

Detection of Sleep Apnea using Machine Learning Algorithms based on ECG Signals: A comprehensive Systematic Review

1st: Nader Salari (PhD)

Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.
E-mail: n_s_514@yahoo.com

2nd: Amin Hosseinian-Far (PhD) *

(Corresponding author): Department of Business Systems & Operations, University of Northampton, Northampton, UK. E-mail: amin.hosseinian-far@northampton.ac.uk

3rd: Masoud Mohammadi (MSc)

Department of Nursing, School of Nursing and Midwifery, Kermanshah University of Medical Sciences, Kermanshah, Iran E-mail: Masoud.mohammadi1989@yahoo.com

4th: Hooman Ghasemi (BSc)*

(Corresponding author): Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran. E-mail: humanghasemi2010@gmail.com

5th: Habibolah Khazaie (MD)

Sleep Disorders Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran E-mail: Hakhzaei@gmail.com

6th: Alireza Daneshkhah (PhD)

School of Computing, Electronics and Maths, **Coventry** University, London, United Kingdom.
E-mail: ac5916@coventry.ac.uk

7th: Arash Ahmadi (PhD)

Department of Electrical and Computer Engineering, Faculty of Engineering, Windsor University, Windsor, Canada. E-mail: aahmadi@uwindsor.ca

Abstract. Sleep apnea (SA) is a common sleep disorder that is not easy to detect among several patients. Recent studies have highlighted ECG analysis as a method of diagnosing SA. Because the changes caused by SA on the ECG are so subtle, the need for new methods in diagnosing the disease is required more than ever. Machine Learning (ML) techniques are recognized as one of the most successful methods of computer aided diagnosis. ML uses new methods to diagnose diseases using past clinical results. The purpose of this study is to evaluate studies using ML algorithms and based on ECG characteristics to evaluate ~~people with SA~~ patients. In this study, ~~English~~ articles written in English and indexed in PubMed, Scopus, Web of Science, and IEEE databases, were searched with no lower time limit and until October 2020, and were systematically reviewed. Finally, 48 articles were approved for review ~~in this study~~. ~~In~~ Within the selected articles, different ML methods were ~~used~~ adopted for classification. All of these studies were binary, and SA was detected from the normal state based on a full ECG stripe (per record), or based on one-minute segments (per segment). Our ~~studies analyses showed~~ show that the most common features used in the studies were frequency, time series and statistical features. Support-Vector Machine (SVM) and deep learning-based neural network (i.e. CNN, DNN) performed best in full record data detection. The highest accuracy, sensitivity and specificity reported between the selected studies were 100%, which was obtained by an SVM. In another ~~easestudy~~, the classification was conducted based on ECG segments, and accordingly, the highest classification accuracy was observed in the residual neural network algorithm (RNN). The accuracy, sensitivity and specificity of this algorithm were reported to be 99%. In general, it can be stated that ML techniques based on ECG characteristics have a high capability in diagnosing SA. This can increase the diagnosis of patients with SA or the detection of SA episodes on ECG record, and can potentially prevent complications of the disease at later stages.

Keywords: Sleep Apnea, Machine Learning, Polysomnography, Electrocardiogram, Accuracy, Systematic Review

1- Introduction

1.1- Sleep apnea definition

Sleep Apnea (SA¹) refers to the periodic cessation or reduction of airflow during sleep (Peppard et al., 2013). SA occurs due to complete or partial obstruction of the upper airways (i.e. Obstructive Sleep Apnea (OSA²)), reduction or cessation of brainstem respiratory motor output (i.e. Central Apnea), or both (Dempsey, Veasey, Morgan, & O'Donnell, 2010). Complete cessation of respiration (i.e. apnea) or decreased airflow (i.e. hypopnea) are two respiratory events observed in SA. These events reduce oxygen levels, i.e. hypercapnia, and increase sympathetic nerve activity, and fluctuations in blood pressure, and heart rate. Such physiological changes also affect patients' sleep cycle. It causes brain arousal, disruption of various stages of sleep, and sleep fragmentation (Young, Skatrud, & Peppard, 2004; Parati, Lombardi, & Narkiewicz, 2007; Veasey & Rosen, 2019; Young, Skatrud, & Peppard, 2004). It is estimated that about 10% of middle-aged people are affected by SA (Peppard et al., 2013). Despite the high prevalence of this disorder, most patients are unaware of the effect of SA on their respiratory pattern. And because of this, many patients do not seek professional treatment (Veasey & Rosen, 2019). Many-Several studies have examined morbidity of SA. The results

¹ Sleep Apnea

² Obstructive Sleep Apnea

of these studies show that failure to diagnose and treat SA in a timely manner can cause daily drowsiness (Kainulainen et al., 2019), cognitive dysfunction (Ferini-Strambi et al., 2003), cardiovascular diseases such as hypertension (Torres, Sánchez-de-la-Torre, & Barbé, 2015), coronary artery disease (Torres-Alba et al., 2013), heart failure (H-Wang et al., 2007), stroke (Dyken & Im, 2009), and metabolic diseases such as diabetes (Kendzerska, Gershon, Hawker, Tomlinson, & Leung, 2014). Therefore, the early detection of SA is a crucial task, and can prevent subsequent ~~compleiations~~complications.

1.2- Sleep apnea detection

Polysomnography (PSG³) is known as the standard SA diagnostic test. Accordingly, PSG examines sleep and respiration parameters using electroencephalogram, electrocardiogram (ECG⁴), electrochogram, electromyogram, pulse oximetry, airflow measurement, and respiratory effort (Rundo & Downey, 2019; Gottlieb & Punjabi, 2020; ~~Rundo & Downey, 2019~~). PSG has a high diagnostic accuracy (Ali, Khalid, & Belhaouari, 2019), however, factors such as high cost, patient inconvenience, cumbersome data recording, and difficult interpretation of data are some of the disadvantages of this method. Moreover, the long waiting list for evaluating patients with a PSG device increases the possibility of not diagnosing and treating SA in time (Portier et al., 2000). Therefore, it is necessary to provide an alternative method to enhance patients' convenience, and reduce costs to diagnose SA at an early stage (Pombo, Silva, Pinho, & Garcia, 2020).

Different strategies have been proposed and adopted to diagnose SA without the use of PSG (Bozkurt, Bostanci, & Turhan, 2017; Wang, Lu, & Shen, 2019). Nonetheless, the use of ECG signals has received much attention (Wang, Lu, Shen, & Hong, 2019). ECG is not stressful for the patients compared to PSG, and the related ECG equipment are less technical. It has also been observed that the ECG with a signal strength of 1-2 mV has the best signal-to-noise ratio among all physiological signals (Kesper, Canisius, Penzel, Ploch, & Cassel, 2012). On the other hand, the parameters extracted from the ECG signal curve allow the extraction of the respiratory effort curve (ECG-induced respiration or EDR) (Janbakhshi & Shamsollahi, 2018).

SA also causes cyclical changes in heart rate (HR) (Hayano et al., 2011). In other words, apnea or hypopnea reduces the patient's oxygen level, and accordingly, less oxygen reaches the heart cells. This reduces the heart rate, and increases the distance between the R peaks. As a result of this event, the brain increases the respiration rate by using immediate pulses to the respiratory system, and subsequently, the heart rate will increase again. Thus, such respiratory arrests affect the ECG waves, and episodes of apnea are recorded during an ECG record (Ali & Hossen, 2020). HR changes are determined by the analysis of heart rate variability (HRV). Considering this, many studies distinguish apnea episodes from ECG record by extracting features such as LF (low frequency) and VF (high frequency) band power distributions from HRV (Babaeizadeh, White, Pittman, & Zhou, 2010; Kesper et al., 2012). With these interpretations, since the changes caused by SA in the ECG are very varied and subtle, the diagnosis of SA based on ECG data is a very complex task. One of the methods that has been

³ Polysomnography

⁴ electrocardiogram

considered to solve this problem is the use of computer algorithms (Faust, Acharya, NG, & Fijita, 2016).

1.3- Machine learning ~~to~~ for SA detection SA

Machine Learning (ML⁵) techniques have been considered as one of the successful methods of computer-aided diagnosis (M-Li & Zhou, 2007). ~~According to the pattern in~~ Considering **Figure 1**, ML is an evolving branch of the computational science designed to simulate human intelligence by learning from the environment (El Naqa & Murphy, 2015). ML is used when it is not possible to interpret a particular pattern or extract relevant information (A-Dey, 2016). When using ML algorithms, SA automatic detection is based on a large number of pre-detected samples. In other words, ML uses data from previous examinations in which the physician has diagnosed the presence or absence of thea disease (M-Li & Zhou, 2007).

Various ML techniques have been used in the diagnosis of SA. In a study by Bozkurt et al. (F-Bozkurt, Ucar, Bozkurt, & Bilgin, 2020), the authors used electrocardiography of 10 patients with OSA against 10 healthy controls. This study first extracted HRV from ECG, and then extracted the QRS component at different frequencies using a digital filter, and then selected the feature using Principal Component Analysis (PCA⁶). Classification was performed by k-Nearest Neighbours (k-NN⁷) algorithm. The results of this study showed that when using 3 features, the classification accuracy was 82.11%, and when using 13 features, this value was 85.12%. In another study (Erdenebayar, Kim, Park, Joo, & Lee, 2019), data collected from 86 patients were used, of which 69 were used in training and 17 in test. The Residual Neural Network (RNN⁸) algorithm was reported to offer the highest accuracy of 99%. Moreover, the study highlighted that deep learning techniques are very useful for automatic detection of SA. In another study (Nguyen, Wilkins, Cheng, & Benjamin, 2014), HRV data were used to automatically detect SA. Then a feature selection algorithm was used-adopted to select the best features. In this study, classification was performed using support vector machine (SVM⁹), artificial neural network (ANN¹⁰) and a combination of these two algorithms. The results of this study indicated that the proposed methods have a high capability in detecting SA from the healthy state. As mentioned, different ML algorithms have been used in detecting SA. Therefore, this systematic review was designed to evaluate the ability of ML algorithms to detect SA. Furthermore, as part of our work, the methods used within selected studies were compared with each other.

(Figure1 here)

2- Methodology

2.1- Search strategy and inclusion criteria

The protocol and reporting used in this systematic review were performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA¹¹)

⁵ Machine Learning

⁶ Principal Component Analysis

⁷ k-nearest neighbours

⁸ Residual Neural Network

⁹ Support Vector Machine

¹⁰ artificial neural network

¹¹ Preferred Reporting Items for Systematic Reviews and Meta-analysis

guidelines (Moher, Liberati, Tetzlaff, & Altman, & Group, 2009). The select relevant studies, the four databases of PubMed, Web of Science (WoS), Scopus and IEEE Xplore were searched. All searches were performed on 24th September 2020, and were subsequently updated on 24th October 2020. Searches were performed using keywords related to sleep apnea, machine learning and electrocardiogram, and in accordance with the search strategy for each of the databases; **Table 1** outlines the keywords and search strategies used for each of the databases (Table1). Articles selected for this systematic review entails studies that distinguish SA from healthy states, in a binary form. A number of diagnostic studies for apnea, hypopnea, and healthy states that followed a multiclass approach were excluded from this systematic review. Within all the included studies, various ML algorithms were used, and all classifications were based on electrocardiogram data. Exclusion criteria of this systematic review were: lack of access to the full text of articles, conference proceedings, and articles written in a language other than English. Furthermore, research works that used a method other than electrocardiogram were excluded from this work.

(Table 1 here)

2.2- Study selection and data extraction

After removing duplicate articles ~~using within~~ the EndNote ~~bibliography reference~~ management software, one of the authors (HGH) reviewed the title and abstract of ~~collected~~ ~~the remaining~~ articles. Accordingly, the studies that did not meet the inclusion criteria were omitted. Two authors (HGH, MM) then reviewed the full text of the remaining articles based on the inclusion and exclusion criteria. These assessment activities were performed independently, and using blind copies. After completing the reviews at this stage, both reviewers reviewed the comments made regarding the approval or rejection of the articles. In case of a disagreement, another author (NS), as a senior reviewer, made the inclusion or exclusion judgement. In cases where ~~the~~ full text of an article was not available, a request for the full text was sent to the corresponding author via email or ~~through~~ ResearchGate. Finally, if the ~~full~~ text was not available or could not be secured, the study had to be excluded. The following descriptive information were extracted from the remaining studies: a) year of study, b) country, c) dataset, d) pre-processing, e) feature extraction/selection, f) ML algorithm, and g) parameters reported in the study that represent the performance of the algorithms used. The collected studies and their data were reviewed systematically, yet conducting meta-analyses was not deemed appropriate in this work.

