signal. A combination of the respiratory signals with an SaO2 signal resulted in good detection accuracy (Otero et al 2012, Huang et al 2017).

Evidently, signals from both single-channel and dual-channel devices are used for binary decision-making (e.g. decision-making between apnea and normal) but not for multi decision-making (e.g. decision-making between CSA, OSA, MSA, and normal). Signals from multi-channel devices would be more effective because multi decision-making is also applied automatically to binary decision-making. Thus, the application of multisignals from multi-parameter systems (e.g. acquisition of AF, TE, and AE with/without SaO2) would be systematically better than single channel (e.g. acquisition of AF or SaO2) or dual channel (e.g. acquisition of AF and SaO2) devices.

In addition to the use of respiratory and oximetry signals to detect SA, sample size and classification methods are the two major concerns for the evaluation of any SA detective systems. This review found intensive use of ML methods for SA detection using respiratory and oximetry signals. Some ML methods (SVM, RFC, AdaBoost, and KNN) were comparatively better than other methods (FIS, ANN, LRA, and LDA). The choice of appropriate ML methods is critical with the use of respiratory and oximetry signals to get high detection accuracy. Despite high detection accuracy with certain classifiers, the sample size used was often small (Guijarro-Berdiñas et al 2012, Avci and Akbaş 2015, Jin and Sánchez-Sinencio 2015). Studies that used highly specific datasets, which were con-
fined to small samples, suffered from limitations of generalizability of results and thus, further investigation is needed to validate the generality of classification models to detect SA on large datasets or within different populations. Apart from yielding high performance, an automated system with acceptable accuracy remains a major concern.

The main benefit of using respiratory and oximetry signals to detect SA is its ease of use. Acquisition of respiratory and oximetry signals is easy from human patients from overnight recordings. These signals without artefacts are reliable. The number of electrical cables required to acquire respiratory and oximetry signals is less than the other physiological signals such as EEG, ECG, and EMG. Moreover, the latter signals are easily affected by noise and complex processing is required to remove signal noise. On the other hand, the reliability of using respiratory and oximetry signals for SA detection is very high. Different devices have been developed which can record overnight respiratory and oximetry signals with good signal quality and less noise. Acquisition of good quality respiratory and oximetry signals would result in better detection of SA events and diagnosis.

PSG requires an exhaustive test in a hospital setting, skilled experts, high cost, and discomfort to the patient, so the implementation of a non-invasive, accurate, and home-based automated technique based on a simple set of respiratory and oximetry signals would be recommended. The main challenge of using respiratory and oximetry signals is to develop an accurate automated system to detect SA events. Acquisition of reliable, noise- and distortion-free respiratory and oximetry signals is paramount for an accurate SA detection system. Acquisition of NAF using a nasal pressure transducer is superior to using thermal-based oronasal AF sensors. Thermal sensors detect a change in the temperature of exhaled air. However, they may fail to detect minor, although significant, changes in AF. Thus, they may underestimate hypopneas (Norman et al. 1997). SaO2 acquisition usually uses finger-based sensors rather than a forehead reflectance oximeter. Overnight recording of SaO2 using forehead sensors is quite challenging and can cause patient discomfort and thus affect sleep quality. On the other hand, a finger-based oximeter is less obtrusive and easy to incorporate in a home-based SA detection system.

Despite a higher SA detection rate reported in several articles using different classification methods (Guijarro-Berdiñas et al. 2012, Avcı and Akbaş 2015, Jin and Sánchez-Sinencio 2015), some major challenges exist with ML techniques. Firstly, a high detection rate was reported mostly in articles that used a small sample size or a fixed number of records from a fixed database. The overall detection accuracy changes with the sample size and database used. For this reason, the accuracy of those detection methods may deviate when the same methods are applied to other datasets or a greater proportion of records. Secondly, each ML technique is linked to a basic step of features extraction. Features selection is an optional stage used when the number of features extracted is numerous and some of them are redundant. Extraction and selection of a large number of weakly relevant and redundant features are the key reasons for poor detection accuracy. On the other hand, efficient features extraction and sometimes robust features selection should result in good detection of SA events. It is difficult to extract relevant and distinguished features as well as to manage efficient and robust features from a wide range of features sets. Thirdly, selecting an appropriate classification method to provide reasonable, reliable, and consistent decisions is very critical. The selection of any classification method is need-based and depends on the nature of the extracted and selected features set. Finally, appropriate training of ML classifiers is a pre-requisite to getting better testing accuracy. Inappropriate training of ML classifiers results in poor performance. If the ML classifiers are poorly trained or overtrained, the testing accuracy will be affected. In addition, parameter selection and training time are two crucial points when training a classifier. Training time increases when new samples are added and thus affect the performance of the classifier. It is challenging to manage the above criteria that deal with ML classifiers to obtain acceptable detection accuracy.

Several automated methods of SA detection have already been developed (Marcos et al. 2010, Bianchi et al. 2014, Giolek et al. 2015) but the reliability reported is not quite high enough to implement in practical cases. In addition, many researchers have employed a single oximetry signal to detect SA (Álvarez et al. 2006a, 2010b, Marcos et al. 2007, Marcos et al. 2008a, Marcos et al. 2009a) and reported accuracy close to 90%. It is challenging to achieve a higher accuracy (e.g. over 90%). However, further improvement in accuracy would be possible, for example, by (1) employing multiple signals such as AF and SaO2. In addition, in the case of apneic type detection, inclusion of respiratory efforts (TE and/or AE) with an AF signal is mandatory in designing an automatic algorithm and (2) incorporating multiple logics into automatic algorithm. Multiple logics should be designed in such a way that they can accurately detect the apneic changes readily. Inclusion of accurate and multiple logics based on updated scoring rules of respiratory events (Berry et al. 2012) is mandatory to design more efficient automatic algorithms for SA detection, and (3) by using large clinical datasets for validation of automated system performance.