3- Results

3.1 Study selection

After searching the databases, a total of 391 studies were collected. Subsequently, duplicates were omitted and only a copy was retained. Then, the remaining 318 articles were evaluated based on the inclusion and exclusion criteria, leaving 68 studies. These 68 articles were reviewed for eligibility, and finally 48 articles were included in the systematic review (Figure 2). ~~Within the final studies,~~ ~~B~~by applying different methods, feature extraction was performed from ECG signals. These features were then used to construct training and test sets and to classify data. In these studies, SA was detected from healthy controls based on ‘per record’,

‘per segment’, or both. ~~The studies performed on the studies~~Our assessment and analysis ~~showshowed~~ that in 23 studies the diagnosis was made based on per record, and in 33 studies the diagnosis was made based on per segment. In the record mode, a complete ECG strip was analyzed to distinguish SA from healthy states. In the other case (per segment), the ECG strip was first divided into smaller pieces and named by experts as apnea ~~and-or~~ healthy ~~parts~~. Classification was then performed based on the characteristics extracted from these components.

(Figure 2 here)

All included studies used single-lead ECG records for classification. Due to changes in the autonomic nerve activity, sleep apnea pathology is associated with rapid bradycardia and tachycardia on the ECG signal (Almazaydeh, Elleithy, & Faezipour, 2012). Therefore, these signals contain useful information about the cardiovascular activity, and other systems within the body. It is important to note that even if SA does not cause cardiovascular complications, its effects and episodes of apnea can be seen on the ECG record (Atri & Mohebbi, 2015). Most of the present studies used the ECG record obtained from polysomnography of patients referred to sleep clinics. A small number of studies were from ECG records obtained from portable devices (Baek, Kim, Kim, & Lee, 2014).

3.2- Dataset

The datasets used in most studies overlapped. These studies used samples from the Physionet Apnea-ECG Database (Penzel, Moody, Mark, Goldberger, & Peter, 2000). Nonetheless, among the remaining studies, one study used data from Yildirim Beyazit (Akşahin, Erdamar, Firat, Ardiç, & Eroñul, 2015) and another study used samples from Sleep laboratories at Chest Diseases Clinics in Sakarya, Turkey (Bozkurt et al., 2020). Referrals to Sultan Qaboos University Hospital in Oman, Samsung Medical Center in South Korea, and University of Heidelberg Hospital in Jordan were also examined in other studies. Two other research works used data from the Sleep Heart Health Study cohort (Al-Angari & Sahakian, 2012; Eiseman, Westover, Mietus, Thomas, & Bianchi, 2012). In 6 studies, data from more than one dataset were examined. In 4 studies, data from both Physionet Apnea-ECG Database, and St. Vincent’s University Hospital/University College Dublin were used ([Travieso, Alonso, del Pozo, Ticay, & Castellanos-Dominguez, 2014](#); Hassan & Haque, 2017; Rekha, Kandaswamy, & Ramanathan, 2018; ~~Travieso, Alonso, del Pozo, Ticay, & Castellanos-Dominguez, 2014~~; ~~T. Wang, C.-H. Lu, & G.-H. Shen, 2019~~). One other research work examined data from Physionet and University Hospital Leuven (Varon, Caicedo, Testelmans, Buyse, & Van Huffel, 2015), and one study examined 3 different datasets (Khandoker, Palaniswami, & Karmakar, 2009). The remaining two studies (Baek, Kim, Kim, & Lee, 2014; Smruthy & Suchetha, 2017) had not provided clear information in relation to the use of specific databases (Supplementary appendix A).

3.3- Pre-processing

In general, the feature extraction process was very different among the studies. In almost all studies, pre-processing was performed with the aim of breaking down the ECG waves into

smaller chunks, and clearing them of junk data. R peaks are known as one of the common features in the diagnosis of apnea, which was detected by different algorithms. In two studies, the diagnosis was made based on ~~expertsexperts'~~ opinions. One of the most common methods in studies for ECG signal analysis was the Pan Tompkins algorithm (PTA¹²) (Al-Angari & Sahakian, 2012; ~~Ali & Hossen, 2020~~; Baek et al., 2014; ~~Varon et al., 2015~~; ~~Tripathy, 2018~~; Bali, Nandi, Hiremath, & Patil, 2018; ~~K-Li, Pan, Li, Jiang, & Liu, 2018~~; ~~H-Sharma & Sharma, 2016, 2020~~; ~~Ali & Hossen, 2020~~~~Tripathy, 2018~~; ~~Varon et al., 2015~~). This algorithm uses the amplitude, slope, and the width of an integrated window to distinguish peaks of R from the QRS complex. Pan Tompkins is known as the QRS detection algorithm in real-time approaches (Fariha, Ikeura, Hayakawa, & Tsutsumi, 2020). Other studies (~~Mendez et al., 2007~~; Khandoker, Karmakar, & Palaniswami, 2009; Khandoker, Palaniswami, et al., 2009; Mendez, Bianchi, Matteucci, Cerutti, & Penzel, 2009; ~~Mendez et al., 2007~~) have used real-time algorithms. ~~y~~ Yet the specific types of the algorithms used ~~were~~ were not reported.

Due to the oscillating nature of the ECG signal, wavelet-based algorithms and analyzes were used in several works. Tunable-Q factor wavelet transform (TQWT¹³) is known as one of the wavelet analyzes that was adopted in four research works (Hassan, 2016; Hassan & Haque, 2016b, 2017; Nishad, Pachori, & Acharya, 2018). This method breaks down the ECG signal into a number of sub-bands signals to extract features. Daubechies (Db) wavelet is another method that was adopted in some works. In this method, the ECG signal is decomposed into several segments. Moreover, some studies have used 4 Db (Bsoul, Minn, & Tamil, 2011; Rachim, Li, & Chung, 2014), 6 Db (~~Yildiz, Akin, & Poyraz, 2011~~; Atri & Mohebbi, 2015; ~~Yildiz, Akin, & Poyraz, 2011~~) and 14 Db (Khandoker, Karmakar, et al., 2009; Khandoker, Palaniswami, et al., 2009) to analyze or extract features. In another research work (Travieso et al., 2014), the continuous wavelet transform method was applied to ECG signals to detect R Waves and peaks. In two other studies (~~M-Sharma, Agarwal, & Acharya, 2018~~; ~~M-Sharma, Raval, & Acharya, 2019~~), the optimal biorthogonal antisymmetric wavelet filter bank was adopted to differentiate SA from healthy. These studies first decomposed the ECG signal into 5 levels by conducting the wavelet decomposition to extract the features.

Studies have used other algorithms to segment ECG signals. In two works (Akşahin et al., 2015; Pinho, Pombo, Silva, Bousson, & Garcia, 2019), due to the non-linear nature and energy of ECG waves, the Teager Energy Operator (TEO¹⁴) method was adopted. Fourier decomposition was another method used in ECG analysis (Fatimah, Singh, Singhal, & Pachori, 2020). A study (Hassan & Haque, 2016a) adopted Empirical Mode Decomposition (EMD¹⁵) to analyze signals. EMD is known as a data-adaptive signal processing method that performs highly localized time-frequency estimations. Another study (Rekha et al., 2018) identified the QRS complex using the Hilbert algorithm. Variational Mode Decomposition (Smruthy & Suchetha, 2017), dynamic autoregressive (AR) representation model (~~L-Wang, Lin, & Wang, 2019~~) and Iterated Cumulative Sums of Squares (ICSS¹⁶) (Chen, Zhang, & Song, 2015) were other methods to detect the RR interval. In another study, the BIOSIG-toolbox was used to detect RRI and EDR (Song, Liu, Zhang, Chen, & Xian, 2016). In two studies, the diagnosis of

¹² Pan Tompkins algorithm

¹³ Tunable-Q factor Wavelet Transform

¹⁴ Teager Energy Operator

¹⁵ Empirical Mode Decomposition

¹⁶ Iterated Cumulative Sums of Squares

QRS complex ([H. Sharma & Sharma, 2016](#)) and EDR ([H. Sharma & Sharma, 2020](#)) were performed based on Hermite algorithm.

The filters have been applied to ECG waves with the aim of clearing ECG from noise, ground drift and baseline drift, and detecting HRV and EDR. Moving average filter was applied with the aim of limiting and adjusting the lower and upper limits of waves and eliminating false and unexplained points ([Ali & Hossen, 2020](#); [Khandoker, Karmakar, et al., 2009](#); [Khandoker, Palaniswami, et al., 2009](#); [T. Wang, C. H. Lu, G. H. Shen, et al., 2019](#); [Ali & Hossen, 2020](#)). This filter was also used to detect EDR ([Song et al., 2016](#)), RR distance ([Chen et al., 2015](#)) and mean RR value ([Mendez et al., 2007](#); [Mendez et al., 2009](#); [Mendez et al., 2007](#)). One work ([Tripathy, 2018](#)) adopted a bandwidth filter to detect EDR. Low pass filter band was adopted in a number of other studies. This particular filter was used to remove noise from the ECG strip in [another-another study article](#) ([Atri & Mohebbi, 2015](#)). In a research work ([Baek et al., 2014](#)) the low pass and high pass filters, and in another work ([Bali et al., 2018](#)) low pass, high pass and band pass filters were used to eliminate noise. Chebyshev bandpass filter types I and II ([F. Bozkurt et al., 2020](#); [Lweesy, Fraiwan, Khasawneh, & Dickhaus, 2011](#); [Rekha et al., 2018](#); [Bozkurt et al., 2020](#)), Butterworth pass filter ([Chang, Yeh, Lee, & Lin, 2020](#)), FIR band pass ([Erdenebayar et al., 2019](#); [T. Wang, C. H. Lu, & G. H. Shen, 2019](#)), powerline filter ([Varon et al., 2015](#)) and Savitzky–Golay filter ([Pinho et al., 2019](#)) have been adopted to clean signals from junk data. In one article, the type of filter used to delete junk data was not reported ([Eiseman et al., 2012](#)) (Supplementary appendix A).

3.4- Feature Extraction/Selection

In the selected studies, after ECG analysis, features were extracted from different sections such as RR Interval (RRI¹⁷), HRV, EDR, R wave, and P wave. Fourier transform has been one of the most common methods for feature extraction. This method was used in most studies to extract Power Spectral Density (PSD¹⁸) of signals, and the frequency feature. PSD shows energy changes as a function of frequency. In a number of works ([Mendez et al., 2007](#); [Bali et al., 2018](#); [Bsoul et al., 2011](#); [Song et al., 2016](#); [Mendez et al., 2007](#); [Bali et al., 2018](#); [Rekha et al., 2018](#); [Nakayama et al., 2019](#); [Pinho et al., 2019](#); [Rekha et al., 2018](#); [Song et al., 2016](#); [T. Wang, C. H. Lu, & G. H. Shen, 2019](#)), in addition to the frequency feature, Fourier transform was adopted to extract the time domain feature. Another study (Ali & Hossen, 2020) extracted the frequency feature using wavelet packet decomposition. The Lomb algorithm was also used to obtain signals PSD (Babaeizadeh et al., 2010). In one article, the Hilbert algorithm was applied to extract the time and frequency features (Rekha et al., 2018).

In 6 articles ([Mendez et al., 2009](#); [Bali et al., 2018](#); [Bsoul et al., 2011](#); [Song et al., 2016](#); [Bali et al., 2018](#); [Mendez et al., 2009](#); [Nakayama et al., 2019](#); [Song et al., 2016](#); [T. Wang, C. H. Lu, & G. H. Shen, 2019](#)), frequency and time domain features were extracted from ECG signals simultaneously. Eight other studies ([Mendez et al., 2007](#); [Babaeizadeh et al., 2010](#); [Yildiz et al., 2011](#); [Aksahin et al., 2015](#); [Al-Angari & Sahakian, 2012](#); [Babaeizadeh et al., 2010](#); [Chen et al., 2015](#); [Eiseman et al., 2012](#); [Aksahin et al., 2015](#); [Mendez et al., 2007](#); [Varon et al., 2015](#); [Yildiz et al., 2011](#)) adopted the frequency feature only, for the classification of SA and healthy individuals states. One research work also used only the time domain feature for classification. In another work (Atri & Mohebbi, 2015), PSD was extracted along with bispectral feature. Another article (Baek et al., 2014) that applied a portable accelerometer with 3 electrodes used the frequency along with the data accelerometer as a feature. Another work (Jafari, 2013) adopted the frequency feature along with the RPS based feature for classification. One piece of research ([T. Wang, C. H. Lu, & G. H. Shen, 2019](#)) also extracted frequency along with the amplitude feature. Frequency and statistical features were also extracted in one article (Hassan & Haque, 2016b). Moreover, 7 studies ([Khandoker, Karmakar, et al., 2009](#); [Khandoker, Palaniswami, et al., 2009](#); [Nguyen et al., 2014](#); [F. Bozkurt et al., 2020](#); [Hassan & Haque, 2016a, 2017](#); [Khandoker, Karmakar, et al., 2009](#); [Khandoker, Palaniswami, et al., 2009](#); [Nguyen et al., 2014](#); [Rekha et al., 2018](#); [Bozkurt et al., 2020](#)) considered statistical features, 9 studies ([Fatimah et al., 2020](#); [Lweesy et al., 2011](#); [Nishad et al., 2018](#); [Rachim et al., 2014](#); [Smruthy & Suchetha, 2017](#); [H. Sharma & Sharma, 2020](#); [Nishad et al., 2018](#); [M. Sharma et al., 2018](#); [M. Sharma et al., 2019](#); [Smruthy & Suchetha, 2017](#); [Tripathy, 2018](#); [Sharma & Sharma, 2020](#); [Fatimah et al., 2020](#)) considered the characteristics related to wave energy and entropy, and a study (Lweesy et al., 2011) used the properties of P and T wave analyses. Four articles (Nguyen et al., 2014; Travieso et al., 2014; [Varon et al., 2015](#); [Tripathy, 2018](#); [Varon et al., 2015](#)) adopted various features for classification, yet the specific features used in these studies were not reported. Also, a number of other studies ([Chang et al., 2020](#); [D. Dey, Chaudhuri, & Munshi, 2018](#); [Li et al., 2018](#); [Erdenebayar et al., 2019](#); [Farouk, Anwar, & Zakaria, 2019](#); [K. Li et al., 2018](#); [L. Wang et al., 2019](#); [T. Wang, C. H. Lu, G. H. Shen, et al., 2019](#); [Chang et al., 2020](#); [X. W. Wang](#)

¹⁷ RR Interval

¹⁸ Power Spectral Density

et al., 2020) prepared data for classification by applying different layers within the Neural Network (NN¹⁹) (Supplementary appendix A).

3.5- Per record classification

In 23 studies, data were classified as per record. The data used in the studies were between 20 and 4647 ECG records. Classification was performed based on 15 different ML algorithms. The most commonly used algorithm was SVM, which was used in Restricted Boltzmann Machine (RBF²⁰), and linear and polynomial forms; the accuracy of this algorithm in detecting SA from healthy individuals was reported to be between 65.4% and 100%. The sensitivity and specificity of this algorithm were also reported (43.4%-100%) and (36.4%-100%) respectively. Moreover, k-nearest neighbors (~~kNN~~k-NN²¹) algorithm was applied in 7 studies; the lowest accuracy of this algorithm was reported 77.3%, and the highest accuracy was 97.14%. In all of these studies, the classification sensitivity was equal to or greater than 80%. The specificity of the ~~kNN~~k-NN algorithm also showed that, with the exception of one work (~~T. Wang, C. H. Lu, G. H. Shen, et al., 2019~~), in other studies, at least half of the healthy individuals were correctly classified, thus, the specificity was 50%. Logistic ~~regression~~ Regression (LR²²) was another common algorithm for classifying data per record. The accuracy of this method in diagnosing SA was reported to be in the range of 74.3% to 97.14%. The classification sensitivity of this algorithm was reported to be 100% in all studies. However, the specificity of the LR algorithm was very scattered, and was reported in the range of 18.2% to 90.91%.

Other methods used in the [reviewed](#) studies were neural network algorithms. These algorithms include Feed ~~forward~~ Forward ~~neural~~ Neural ~~network~~ Network (FFNN²³) in two studies (Akşahin et al., 2015; Ali & Hossen, 2020), PNN in three studies ([Khandoker, Karmakar, et al., 2009](#); [Sharma & Sharma, 2016](#); Ali & Hossen, 2020; [Khandoker, Karmakar, et al., 2009](#); [H. Sharma & Sharma, 2016](#)), ~~Ce~~onvolutional ~~N~~neural ~~N~~network (CNN²⁴) in 2 studies (Chang et al., 2020; ~~T. Wang, C. H. Lu, G. H. Shen, et al., 2019~~), Deep ~~N~~neural ~~N~~network (DNN²⁵) (Kaguara, Nam, & Reddy, 2014), and Artificial ~~N~~neural ~~N~~network (ANN²⁶) (Bali et al., 2018) each in ~~one a single~~ study. The results showed that neural network algorithms perform classification with a very high accuracy. Classification accuracy was also reported in these studies (80%-99%). The sensitivity and specificity of these algorithms were investigated in 7 research works. Sensitivity greater than 85% and specificity greater than or equal to 80% were reported in all, but one study (Khandoker, Karmakar, et al., 2009). Among the other techniques used, Naive Bayes, ~~L~~east-~~S~~square ~~S~~upport ~~V~~ector ~~M~~achine (LS-SVM²⁷), ~~L~~inear Discriminant Analysis (LDA²⁸), Bagged Decision Tree (BDT²⁹), ~~Q~~quadratic ~~C~~lassifier (QC³⁰) could be mentioned. The minimum classification accuracy was obtained in the study of

¹⁹ Neural network

²⁰ Restricted Boltzmann Machine

²¹ k-Nearest Neighbors

²² Logistic regression

²³ Feed Forward Neural Network

²⁴ Convolutional Neural Network

²⁵ Deep neural network

²⁶ Artificial neural network

²⁷ Least-Square Support Vector Machine

²⁸ linear Discriminant Analysis

²⁹ Bagged Decision Tree

³⁰ Quadratic Classifier

Eiseman et al. (Eiseman et al., 2012) using Naive Bayes for which the value was reported to be 63.02%. Study of Song et al. (Song et al., 2016) also reported SA detection accuracy based on ANN algorithm as 68.6%. The results of other articles reported a high accuracy of 75% in the classification based on per record data (Table 3).

3.6- Per segment classification

Classification was performed per segment in 33 articles. These studies used segments ranging from 1500 to 43522 for their classification. In 9 studies ([Nguyen et al., 2014](#); [Rachim et al., 2014](#); [Hassan, 2016](#); Hassan & Haque, 2016a, 2016b, 2017; [Hassan, 2016](#); [Nguyen et al., 2014](#); Nishad et al., 2018; [Rachim et al., 2014](#); Rekha et al., 2018; [T. Wang, C. H. Lu, G. H. Shen, et al., 2019](#)), the number of examined segments was not reported. However, in all of these articles, it was stated that one-minute ECG components were labeled by experts, indicating a classification based on one-minute components in these studies. Furthermore, SVM was used as the most common algorithm. In total, among the selected studies, 31 different ML algorithms were adopted to classify SA and healthy [individuals](#)state.

SVM algorithm has been used in linear, polynomial, RBF, and nonlinear forms. The accuracy of this algorithm was reported in the range of 59.22% to 93.91%. The sensitivity of the algorithm was also reported in the range of 32.7% to 95.2%. Specificity [studies-analyses](#) of SVM algorithm showed that the minimum specificity was 47.32%, with the maximum being 95.42%. In 3 articles, Hidden Markov Model (HMM³¹) was applied to SVM algorithm, leading to the accuracy of the algorithm to be between 80% and 100%. Another LS-SVM algorithm was used in a number of research works. The lowest [reported](#) accuracy [reported](#) was in the study of Hassan et al. (Hassan & Haque, 2017), which was 31.88%. Other studies reported an accuracy of more than 70% for the LS-SVM algorithm. The highest accuracy was reported in the study of Atri et al. (Atri & Mohebbi, 2015) with 95.57%. The sensitivity and specificity of this algorithm in this [study-work](#) were [reported](#) 98.64% and 92.51%, respectively.

NN algorithms have also been used in several related research works. CNN, DNN, RNN, ANN and PNN were the specific algorithms used. Eight studies ([Mendez et al., 2009](#); [Lweesy et al., 2011](#); [Nguyen et al., 2014](#); Atri & Mohebbi, 2015; Hassan, 2016; Hassan & Haque, 2016a; K. Li et al., 2018; [Lweesy et al., 2011](#); [Mendez et al., 2009](#); [Nguyen et al., 2014](#); Pinho et al., 2019; [H. Sharma & Sharma, 2020](#)) adopted the ANN algorithm. The highest accuracy was reported at 92.3%, and the lowest at 68.52%. CNN was another algorithm that was examined in 6 articles ([Chang et al., 2020](#); [D. Dey et al., 2018](#); Erdenebayar et al., 2019; Farouk et al., 2019; [L. Wang et al., 2019](#); [T. Wang, C. H. Lu, G. H. Shen, et al., 2019](#); [Chang et al., 2020](#)). The accuracy reported in these studies was 78.2% to 98.91%. Sensitivity and specificity were also calculated to be more than 80% in most research works. In only one article ([T. Wang, C. H. Lu, G. H. Shen, et al., 2019](#)), the reported classification sensitivity was 26.6%. This study had the lowest accuracy among all articles in terms of CNN algorithms. Specificity reported as 86.9%. DNN and RNN algorithms were adopted in the study of Erdenebayar et al. (Erdenebayar et al., 2019). The results of this [study-research](#) showed that the accuracy of the DNN algorithm in detecting and classifying SA's and healthies was 93.1%. Moreover, in this research work, the accuracy of the RNN algorithm was 99%. The sensitivity and specificity of DNN and RNN algorithms were above 90%. PNN was also investigated in a study with an

³¹ Hidden Markov Model

accuracy of 60.95%. This work (Hassan & Haque, 2017) did not report sensitivity and specificity for the PNN algorithm.

Random forest (RF³²), LR, LDA and ~~kNNk-NN~~ were 4 other algorithms used to classify data. ~~kNNk-NN~~ was applied to selected datasets in 12 studies (~~Mendez et al., 2009; Hassan, 2016; Song et al., 2016; Hassan & Haque, 2016a, 2017; Sharma & Sharma, 2016, 2020; M. Sharma et al., 2018; M. Sharma et al., 2019; Wang, Lu, Shen, et al., 2019; F. Bozkurt et al., 2020; Fatimah et al., 2020; Hassan, 2016; Hassan & Haque, 2016a, 2017; Mendez et al., 2009; H. Sharma & Sharma, 2016, 2020; M. Sharma et al., 2018; M. Sharma et al., 2019; Song et al., 2016; T. Wang, C. H. Lu, G. H. Shen, et al., 2019~~). Examination of the results in these research works showed that the accuracy of the ~~kNNk-NN~~ algorithm is in the range of 66.1% to 90.57%. In these works, with increasing accuracy, sensitivity and specificity also increased. The accuracy of RF, LR and LD algorithms were reported to be in the ranges of 79.26%-92.78%, 66%-85.6% and 62.93%-83.72% respectively, which shows the better performance of RF algorithm in data classification. Information on other algorithms used are provided in Table 3.

As highlighted earlier, the selected research works have used different methods to classify SA's from healthy ~~individualsstates~~. Therefore, despite the use of the same datasets in most studies, the main difference was observed between the performance of the selected algorithms. Applying the preprocessing step to data seems to increase the accuracy of the classification. In a number of research works, it has been observed that applying different filters eliminates noise, and improves the system's performance (~~Ali & Hossen, 2020; Atri & Mohebbi, 2015; Bali et al., 2018; Chen et al., 2015; Erdenebayar et al., 2019; Khandoker, Karmakar, et al., 2009; Khandoker, Palaniswami, et al., 2009; Lweesy et al., 2011; Song et al., 2016; Li et al., 2018; Bali et al., 2018; Lweesy et al., 2011; Rekha et al., 2018; Sharma & Sharma, 2020; Song et al., 2016; Wang et al., 2019; Wang, Lu, & Shen, 2019; Sharma & Sharma, 2020; Ali & Hossen, 2020~~). Accordingly, the study of Bali et al. argues that the use of filters and wave purification improves the performance of the classification system by about 12% (Bali et al., 2018).