Some limitations of this review should be mentioned. Firstly, some studies included in this review did not report clearly on the performance metrics used for SA detection. Secondly, only English-language publications were included.
5. Conclusions and future directions

This systematic literature review has synthesized and summarized the existing classification methods based on respiratory and oximetry signals to detect SA. Sixty-two studies were examined and the main findings are summarized as follows.

A single respiratory signal, AF or SaO₂, provided good support for binary class decision-making, whereas multiple respiratory signals (AE, TE, and AE) combined with an SaO₂ signal resulted in better in multi-class decision-making in SA detection.

Several ML techniques, specifically the SVM and KNN, were by far more accurate than other methods and thus selection of appropriate ML approaches with appropriately selected respiratory and oximetry signals would be effective for SA detection.

Despite certain benefits associated with the use of respiratory and oximetry signals, major concerns remain: high accuracy is yet to be achieved with the automated detection technique. In addition, large and/or multiple samples of data should be included especially from a clinical perspective.

There remain some current gaps in the existing literature. A focus on SA detection using respiratory and oximetry signals would be the first priority. Acquisition of noise- and distortion-free signals is a prerequisite for accurate detection. Acquisition of NAF and SaO₂ signals using a pressure transducer and oximetry, respectively, would be mandatory. An automated detection approach is still required that will detect SA events with high accuracy.

ML methods are becoming more popular in view of big data analytics. These methods can be used to an advantage by applying relevant features extraction, robust features selection and good selection of classification methods, and optimal training of the classifiers. In addition, newly developed ML methods (deep learning) would be more appropriate than other ML approaches.

In parallel to finding new ML approaches, other detection approaches such as automated SA detection algorithms are promising. An automated and accurate SA detection approach would be needed, since epoch-based detection is challenging. Appropriate and multiple logics can be incorporated into an automated detection algorithm that will analyze each sample of the record one by one and detect apneic events accurately. Consideration of multiple signals (respiratory and oximetry signals) in designing a new automated SA detection algorithm would result in a higher detection rate. Finally, the designed automated SA detection algorithm should be validated or tested using large standardized datasets.

References

Alsubie H S and BaHammam A S 2017 Obstructive sleep apnea: children are not little adults Paediatr. Respir. Rev. 21 72–9
Álvarez D, Hornero R, Abásolo D, Del Campo F and Zamarrón C 2006a Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection Physiol. Meas. 27 399–412
Álvarez-Estévez D and Moret-Bonillo V 2009 Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome Expert Syst. Appl. 36 7778–85
Ben-Israel N, Tarasiski A and Zigal Y 2012 Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults Sleep 35 1299–305
Berger M, Oksenberg A, Silverberg S, Arons E, Radwan H and Iaina A 1997 Avoiding the supine position during sleep lowers 24 h blood pressure in obstructive sleep apnea (OSA) patients J. Hum. Hypertens. 11 657–64
/Bucklin C L, Das M and Luo S L 2010 An inexpensive accelerometer-based sleep-apnea screening technique

NAECON

Chang S J and Chae K Y 2010 Obstructive sleep apnea syndrome in children: epidemiology, pathophysiology, diagnosis and sequelae


Breiman L 1996 Bagging predictors Mach. Learn. 24 123–40


Cassaret D and Wang Y C 2005 Automatic target recognition using new support vector machine IJCNN ’05: Proc. of the IEEE Int. Joint Conf. on Neural Networks (Montreal, Canada) pp 84–9


Coenen F and Leng P 2007 The effect of threshold values on association rule based classification accuracy Data Knowl. Eng. 60 345–60

Cybenko G 1989 Approximation by superpositions of a sigmoidal function Math. Control Signals Syst. 2 303–14


Kaimakis E, Bratsas C, Sichlitis L, Karvounis C and Maglaveras N 2009 Screening of patients with obstructive sleep apnea syndrome using C4.5 algorithm based on non linear analysis of respiratory signals during sleep EMRS 09: 31st Annual Int. Conf. of IEEE Engineering in Medicine and Biology Society (Minneapolis) pp 3465–9


Köves P 1999 Obstructive Sleep Apnea Syndrome (Berlin: Springer).


Specht D F 1990 Probabilistic neural networks Neural Netw. 3 109–18
Standards of Practice Committee of the American Sleep Disorders Association 1997 Practice parameters for the indications for polysomnography and related procedures Sleep 20 406–22
Suykens J A and Vandewalle J 1999 Least squares support vector machine classifiers Neural Process. Lett. 9 293–300
Tagluk M E, Akin M and Sezgin N 2010 Classification of sleep apnea by using wavelet transform and artificial neural networks Expert Syst. Appl. 37 1600–7
Tagluk M E and Sezgin N 2010 Classification of sleep apnea through sub-band energy of abdominal effort signal using wavelets + neural networks J. Med. Syst. 34 1111–9
Zadeh L A 1965 Fuzzy sets Inf. Control 8 338–53