A different factor that improves systems' performance is related to how data would be selected for training and testing. It has been observed that most existing systems use equal sized set of data (i.e. 50% training and 50% test). It has also been observed that with increasing severity and frequency of apnea, diagnosis, and classification were performed with much higher accuracy (~~Akşahin et al., 2015; Ali & Hossen, 2020; Khandoker, Karmakar, et al., 2009; Akşahin et al., 2015; Nakayama et al., 2019; Ali & Hossen, 2020~~). In the study of Ali et al. (2020), it was outlined that the accuracy of diagnosing severe apnea from healthy states is 95%. While the accuracy of patient and healthy classification in this study was 87.5%. Moreover, in a different work conducted by Khandoker, et al. (2009), the classification was based on healthy individuals and those who had experienced apnea for more than 100 minutes. In the study, people who experienced apnea between 5-100 minutes were known as borderline, and were excluded from the research. Therefore, it can be argued that in the ~~researchis study~~, patients with moderate to severe SA only were examined. Classification in the ~~sameis~~ work was reported as 100%, using the SVM algorithm. Further, Eiseman et al. (2012) examined 4647 participants in the Sleep Heart Health Study cohort; similar to many other studies, this work also used filters to remove noise. PSD features were also adopted in the work. However, the

³² Random forest

accuracy of the system in this study was lower than other research works. This may be due to the fact that many of the patients in the research had mild apnea.

According to the reported results of the selected studies, the use of wavelet-based methods is one of the successful methods for feature extraction ([Khandoker, Palaniswami, et al., 2009](#); [Yildiz et al., 2011](#); [Bsoul et al., 2011](#); [Rachim et al., 2014](#); [Khandoker, Palaniswami, et al., 2009](#); [Nishad et al., 2018](#); [Rachim et al., 2014](#); [Sharma et al., 2018](#); [Sharma et al., 2019](#); [Yildiz et al., 2011](#)). These methods are very useful in oscillating waves. They break down waves into smaller parts at specific frequencies (Rachim et al., 2014). Another method that had a significant impact on feature selection was PCA, in a way that all studies using PCA for feature reduction and effective feature selection performed very well ([Rachim et al., 2014](#); [Varon et al., 2015](#); [Bali et al., 2018](#); [Rekha et al., 2018](#); [Tripathy, 2018](#); [Bozkurt et al., 2020](#); [Rachim et al., 2014](#); [Rekha et al., 2018](#); [Tripathy, 2018](#); [Varon et al., 2015](#)). Another successful method that has performed very well in studies is the deep learning method; all [research works the studies](#) that used this method to train the system and select the feature, reported a high accuracy, sensitivity and specificity (in the range of 90-100%) in both [parts](#)-per record and per segment. Among these, deep learning based neural networks such as DNN and CNN were the most adopted ([Kaguara et al., 2014](#); [Chang et al., 2020](#); [Dey et al., 2018](#); [Erdenebayar et al., 2019](#); [Farouk et al., 2019](#); [Kaguara et al., 2014](#); [K. Li et al., 2018](#); [Erdenebayar et al., 2019](#); [Farouk et al., 2019](#); [Wang et al., 2019](#); [Wang, Lu, & Shen, 2019](#); [Wang, Lu, Shen, et al., 2019](#); [Chang et al., 2020](#)). Older NN methods such as FFNN, PNN and ANN also had acceptable performances ([Mendez et al., 2009](#); [Lweesy et al., 2011](#); [Akşahin et al., 2015](#); [Ali & Hossen, 2020](#); [Bali et al., 2018](#); [Ali & Hossen, 2020](#); [Lweesy et al., 2011](#); [Mendez et al., 2009](#)). System performance seems to improve with increasing neural layers in these works (Mendez et al., 2009). Another common algorithm that performed very well in most studies was the SVM algorithm, and its derivative model (LS-SVM). It was also observed that all studies that apply HMM to their models perform classification with excellent accuracy and performance ([Travieso et al., 2014](#); [Song et al., 2016](#); [Li et al., 2018](#); [Song et al., 2016](#); [Travieso et al., 2014](#)). In the work of Song et al. (Song et al., 2016), it was observed that the application of HMM on various algorithms such as SVM, LR, and LDA greatly increases the accuracy of classification. This result was also observed in the studies of Travieso et al. (2014) and Li et al. (2018).

(Table 2 here)

4- Discussion

.41- Comparison of studies

In this study, different ML algorithms used in the diagnosis of SA were examined. Our analyses show that SA detection using both per record and per segment methods has high accuracy, sensitivity and specificity. Various features of the ECG were also extracted and used in the classification. The occurrence of autonomic nervous system dysregulation is common in patients with SA and is known as one of the most important mechanisms affecting the cardiovascular outcomes caused by SA. There are several methods for assessing these disorders, one of which is HRV analysis (Lombardi, Pengo, & Parati, 2019). In other words, HRV is a reflection of the state of autonomic nervous system activity, and is a measure of overall heart health (Ali & Hossen, 2020). Thus, the widespread use of HRV fluctuations in apnea episodes in an ECG record can be observed in various studies (Bušek, Vaňková, Opavský, Salinger, & Nevšimalová, 2005).

EDR is another feature that was commonly monitored. This feature is important since the electrodes on the surface of the body move due to the sequential filling and emptying of lungs. This movement shifts the axis relative to the heart, and a continuous ECG-derived respiration is extracted for each normal QRS complex (Moody et al., 1986). As a result, many classification studies were performed based on EDR and HRV characteristics.

The number and selection of appropriate attributes are crucial for the success of classification (Foster, Koprowski, & Skufca, 2014). PSD signal alone (or in combination with other features) was one of the most common features used in most research works. Time domain and statistical features were other features that were used for classification. In general, our review ~~showed~~ shows that all the features used in the selected studies were applicable to the diagnosis of SA. However, it was not possible to determine the best feature to diagnose SA. This is due to the high accuracy of classification in various studies. Moreover, the combined use of the features in some research works prevents a tailored and detailed examination.

The highest classification accuracy (100%) when performing the analysis based on a complete ECG strip was reported in the study of Babaeizadeh et al. (2010). ~~In, in~~ which the authors used the QC algorithm to classify data. In their work, sensitivity and specificity were also reported to be 100%. Moreover, the study used a hybrid classification system. Additionally, PSD features extracted from HRV were used for classification. Other studies such as (Khandoker, et al., 2009a; Khandoker, et al., 2009b; Sharma & Sharma, 2020) also ~~provide-reported~~ accuracy, sensitivity, and specificity and reported as 100% ~~specificity~~. In the first two studies, HRV and EDR were broken down into smaller sections using a wave decomposition method. Then, a number of features were selected and classified. In both studies, the SVM algorithm performed a more accurate classification than the other algorithms. The better performance of SVM than other algorithms has also been reported in some other studies (Sharma & Sharma, 2016; Song et al., 2016). Although in the study of Sharma et al. (2020), the SVM algorithm was highly accurate, the highest classification accuracy was observed using the LDA algorithm. In this study, classification was performed based on energy, entropy and standard deviation characteristics in HRV and EDR. Overall, existing research works have shown that classifying data per record is a very suitable approach to diagnose and classify SA patients ~~from patients from~~ healthy individuals. Classification using full ECG tape seems to provide

more accurate diagnoses of SA than ECG components, which was confirmed in all studies ~~that had comparing compared~~ the two methods (Babaeizadeh et al., 2010; Al-Angari & Sahakian, 2012; Rachim et al., 2014; Sharma & Sharma, 2016, 2020; Song et al., 2016; Babaeizadeh et al., 2010; Chang et al., 2020; Li et al., 2018; ~~Rachim et al., 2014; Sharma & Sharma, 2016, 2020; Song et al., 2016;~~ Wang, Lu, & Shen, 2019; Wang, Lu, Shen, et al., 2019; Chang et al., 2020). However, this does not suggest that segmentation is not performing well.

Unlike per record data classification, none of the classifications based on ECG components had 100% accuracy, sensitivity and specificity. However, a review of all the presented results shows that more than 80% of the algorithms applied to the ECG components have a high accuracy that is in the range of 70% to 99%. This indicates that the use of small ECG components also performs very well in the diagnosis and classification of SA. Our review also highlighted that the best algorithms in segment classification were Neural Network, RF and SVM. Erdenebayar et al. (Erdenebayar et al., 2019) used neural network algorithms in their work. The results of their study showed that the best algorithm in segmentation is RNN. DNN and CNN algorithms were also highly accurate. Such high accuracies were also observed in a number of other ~~studies-research works~~ (Lweesy et al., 2011; Chang et al., 2020; D. Dey et al., 2018; Farouk et al., 2019; ~~Lweesy et al., 2011;~~ L. Wang et al., 2019; ~~T. Wang, C. H. Lu, G. H. Shen, et al., 2019;~~ Chang et al., 2020).

In general, the results of selected studies indicate that ML techniques are useful in diagnosing SA. Unlike the present study, which only diagnoses SA from binary healthy individuals, some studies have classified SA patients, based on the severity of the disease, the results of which again showed the high accuracy of ML techniques. Another approach was to evaluate the ability of models to detect the exact value of Apnea Hypopnea Index (AHI), and the results of these research works were also of acceptable accuracy (Mencar et al., 2019).

Overall, it can be reported that classification of SA patients of healthy individuals using physiological signals is very accurate. In fact, this study showed that in a simpler way than in the past, a portable and wearable system can be designed ~~that has offering~~ many applications in sleep medicine. Other features of these designed systems can be their non-invasiveness and low cost (Atri & Mohebbi, 2015). ~~In fact~~ Additionally, home screening of patients with these methods is straightforward, allowing for the diagnose of SA in time, to prevent more serious medical complications (Ali & Hossen, 2020). Moreover, the capability of the methods in accurate clinical diagnoses has also been considered (Akşahin et al., 2015). Therefore, the results of ML techniques in diagnosing the disease can assist physicians with the selection of ~~the~~ best diagnosis methods. The ~~results-analysis of the reported results in~~ collected studies show~~ed~~ that the use of morphological changes in ECG waves and the features that can be extracted from it, especially PSD waves, can distinguish patients from the healthy individuals. Other characteristics based on time, energy and entropy were the other effective characteristics in classifying patients from healthy. Another important point is that the use of techniques based on deep learning can be very effective in training the system. The selection of features in these methods is also performed with a high accuracy. It is recommended to use deep learning neural networks or SVM and its derivatives (LS-SVM, SVM-HMM) to design new systems. ~~The Our~~ results also demonstrat~~e~~d that increasing the layers of the neural network can have a positive effect on the performance of the classification system.

Among the challenges for the designed classification systems is the accuracy of the systems distinguishing mild apnea of healthy. Another issue to consider is that almost all existing systems are based on OSA, ~~thus,~~ it seems that their performance still requires further investigation to identify and classify other types of SA. Another issue to consider is that ML faces challenges that may affect the results obtained. One of the issues with ML algorithms is that they use random models to train their data. This means that if the same model is retrained with the same data, different values of the parameters may be reported. In other words, the reproducibility of the models is one of the issues that should be considered (Beam, Manrai, & Ghassemi, 2020).

4.2- Limitations

One of the limitations of this work is that most of the selected studies had used the same dataset. This made it impossible to review data from the same area, so the impact of environmental and individual factors on the incidence of SA was not measured. The small samples used in these datasets is another factor that can affect the generalizability of the results. Another limitation observed was the lack of reporting true positives, true negatives, false positives and false negatives in most studies. Accordingly, it was not possible to perform meta-analysis on the data.

4.3- Conclusion

A comprehensive review of selected existing studies ~~have~~has confirmed the effectiveness of ML techniques in diagnosing SA. Since neurological, hormonal, and respiratory changes are quite effective on the ECG, HRV and EDR were the most common ECG features extracted for classification. The features of PSD waves obtained from ECG analysis appear to be very useful in the diagnosis of SA. It was also observed that SVM and Neural Network algorithms are highly accurate in detecting SA. Additionally, other ML techniques such as ~~KNN~~k-NN, RF, and LR performed well in classifying SA-related data. There was no significant difference between the parameters related to the classification based on the complete ECG record and ECG components. However, the ML performance seems to have been better in the full record classification. In future studies, more up-to-date datasets can be used to classify SA. Furthermore, datasets with higher number of records, and the use of samples in various geographical areas are other research areas that can be explored as part of the future of work.

Declarations

Ethics approval and consent to participate

Ethics approval (IR.KUMS.REC.1400.027) was received from the ethics committee of deputy of research and technology, Kermanshah University of Medical Sciences (3011544).

Consent for publication

Not applicable.

Availability of data and materials

Datasets are available through the corresponding author, upon reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

By Deputy for Research and Technology, Kermanshah University of Medical Sciences (IR) (3011544). The funder had no role in the study process.

Authors' contributions

NS, MM, HGH and HKH contributed to the design the study and participated in most of the study stages. HGH and MM extracted the data. AH, AA and AD produced the manuscript. And NS and AH contributed to the interpretation of the study. All authors have read and approved the manuscript.

Acknowledgements

The authors would like to thank the Student Research Committee of Kermanshah University of Medical Sciences. This study is the result of a project registered with the code 3011544 in the Student Research Committee of Kermanshah University of Medical Sciences.

References

- Akşahin, M., Erdamar, A., Firat, H., Ardiç, S., & Erođul, O. (2015). Obstructive sleep apnea classification with artificial neural network based on two synchronic HRV series. *Biomedical Engineering - Applications, Basis and Communications*, 27(2). doi:10.4015/S1016237215500118
- Al-Angari, H. M., & Sahakian, A. V. (2012). Automated Recognition of Obstructive Sleep Apnea Syndrome Using Support Vector Machine Classifier. *Ieee Transactions on Information Technology in Biomedicine*, 16(3), 463-468. doi:10.1109/titb.2012.2185809
- Ali, S. Q., & Hossen, A. (2020). Identification of obstructive sleep apnea using artificial neural networks and wavelet packet decomposition of the HRV signal. *Journal of Engineering Research*, 17(1), 24-33. doi:10.24200/tjer.vol17iss1pp24-33
- Ali, S. Q., Khalid, S., & Belhaouari, S. B. (2019). A Novel Technique to Diagnose Sleep Apnea in Suspected Patients Using Their ECG Data. *Ieee Access*, 7, 35184-35194. doi:10.1109/ACCESS.2019.2904601
- Atri, R., & Mohebbi, M. (2015). Obstructive sleep apnea detection using spectrum and bispectrum analysis of single-lead ECG signal. *Physiol Meas*, 36(9), 1963-1980. doi:10.1088/0967-3334/36/9/1963
- Babaeizadeh, S., White, D. P., Pittman, S. D., & Zhou, S. H. (2010). Automatic detection and quantification of sleep apnea using heart rate variability. *J Electrocardiol*, 43(6), 535-541. doi:10.1016/j.jelectrocard.2010.07.003
- Baek, J. W., Kim, Y. N., Kim, D. E., & Lee, J. H. (2014). Computer-aided detection with a portable electrocardiographic recorder and acceleration sensors for monitoring obstructive sleep apnea. *Sensors and Transducers*, 167(3), 80-87.
- Bali, J., Nandi, A., Hiremath, P. S., & Patil, P. G. (2018). Detection of sleep apnea in ecg signal using pan-tompkins algorithm and ann classifiers. *Compusoft*, 7(11), 2852-2861.
- Bozkurt, F., Ucar, M. K., Bozkurt, M. R., & Bilgin, C. (2020). Detection of Abnormal Respiratory Events with Single Channel ECG and Hybrid Machine Learning Model in Patients with Obstructive Sleep Apnea. *Irbm*, 41(5), 241-251. doi:10.1016/j.irbm.2020.05.006
- Bozkurt, S., Bostanci, A., & Turhan, M. (2017). Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf Med*, 56(4), 308-318. doi:10.3414/me16-01-0084
- Bsoul, M., Minn, H., & Tamil, L. (2011). Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG. *Ieee Transactions on Information Technology in Biomedicine*, 15(3), 416-427. doi:10.1109/titb.2010.2087386

- Chang, H. Y., Yeh, C. Y., Lee, C. T., & Lin, C. C. (2020). A Sleep Apnea Detection System Based on a One-Dimensional Deep Convolution Neural Network Model Using Single-Lead Electrocardiogram. *Sensors (Basel)*, *20*(15). doi:10.3390/s20154157
- Chen, L., Zhang, X., & Song, C. (2015). An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Transactions on Automation Science and Engineering*, *12*(1), 106-115. doi:10.1109/TASE.2014.2345667
- Dempsey, J. A., Veasey, S. C., Morgan, B. J., & O'Donnell, C. P. (2010). Pathophysiology of Sleep Apnea. *Physiological Reviews*, *90*(1), 47-112. doi:10.1152/physrev.00043.2008
- Dey, A. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174-1179.
- Dey, D., Chaudhuri, S., & Munshi, S. (2018). Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomedical Engineering Letters*, *8*(1), 95-100. doi:10.1007/s13534-017-0055-y
- Dyken, M. E., & Im, K. B. (2009). Obstructive sleep apnea and stroke. *Chest*, *136*(6), 1668-1677.
- Eiseman, N. A., Westover, M. B., Mietus, J. E., Thomas, R. J., & Bianchi, M. T. (2012). Classification algorithms for predicting sleepiness and sleep apnea severity. *Journal of Sleep Research*, *21*(1), 101-112. doi:10.1111/j.1365-2869.2011.00935.x
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In *machine learning in radiation oncology* (pp. 3-11): Springer.
- Erdenebayar, U., Kim, Y. J., Park, J. U., Joo, E. Y., & Lee, K. J. (2019). Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput Methods Programs Biomed*, *180*, 105001. doi:10.1016/j.cmpb.2019.105001
- Fariha, M. A. Z., Ikeura, R., Hayakawa, S., & Tsutsumi, S. (2020). Analysis of Pan-Tompkins Algorithm Performance with Noisy ECG Signals. *Journal of Physics: Conference Series*, *1532*, 012022. doi:10.1088/1742-6596/1532/1/012022
- Farouk, F. N. B. M., Anwar, T., & Zakaria, N. B. (2019). Hybrid bayesian network in neural network based deep learning framework for detection of obstructive sleep apnea syndrome. *International Journal of Engineering and Advanced Technology*, *9*(1), 4922-4926. doi:10.35940/ijeat.A2077.109119
- Fatimah, B., Singh, P., Singhal, A., & Pachori, R. B. (2020). Detection of apnea events from ECG segments using Fourier decomposition method. *Biomedical Signal Processing and Control*, *61*. doi:10.1016/j.bspc.2020.102005
- FAUST, O., ACHARYA, U. R., NG, E. Y. K., & FUJITA, H. (2016). A REVIEW OF ECG-BASED DIAGNOSIS SUPPORT SYSTEMS FOR OBSTRUCTIVE SLEEP APNEA. *Journal of Mechanics in Medicine and Biology*, *16*(01), 1640004. doi:10.1142/s0219519416400042
- Ferini-Strambi, L., Baietto, C., Di Gioia, M., Castaldi, P., Castronovo, C., Zucconi, M., & Cappa, S. (2003). Cognitive dysfunction in patients with obstructive sleep apnea (OSA): partial reversibility after continuous positive airway pressure (CPAP). *Brain research bulletin*, *61*(1), 87-92.
- Foster, K. R., Koprowski, R., & Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online*, *13*(1), 94. doi:10.1186/1475-925X-13-94
- Gottlieb, D. J., & Punjabi, N. M. (2020). Diagnosis and Management of Obstructive Sleep Apnea: A Review. *JAMA*, *323*(14), 1389-1400. doi:10.1001/jama.2020.3514
- Hassan, A. R. (2016). Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting. *Biomedical Signal Processing and Control*, *29*, 22-30. doi:10.1016/j.bspc.2016.05.009
- Hassan, A. R., & Haque, M. A. (2016a). Computer-aided obstructive sleep apnea identification using statistical features in the emd domain and extreme learning machine. *Biomedical Physics and Engineering Express*, *2*(3). doi:10.1088/2057-1976/2/3/035003

- Hassan, A. R., & Haque, M. A. (2016b). Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating. *Biocybernetics and Biomedical Engineering*, *36*(1), 256-266. doi:10.1016/j.bbe.2015.11.003
- Hassan, A. R., & Haque, M. A. (2017). An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting. *Neurocomputing*, *235*, 122-130. doi:10.1016/j.neucom.2016.12.062
- Hayano, J., Watanabe, E., Saito, Y., Sasaki, F., Fujimoto, K., Nomiya, T., . . . Sakakibara, H. (2011). Screening for obstructive sleep apnea by cyclic variation of heart rate. *Circulation: Arrhythmia and Electrophysiology*, *4*(1), 64-72.
- Jafari, A. (2013). Sleep apnoea detection from ECG using features extracted from reconstructed phase space and frequency domain. *Biomedical Signal Processing and Control*, *8*(6), 551-558. doi:10.1016/j.bspc.2013.05.007
- Jaiswal, J., Shastry, A. H., Ramesh, A., Chickabasaviah, Y. T., Arimappamagan, A., & Santosh, V. (2016). Spectrum of primary intracranial tumors at a tertiary care neurological institute: A hospital-based brain tumor registry. *Neurology India*, *64*(3), 494.
- Janbakhshi, P., & Shamsollahi, M. (2018). Sleep apnea detection from single-lead ECG using features based on ECG-derived respiration (EDR) signals. *Irbm*, *39*(3), 206-218.
- Kaguara, A., Nam, K., & Reddy, S. (2014). *A deep neural network classifier for diagnosing sleep apnea from ECG data on smartphones and small embedded systems.*
- Kainulainen, S., Töyräs, J., Oksenberg, A., Korkalainen, H., Sefa, S., Kulkas, A., & Leppänen, T. (2019). Severity of desaturations reflects OSA-related daytime sleepiness better than AHI. *Journal of Clinical Sleep Medicine*, *15*(8), 1135-1142.
- Kendzierska, T., Gershon, A. S., Hawker, G., Tomlinson, G., & Leung, R. S. (2014). Obstructive sleep apnea and incident diabetes. A historical cohort study. *American journal of respiratory and critical care medicine*, *190*(2), 218-225.
- Kesper, K., Canisius, S., Penzel, T., Ploch, T., & Cassel, W. (2012). ECG signal analysis for the assessment of sleep-disordered breathing and sleep pattern. *Medical & Biological Engineering & Computing*, *50*(2), 135-144. doi:10.1007/s11517-011-0853-9
- Khandoker, A. H., Karmakar, C. K., & Palaniswami, M. (2009). Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings. *Comput Biol Med*, *39*(1), 88-96. doi:10.1016/j.combiomed.2008.11.003
- Khandoker, A. H., Palaniswami, M., & Karmakar, C. K. (2009). Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. *IEEE Trans Inf Technol Biomed*, *13*(1), 37-48. doi:10.1109/titb.2008.2004495
- Li, K., Pan, W., Li, Y., Jiang, Q., & Liu, G. (2018). A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal. *Neurocomputing*, *294*, 94-101. doi:10.1016/j.neucom.2018.03.011
- Li, M., & Zhou, Z. (2007). Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *37*(6), 1088-1098. doi:10.1109/TSMCA.2007.904745
- Lweesy, K., Fraiwan, L., Khasawneh, N., & Dickhaus, H. (2011). New automated detection method of OSA based on artificial neural networks using P-wave shape and time changes. *J Med Syst*, *35*(4), 723-734. doi:10.1007/s10916-009-9409-z
- Mencar, C., Gallo, C., Mantero, M., Tarsia, P., Carpagnano, G. E., Foschino Barbaro, M. P., & Lacedonia, D. (2019). Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics Journal*, *26*(1), 298-317. doi:10.1177/1460458218824725
- Mendez, M. O., Bianchi, A. M., Matteucci, M., Cerutti, S., & Penzel, T. (2009). Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE Trans Biomed Eng*, *56*(12), 2838-2850. doi:10.1109/tbme.2009.2029563

- Mendez, M. O., Ruini, D. D., Villantieri, O. P., Matteucci, M., Penzel, T., Cerutti, S., & Bianchi, A. M. (2007). Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model. *Annu Int Conf IEEE Eng Med Biol Soc, 2007*, 6106-6109. doi:10.1109/iembs.2007.4353742
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.
- Moody, G. B., Mark, R. G., Bump, M. A., Weinstein, J. S., Berman, A. D., Mietus, J. E., & Goldberger, A. L. (1986). Clinical validation of the ECG-derived respiration (EDR) technique. *Computers in cardiology, 13*(1), 507-510.
- Nakayama, C., Fujiwara, K., Sumi, Y., Matsuo, M., Kano, M., & Kadotani, H. (2019). Obstructive sleep apnea screening by heart rate variability-based apnea/normal respiration discriminant model. *Physiol Meas, 40*(12), 125001. doi:10.1088/1361-6579/ab57be
- Nguyen, H. D., Wilkins, B. A., Cheng, Q., & Benjamin, B. A. (2014). An online sleep apnea detection method based on recurrence quantification analysis. *IEEE J Biomed Health Inform, 18*(4), 1285-1293. doi:10.1109/jbhi.2013.2292928
- Nishad, A., Pachori, R. B., & Acharya, U. R. (2018). Application of TQWT based filter-bank for sleep apnea screening using ECG signals. *Journal of Ambient Intelligence and Humanized Computing, 1-12*. doi:10.1007/s12652-018-0867-3
- Parati, G., Lombardi, C., & Narkiewicz, K. (2007). Sleep apnea: epidemiology, pathophysiology, and relation to cardiovascular risk. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 293*(4), R1671-R1683. doi:10.1152/ajpregu.00400.2007
- Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A., & Peter, J. H. (2000). *The apnea-ECG database* (Vol. 27).
- Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W., & Hla, K. M. (2013). Increased prevalence of sleep-disordered breathing in adults. *American journal of epidemiology, 177*(9), 1006-1014.
- Pinho, A., Pombo, N., Silva, B. M. C., Bousson, K., & Garcia, N. (2019). Towards an accurate sleep apnea detection based on ECG signal: The quintessential of a wise feature selection. *Applied Soft Computing, 83*. doi:10.1016/j.asoc.2019.105568
- Pombo, N., Silva, B. M. C., Pinho, A. M., & Garcia, N. (2020). Classifier Precision Analysis for Sleep Apnea Detection Using ECG Signals. *Ieee Access, 8*, 200477-200485. doi:10.1109/ACCESS.2020.3036024
- Portier, F., Portmann, A., Czernichow, P., Vascaut, L., Devin, E., Benhamou, D., . . . Muir, J. F. (2000). Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. *American journal of respiratory and critical care medicine, 162*(3), 814-818.
- Rachim, V. P., Li, G., & Chung, W. Y. (2014). Sleep apnea classification using ECG-signal wavelet-PCA features. *Biomed Mater Eng, 24*(6), 2875-2882. doi:10.3233/bme-141106
- Rekha, B. B., Kandaswamy, A., & Ramanathan, R. M. P. L. (2018). Ensemble classification approach for screening of obstructive sleep apnoea using ECG. *International Journal of Biomedical Engineering and Technology, 27*(1-2), 139-150. doi:10.1504/ijbet.2018.093102
- Rundo, J. V., & Downey, R. (2019). Chapter 25 - Polysomnography. In K. H. Levin & P. Chauvel (Eds.), *Handbook of Clinical Neurology* (Vol. 160, pp. 381-392): Elsevier.
- Sharma, H., & Sharma, K. K. (2016). An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions. *Comput Biol Med, 77*, 116-124. doi:10.1016/j.compbimed.2016.08.012
- Sharma, H., & Sharma, K. K. (2020). Sleep apnea detection from ECG using variational mode decomposition. *Biomedical Physics and Engineering Express, 6*(1). doi:10.1088/2057-1976/ab68e9

- Sharma, M., Agarwal, S., & Acharya, U. R. (2018). Application of an optimal class of antisymmetric wavelet filter banks for obstructive sleep apnea diagnosis using ECG signals. *Comput Biol Med*, *100*, 100-113. doi:10.1016/j.compbiomed.2018.06.011
- Sharma, M., Raval, M., & Acharya, U. R. (2019). A new approach to identify obstructive sleep apnea using an optimal orthogonal wavelet filter bank with ECG signals. *Informatics in Medicine Unlocked*, *16*. doi:10.1016/j.imu.2019.100170
- Smruthy, A., & Suchetha, M. (2017). Real-Time Classification of Healthy and Apnea Subjects Using ECG Signals With Variational Mode Decomposition. *IEEE Sensors Journal*, *17*(10), 3092-3099. doi:10.1109/JSEN.2017.2690805
- Song, C., Liu, K., Zhang, X., Chen, L., & Xian, X. (2016). An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals. *IEEE Trans Biomed Eng*, *63*(7), 1532-1542. doi:10.1109/tbme.2015.2498199
- Torres-Alba, D., Gemma, D., Armada-Romero, E., Rey-Blas, J. R., López-de-Sá, E., & López-Sendon, J. L. (2013). Obstructive sleep apnea and coronary artery disease: from pathophysiology to clinical implications. *Pulmonary medicine*, *2013*.
- Torres, G., Sánchez-de-la-Torre, M., & Barbé, F. (2015). Relationship between OSA and hypertension. *Chest*, *148*(3), 824-832.
- Travieso, C. M., Alonso, J. B., del Pozo, M., Ticay, J. R., & Castellanos-Dominguez, G. (2014). Building a Cepstrum-HMM kernel for Apnea identification. *Neurocomputing*, *132*, 159-165. doi:10.1016/j.neucom.2013.04.048
- Tripathy, R. K. (2018). Application of intrinsic band function technique for automated detection of sleep apnea using HRV and EDR signals. *Biocybernetics and Biomedical Engineering*, *38*(1), 136-144. doi:10.1016/j.bbe.2017.11.003
- Varon, C., Caicedo, A., Testelmans, D., Buyse, B., & Van Huffel, S. (2015). A Novel Algorithm for the Automatic Detection of Sleep Apnea From Single-Lead ECG. *IEEE Trans Biomed Eng*, *62*(9), 2269-2278. doi:10.1109/tbme.2015.2422378
- Veasey, S. C., & Rosen, I. M. (2019). Obstructive Sleep Apnea in Adults. *New England Journal of Medicine*, *380*(15), 1442-1449. doi:10.1056/NEJMc1816152
- Wang, H., Parker, J. D., Newton, G. E., Floras, J. S., Mak, S., Chiu, K.-L., . . . Bradley, T. D. (2007). Influence of obstructive sleep apnea on mortality in patients with heart failure. *Journal of the American College of Cardiology*, *49*(15), 1625-1631.
- Wang, L., Lin, Y. F., & Wang, J. (2019). A RR interval based automated apnea detection approach using residual network. *Comput Methods Programs Biomed*, *176*, 93-104. doi:10.1016/j.cmpb.2019.05.002
- Wang, T., Lu, C. H., & Shen, G. H. (2019). Detection of Sleep Apnea from Single-Lead ECG Signal Using a Time Window Artificial Neural Network. *Biomed Res Int*, *2019*. doi:10.1155/2019/9768072
- Wang, T., Lu, C. H., Shen, G. H., & Hong, F. (2019). Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network. *PeerJ*, *7*. doi:10.7717/peerj.7731
- Wang, X. W., Cheng, M. W., Wang, Y. F., Liu, S. H., Tian, Z. H., Jiang, F., & Zhang, H. J. (2020). Obstructive sleep apnea detection using ecg-sensor with convolutional neural networks. *Multimedia Tools and Applications*, *79*(23-24), 15813-15827. doi:10.1007/s11042-018-6161-8
- Yildiz, A., Akin, M., & Poyraz, M. (2011). An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings. *Expert Systems with Applications*, *38*(10), 12880-12890. doi:10.1016/j.eswa.2011.04.080
- Young, T., Skatrud, J., & Peppard, P. E. (2004). Risk Factors for Obstructive Sleep Apnea in Adults. *JAMA*, *291*(16), 2013-2016. doi:10.1001/jama.291.16.2013

Table legend:

Table1: Search strategies, and keywords

Table2: Classification result based on per record or per segment data

Figures legend:

Figure1: An overview of the implementation of machine learning techniques

Figure2: PRISMA flow diagram for study selection

Table 1: Search strategies, and keywords

Database	Search strategy	Date	number
PubMed	(Artificial Intelligence(Jaiswal et al.) OR "machine learning"[tiab]s OR "neural networks"[tiab] OR "Bayesian models"[tiab] OR "deep learning"[tiab] OR "dimensionality reduction"[tiab] OR "decision trees"[tiab] OR "ensemble learning"[tiab] OR "instance based models"[tiab] OR "support vector machines"[tiab]) AND (Sleep Apnea[tiab] OR "Sleep-Disordered Breathing"[tiab] OR Sleep Apnea, Central[tiab] OR Sleep Apnea Syndrome(Jaiswal et al.) OR Sleep Apnea, Obstructive(Jaiswal et al.) OR "Sleep Apnea Hypopnea Syndrome"[tiab] OR OSA[TIAB] OR OSAHS[TIAB] OR "Sleep Apnea Syndromes"[tiab]) AND (Electrocardiography(Jaiswal et al.) OR ECG[tiab] OR EKG[tiab] OR Electrocardiogram[TIAB] OR "electrocardiogram derived respiration"[TIAB])	24/10/2020	59
Scopus	TITLE-ABS-KEY("Artificial Intelligence" OR "machine learning" OR "neural networks" OR "Bayesian models" OR "deep learning" OR "dimensionality reduction" OR "decision trees" OR "ensemble learning" OR "instance based models" OR "support vector machines") AND TITLE-ABS-KEY("Sleep Apnea" OR "Sleep-Disordered Breathing" OR "Sleep Apnea, Central" OR "Sleep Apnea Syndrome" OR "Obstructive Sleep Apnea" OR "Sleep Apnea Hypopnea Syndrome" OR OSA OR OSAHS OR "Sleep Apnea Syndromes") AND TITLE-ABS-KEY(Electrocardiography OR ECG OR EKG OR Electrocardiogram OR "electrocardiogram derived respiration")	24/10/2020	210
WOS	TS=("Artificial Intelligence" OR "machine learning" OR "neural networks" OR "Bayesian models" OR "deep learning" OR "dimensionality reduction" OR "decision trees" OR "ensemble learning" OR "instance based models" OR "support vector machines") AND TS=("Sleep Apnea" OR "Sleep-Disordered Breathing" OR "Sleep Apnea, Central" OR "Sleep Apnea Syndrome" OR "Obstructive Sleep Apnea" OR "Sleep Apnea Hypopnea Syndrome" OR OSA OR OSAHS OR "Sleep Apnea Syndromes") AND TS=(Electrocardiography OR ECG OR EKG OR Electrocardiogram OR "electrocardiogram derived respiration")	24/10/2020	103
IEEE Explore	("Artificial Intelligence" OR "machine learning" OR "neural networks" OR "Bayesian models" OR "deep learning" OR "dimensionality reduction" OR "decision trees" OR "ensemble learning" OR "instance based models" OR "support vector machines") AND ("Sleep Apnea" OR "Sleep-Disordered Breathing" OR "Sleep Apnea, Central" OR "Sleep Apnea Syndrome" OR "Obstructive Sleep Apnea" OR "Sleep Apnea Hypopnea Syndrome" OR OSA OR OSAHS OR "Sleep Apnea Syndromes") AND (Electrocardiography OR ECG OR EKG OR Electrocardiogram OR "electrocardiogram derived respiration")	24/10/2020	29

Table 2: Classification result based on per record or per segment data

		Per segment		
First author, Year,	NO, of segments	classification	ACC%	Other parameters
Al-Angari, H. M. 2012,	Total segments: 39575 (22908 normal, 16607 apnea)	SVM (linear) C= 10	68.8	Sen: 51.6, spec: 81.4
		SVM (polynomial) C= 10	69.5	Sen: 54.8, spec: 80.1
Atri, R. 2015	Total segments: 16479 Use 13,000 segments in 10 fold cross validation	LS-SVM	95.57	Sen: 98.64, Spec: 92.51
Babaeizadeh, S. 2010,	Total segments: 34313 17010 training, 17268 test segments	QC	84.7	Sen: 76.7, Spec: 89.6, PPV: 81.8, NPV:86.3
Bozkurt, F. 2020,	Total segments: 2460 (1242 Apnea, 1218 Normal), 1230 training segments, 1230 test segments	DT (113 feature)	79.11	Sen: 76, sp:82
		kNNk-NN (90 feature)	82.2	Sen: 78, sp:75
		SVM (113 feature)	84.15	Sen: 76, sp:82
		Ensamble (113 feature)	85.12	Sen: 76, sp:82
Bsoul, M. 2011,	Total segments: 14700 segments in 5,10 and 35 fold cross validation	SVM (linear) C=32	91.16	Sen:89.12, Spec: 92.35, F: 90.70
		SVM (poly, d=2) C=0.5, $\gamma=0.5$	89.85	Sen:88.25, Spec: 88.25, F: 90.82
		SVM (RBF) C=2, $\gamma=0.5$	90.86	Sen: 89.02, Spec: 91.94, F: 90.46
		SVM (MLP) C=0.5, $\gamma=0.5$	80.45	Sen:74.66, Spec: 83.96, F: 79.04
Chang, H. Y. 2020,	Total segments: 34230, 17234 test segments, 16979 training set	CNN	87.9	Sen:81.1, Spec:92
Dey, D. 2017,	10787 training segments, 10787 test segments (4987Apnea, 5800 Normal)	CNN	98.1	Sen:97.82, Spec: 99.2, PPV: 99.06, NPV: 98.14
Erdenebayar, U. 2019	Total segments: 43522 , 37338 training segments, 6184 test segments (1623Apnea, 4561 Normal)	DNN	93.1	Sen: 93 Spec:94
		IDCNN	98.5	Sen: 99 Spec:99
		2DCNN	95.9	Sen: 96 Spec:96
		RNN	85.4	Sen: 97 Spec:87
		RNN(LSTM)	98	Sen: 98 Spec:98
		GRU	99	Sen: 99 Spec:99
		CNN	98.91	Sen:97.82, Spec:99.20 PPV:99.06, NPV:98.14
Farouk, F. N. B. M. 2019,	10787 training segments, 10787 test segments	Bagging	91.44	Sen:91.61 Spec: 92.52 Pre: 88.20
		kNNk-NN	90.57	Sen: 89.13 Spec: 91.56 Pre: 86.90
		SVM	92.59	Sen: 89.70 Spec: 94.67 Pre: 91.27
		LogitBoost	85.84	Sen: 79.17 Spec: 89.97 Pre: 83.04
Hassan, A. R.1 2016,	-	RBM	38.79	AdaBoost:
		SVM	59.22	Sen: 81.99, Spec: 90.72
		Naïve Bays	62.15	
		ANN	81.37	
		RF	82.70	
		kNNk-NN	83.32	
		Bagging	83.33	
		LDA	83.72	
		AdaBoost	87.33	
Hassan, A. R. 2, 2016,	-	ANN	68.52	ELM:
		NBC	39.47	Sen: 85.20, Spec: 82.79
		RBM	61.20	
		kNNk-NN	69.72	

		AdaBoost	80.07	
		Bagging	79.82	
		RF	79.26	
		DA	64.60	
		ELM	83.77	
Hassan, A. R. 3, 2016,	-	Bagging	85.97	Sen:84.14, Spec: 86.83
Hassan, A. R. 4, 2017,	-	LS-SVM	31.88	Rusboost
		ELM	53.02	Sen: 87.58, Spec: 91.49
		PRAZEN-PNN	60.95	
		SVM	72.4	
		KNNK-NN	79.77	
		Bagging	84.29	
		RF	84.49	
		Adaboost	86.94	
		Rusboost	88.88	
Jafari, A. 2013,	Total segments: 16711 segments, 10000 training, 6711 test segments	SVM	94.80	Sen:94.16, Spec:95.42
Li, K. 2018,	Total segments: 33979 16857 training set, 17122 test segments (6517A, 10605N)	ANN	78.3	Sen:66.6 Spec:85.4
		ANN-HMM	83	Sen:91.5 Spec:77.7
		SVM	78.6	Sen:66.5 Spec:86.1
		SVM-HMM	84.7	Sen:68.8 Spec:94.5
		Decision fusion	84.7	Sen: 88.9Spec:82.1
Lweesy, K. 2011,	Total segments: 1500 , 1052 training set, 224 test segments	ANN	92.3	Sen:90.1, Spec:94.4,
Mendez, M. O. 2009.	Total segments: 24432 , 12077 training segments, 12355 test segments	KNNK-NN Feature=10	88	Sen:86, Spec: 87
		ANN Feature=10	88	Sen: 89, Spec:86
Nguyen, H. D. 2014,	-	ANN	83.23	Sen:85.57, Spec:79.09
		SVM	84.14	Sen:93.72, Spec:65.88
		Decision fusion	85.26	Sen:86.37, Spec:83.47
Nishad, A. 2018.	-	RF	92.78	Sen: 90.95, Spec:93.91
Pinho, A. 2019,	17401 segments in 10 fold cross validation	ANN Features: 20	82.12	Sen:88.41, Spec:72.29
		SVM Features: 70	75.18	Sen:86.79, Spec:56.45
		LDA Features: 20	62.93	Sen:83.98, Spec:28.40
		PLS Features: 20	64.49	Sen:57.78, Spec:66.05
		REG Features: 20	65.13	Sen:62.23, Spec:65.65
		WienerHopf Features: 20	64.05	Sen:58.14, Spec:65.07
		aNBC Features: 44	62.12	Sen:0, Spec:62.12
		PLA Features: 6	61.36	Sen:36.84, Spec:61.70
		LMS Features: 84	61.72	Sen:28.70, Spec:62.35
Rachim, V. P. 2014	-	SVM (RBF), C=10, $\sigma=0.5$, PCA=5	93.91	Sen: 95.20, Spec: 92.65
Rekha, B. B. 2018.	-	SVM (without feature reduction)	91	Sen: 90.38, Spec: 91.54
		RF (with feature reduction)	94.32	Sen:92.98, spec: 94.77
Sharma, H. 2016,	Total segments: 32727 segments, 16845 training segments, 15873 test segments	KNNK-NN	73.3	Sen:72.5, Spec:73.8 AUC:73.8
		MLPNN	81.2	Sen:77.5, Spec:83.4 AUC:80.7
		LS-SVM	82.6	Sen:76.7, Spec:88.2 AUC:82
		SVM	83.8	Sen:79.5, Spec:88.4 AUC:83.4

Sharma, H. 2020,	Total segments: 34313, 17045 training, 17268 test segments	LR	85.6	Sen:82.1, SP:90.2 AUC: 0.93
		LDA	82.5	Sen: 73, Spec: 88.5 AUC: 0.89
		BDT	84.7	SEN:79.5, Spec: 87.5 AUC: 0.92
		ADT	81.6	Sen: 74, Spec: 86.3 AUC: 0.88
		KNNK-NN	87.5	Sen: 84.9, Spec: 88.2 AUC: 0.93
		ANN	86.1	Sen: 84, Spec: 86.9, AUC: 0.82
		LS-SVM	86.2	Sen: 81.3, Spec: 87.7 AUC: 0.91
		SVM	85.3	Sen: 82.5, Spec: 88 AUC: 0.91
Sharma, M. 2018,	Total segments: 16993 in 35 fold cross validation	Weighted KNNK-NN	89.1	Sen: 91.8, Spec: 84.9, PPV: 90.36, NPV:87
		CT	83.5	Sen: 86.3, Spec:78.9, PPV: 87.1, NPV:77.8
		LD	66.1	Sen:70.9, Spec: 56.7, PPV: 76.6, NPV: 49.8
		LR	66.5	Sen: 70, Spec: 58.1 PPV:80, NPV: 44.8
		LS-SVM(kernel)	90.11	Sen: 90.9, Spec:88.9 PPV: 93, NPV: 85.8
Sharma, M. 2019,	Total segments: 16993 in 35 fold cross validation	KNNK-NN	90.3	Sen: 91.5, Spec: 88.5, PPV: 92.8, NPV:86.6
		CT	82.9	Sen: 91, Spec:76.8, PPV: 85.8, NPV:78.1
		LD	68.9	Sen:86.9, Spec: 40, PPV:70, NPV: 65.4
		LR	71.1	Sen: 86, Spec: 47.2 PPV:72.4, NPV: 67.7
		Gaussian SVM	90.87	Sen: 92.43, Spec:88.33 PPV: 92.8, NPV: 88.3
Song, C. 2016,	17268 test segments (6550Apnea, 10718Normal)	SVM	81.2	Sen: 75.7, Spec: 84.7
		SVM-HMM	86.2	Sen: 82.6, Spec: 88.4
		LR	81.2	Sen: 74.4, Spec: 85.4
		LR+HMM	86.2	Sen: 80, Spec: 89.9
		LDA	80.5	Sen: 83.1, Spec:78.9
		LDA+HMM	85.3	Sen: 77.5, Spec: 90.1
		KNNK-NN	80.7	Sen:75.3, Spec: 83.9
		KNNK-NN +HMM	84.5	Sen: 74, Spec: 90.8
Travieso, C. M. 2014	Total segments: 21865, 10599 training, 11266 test segments (4375A, 6891N)	SVM-HMM	99.2	Sen: 98.8, Spec: 99.5, PPV: 99.2, NPV:99.3
Varon, C. 2015.	Total segments: 32 477 , 6000 training segments, 26477 test segments	LDA	71.43	Sen: 71.74, Spec:71.2
		SVM(LIN)	71.16	Sen: 74.63, Spec: 68.51
		SVM(POLY)	72.6	Sen: 78.36, Spec: 68.22
		SVM(RBF)	73.86	Sen: 78.2, Spec: 70.55
		LS-SVM(LIN)	71.78	Sen: 71.48, Spec: 76.26
		LS-SVM(POLY)	73.43	Sen: 73.4, Spec:73.43
		LS-SVM(RBF)	84.74	Sen: 84.71, Spec: 84.69
		LDA	79.86	Sen: 81.22, Spec: 79.71
Varon, C.2 2015.	Total segments: 5205, 3000 training segments, 2205 test segments	SVM(LIN)	75.83	Sen: 86.46, Spec: 74.6
		SVM(POLY)	77.78	Sen: 88.21, Spec: 76.57
		SVM(RBF)	81.9	Sen: 82.97, Spec: 81.78
		LS-SVM(LIN)	80.05	Sen: 80.35, Spec: 80.01
		LS-SVM(POLY)	81.68	Sen: 81.66, Spec: 81.7
			83.95	Sen: 78.81, Spec: 84.56
		LS-SVM(RBF)		
Wang, L. 2019.	Total segments: 16 988 (6496Apnea , 10492 Normal) in 10 fold cross validation	CNN	90.97	Sen: 83.43, Spec: 85.5
		residual network	94.39	Sen: 93.04, Spec: 94.95
Wang, T. 2019.	Total segments 34,313,	LR	81.5	Sen: 72, Spec: 87.4

	17045 training segments, 17268 test segments.	LDA	81.8	Sen: 70.9, Spec: 88.4
		SVM	80.6	Sen: 72.1, Spec: 85.6
		MLP	81.4	Sen: 74.3, Spec: 85.7
		TW-MLP	87.3	Sen: 85.1, Spec: 88.7
Wang, T.2 2019.	-	SVM	81.4	Sen: 76.9, Spec: 84.3
		LR	80.8	Sen: 75.7, Spec: 84
		KNNK-NN	77.5	Sen: 68.1, Spec: 83.4
		MLP	81.1	Sen: 71.3, Spec: 87.2
		LET-NET5 CNN	87.6	Sen: 83.1, Spec: 90.3
Wang, T.3 2019.	-	SVM	70.6	Sen: 32.7, Spec: 83.3
		LR	69.6	Sen: 34.7, Spec: 81.3
		KNNK-NN	66.1	Sen: 38.1, Spec: 75.4
		MLP	67.2	Sen: 38.5, Spec: 76.8
		-CNN	71.2	Sen: 26.6, Spec: 86.9
Per record				
First author, Year,	No. of records	ML algorithm	ACC%	Other parameters
Akşahin, M. 2015,	Total records: 20 record in 3 and 5 cross validation (10 Apnea, 10 Normal)	FFNN	99	-
Al -Angari, H. M. 2012,	Total records: 100 records, 50 training, 50 test	SVM(linear) C=5	79	Sen: 79.6, spec: 78.4
		SVM(polynomial) C=5, 10	78	Sen: 67.3, spec: 88.2
Ali, S. Q. 2020,	Total record: 80, 40 training, 40 test	FFNN	87.5	Sen:86.67, spec: 90
		PNN	85	Sen:86.67, spec: 80
Babaeizadeh, S. 2010,	Total records : 60 records, 30 training, 30 test record (20 Apnea , 10 Normal)	QC	100	Sen: 100, Spec: 100, PPV: 100, NPV: 100
Baek, J. W. 2014,	Total records : 20 records in 10 fold cross validation (10Apnea, 10Normal)	ML	85	Sen: 90, spec: 80
Bali, J. 2018,	Total record: 70 , 35 training , 35 test record	ANN-LM	91	Sen:91, spec: 92, PR:95
		ANN-SCG	95	Sen:94, spec: 91, PR: 96
Chang, H. Y. 2020,	Total record:70 , 35 training , 35 test record	CNN	97.1	Sen:95.7, spec:100
Chen, L. 2015	Total record: 90 subjects, 59 training, 31 test record (23 Apnea, 8 Normal)	SVM (RBF kernel), C=3	97.41	Sen:98.99, Spec: 92.87
		SVM(polynomial) C=5, Order=1	97.03	Sen:99.16, Spec: 90.91
Kaguara, A. 2014,	Total records: 35 records in 5 fold cross validation	DNN (fold 4)	91	
Eiseman, N. A. 2012,	Total record: 4647 records, (2090 Apnea, 2557 Normal) in 20 fold cross validation	SVM	65.4	Sen: 43.4, Spec: 83.5 PPV: 68.3, NPV: 64.4
		Naive Bayes	63.02	Sen: 39, Spec: 82.7 PPV: 64.8 NPV: 62.4
Khandoker, A. H. 2009,	Total record: 60 records, 30 training , 30 test (20Apnea, 10Normal)	SVM(Poly) Feature:5,6,7or8, C=0.1,1,1	100	Sen: 100, Spec: 100
		SVM(Linear) Feature: 6 or7, C=10	100	Sen: 100, Spec: 100
		LD	90	Sen: 100, Spec: 93
		KNNK-NN (K =1)	80	Sen: 90, Spec: 83
		PNN ($\sigma=0.5$)	80	Sen: 50, Spec: 70
Khandoker, A. H.2. 2009,	Total record: 125 records, 83 training, 42 test	SVM(Polynomial) D= 3, C =0.8	100	Sen: 100, Spec: 100
		SVM(Linear) C =10	98.8	Sen: 100, Spec: 94.44
		SVM(RBF) $\sigma=0.5$, C= 10	96.39	Sen: 98.46, Spec: 88.89

Li, K. 2018,	Total record: 70 records, 35 training, 35 test	DNN	100	Sen: 100, Spec: 100
Mendez, M. O. 2007.	Total record: 50 record, 25 training , 25 test	KNNK-NN	85	Sen: 83.90, Spec:88.50
Nakayama, C. 2019	Total record: 61 record (25Apnea, 36Normal)	RF	85	Sen: 76, Spec: 92
Rachim, V. P. 2014	Total record : 35 record (22Apnea, 13Normal) in 10 fold cross validation	SVM	94.3	Sen: 100, Spec: 81.3
Sharma, H. 2016,	Total records: 70 records, 35 training, 35 test	KNNK-NN	77.3	Sen:100, Spec:63.6, AUC:77.3
		MLPNN	93.4	Sen:95.8, Spec:90.9, AUC:93.4
		LS-SVM(RBF)	97.8	Sen:95.8, Spec:100 AUC:97.8
		SVM(RBF)	97.8	Sen:95.8, Spec:100 AUC:97.8
		LR	97.14	Sen:100, Spec:90.91AUC:97.8
Sharma, H. 2020,	Total records: 70 records, 35 training , 35 test	LDA	100	Sen:100, Spec:100 AUC:0.95
		BDT	97.14	Sen:100, Spec:90.91 AUC:1
		ADT	94.28	Sen:91.67, Spec:100 AUC: 0.95
		KNNK-NN	97.14	Sen:100, Spec: 90.91 AUC: 0.95
		ANN	97.14	Sen:100, Spec: 90.91 AUC: 0.95
		LS-SVM	94.28	Sen:100, Spec: 90.91 AUC: 0.95
		SVM	97.14	Sen:95.8, Spec: 90.91 AUC: 0.93
		SVM	97.5	Sen:95.45, spec: 100, PPV: 100, NPV: 94.7
Smruthy,A. 2017	Total record: 40 record (all healthy participant) in 10 fold cross validation	SVM	97.5	Sen:95.45, spec: 100, PPV: 100, NPV: 94.7
	Total record: 25 record in 10 fold cross validation	SVM	95	Sen: 100, Spec: 80, PPV: 94.12, NPV:1
Song, C. 2016,	Total record: 30 record , 20 training and 10 test	SVM	80	Sen: 100, Spec: 36.4
		SVM-HMM	97.1	Sen: 95.8, Spec:100
		LR	74.3	Sen: 100, Spec: 18.2
		LR+HMM	97.1	Sen:95.8, Spec: 100
		LDA	68.6	Sen:100, Spec: 0
		LDA+HMM	97.1	Sen: 95.8, spec: 100
		KNNK-NN	91.4	Sen: 100, Spec: 72.7
		KNNK-NN +HMM	91.4	Sen: 87.5, Spec: 100
Tripathy, R. K. 2018,	Total records: 31 records in 10 fold cross validation	KELM (RBF) K=5	78.71	Sen:83.45, Spec:73.27
		KELM (LINEAR) K=10	75	Sen:91.26, Spec:58.19
		KELM (POLY) K=10	83.46	Sen: 85.6, Spec: 81.30
		KELM (CWK) K=6	78.71	Sen: 79.06, Spec: 78.33
		LR	91.4	Sen: 100, Spec: 75
Wang, T. 2019.	Total record: 70 records in 7 fold cross validation	LDA	88.6	Sen: 100, Spec: 66.7
		SVM	82.9	Sen: 100, Spec: 50
		MLP	82.9	Sen: 100, Spec: 50
		TW-MLP	97.1	Sen: 100, Spec: 91.7
		SVM	88.6	Sen: 100, Spec: 66.7
Wang, T.2 2019.	Total record: 70 records in 10 fold cross validation	LR	88.6	Sen: 100, Spec: 66.7
		KNNK-NN	82.9	Sen: 100, Spec: 50
		MLP	85.7	Sen: 95.7, Spec: 66.7
		LET-NET5 CNN	97.1	Sen: 100, Spec: 91.7
		SVM	92.3	Sen: 100, Spec: 50
Wang, T.3 2019.	Total record: 25 record in 10 fold cross validation	LR	84.6	Sen: 100, Spec: 50
		KNNK-NN	84.6	Sen: 90.9, Spec: 0
		MLP	92.3	Sen: 100, Spec: 50
		LET-NET5 CNN	92.3	Sen: 90.9, Spec: 100
		LET-NET5 CNN	97.8	
Wang, X. W.	Total record: 30 record	LET-NET5 CNN	97.8	

2020.	20 training and 10 test			
Yildiz, A. 2011.	Total record: 30 record in 10 fold cross validation	LS-SVM(RBF)	100	Sen: 100, Spec: 100
		LS-SVM(POLY)	96.7	Sen: 100, Spec:95
		LS-SVM(LIN)	96.7	Sen: 100, Spec: 95

FFNN: Feed Forward Neural Network, SVM: Support Vector Machine, LS-SVM: least-square support vector machine, QC: quadratic classifier, ANN: Artificial Neural Network, ANN-SCG: Artificial Neural Network- Scaled Conjugate Gradient, ANN-LM: Artificial Neural Network- Levenberg-Marquardt algorithm, CNN: Convolutional neural network, DNN: Deep Neural Network, RNN: Recurrent Neural Networks DT: Decision tree, [kNNk-NN](#): k-nearest neighbors algorithm, GRU: Gated Recurrent Unit, RBF: Restricted Boltzmann Machine, LDA: linear Discriminant Analysis, Bagging: Bootstrap Aggregating, Adaboost: Adaptive boosting, HMM: Hidden Markov Model, PLS: Partial Least Squares Regression, REG: Regression Analysis, aNBC: Augmented Naive Bayesian Classifier, PLA: Perceptron Learning Algorithm, LMS: Least Mean Square, WienerHopf: Wiener-Hopf equation, BDT: Bagged Decision Tree, CT: Complex Tree, LR: Logistic R

Figure 1:

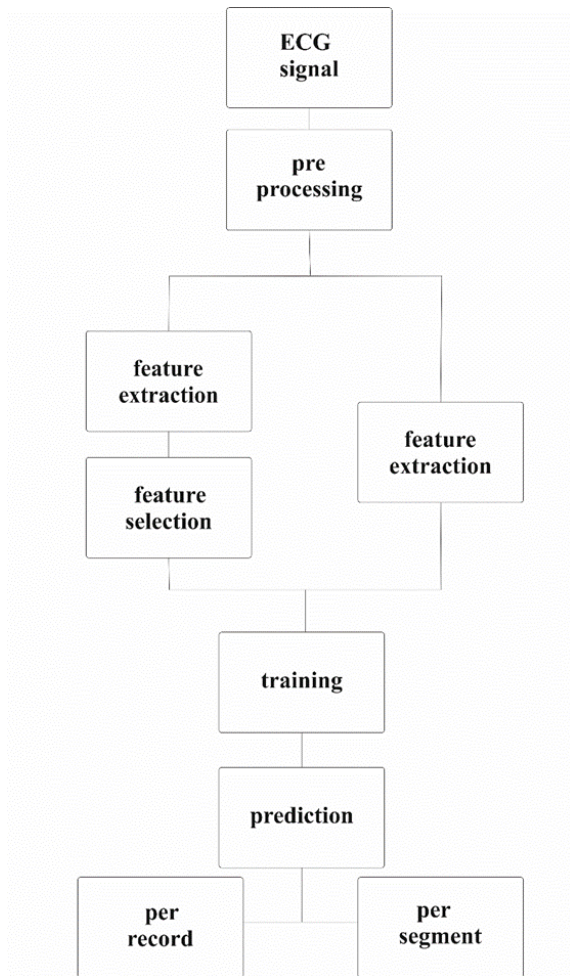


Figure 1: An overview of the implementation of machine learning techniques [within the context](#)

Figure 2:

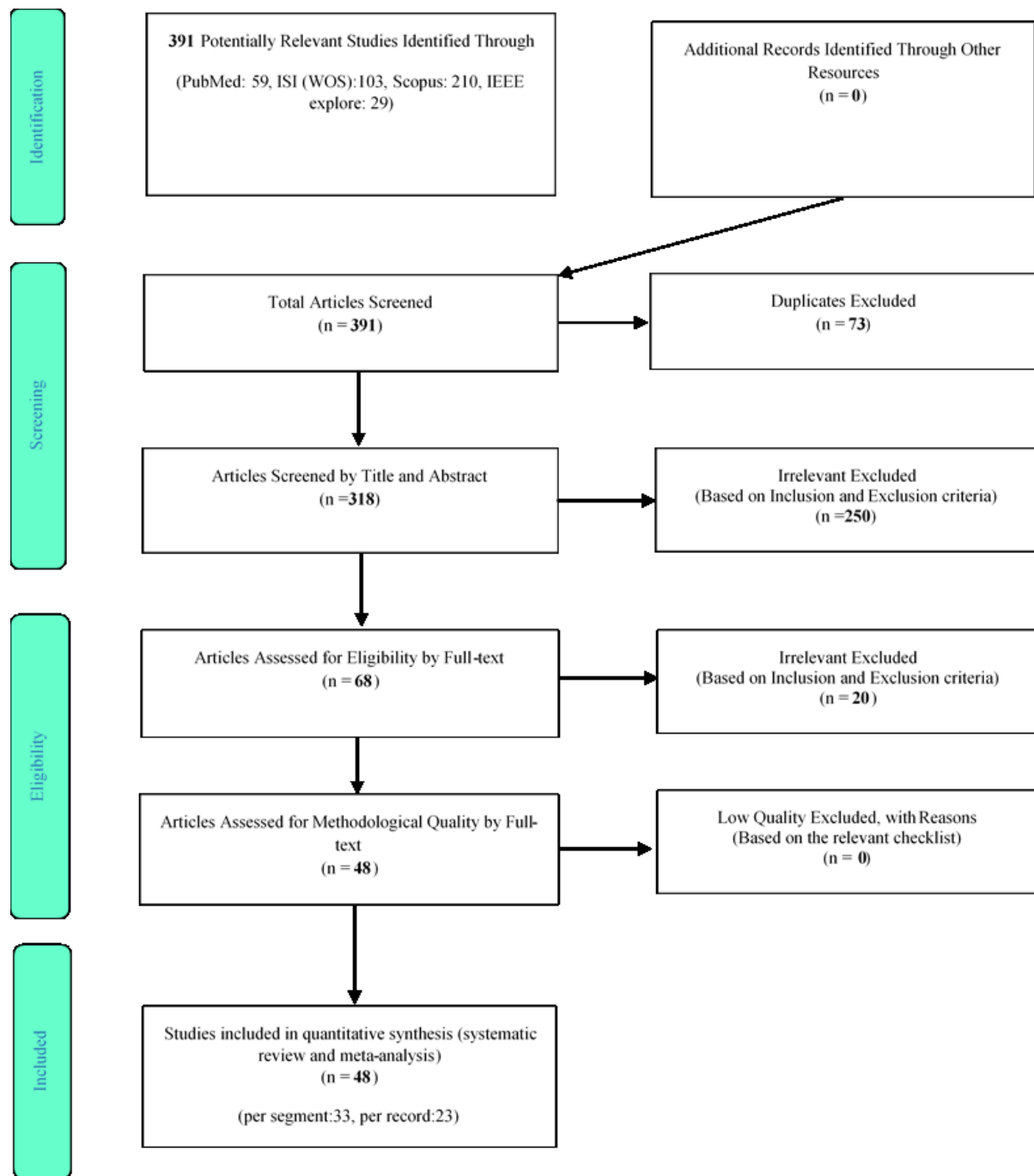


Figure 2: PRISMA flow diagram for study selection