BIOCYBERNETICS AND BIOMEDICAL ENGINEERING 43 (2023) 684—699

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

End-to end decision support system for sleep apnea detection and Apnea-Hypopnea Index calculation using hybrid feature vector and Machine learning

Recep Sinan Arslan [a,*], *Hasan Ulutas* [b], *Ahmet Sertol Köksal* [b], *Mehmet Bakir* [b], *Bülent Çiftçi* [c]

[a] *Department of Computer Engineering, Faculty of Engineering, Architecture and Design, Kayseri University, Kayseri, Turkey*
[b] *Department of Computer Engineering, Faculty of Engineering and Architecture, Bozok University, Yozgat, Turkey*
[c] *Department of Chest Diseases, Faculty of Medicine, Yuksek Ihtisas University, Ankara, Turkey*

A R T I C L E I N F O

Article history:
Received 1 August 2023
Received in revised form
6 September 2023
Accepted 5 October 2023
Available online 13 October 2023

Keywords:
Sleep Apnea-Hypopnea Detection
Machine Learning
Discrete Signal Processing
AHI

A B S T R A C T

Sleep apnea is a disease that occurs due to the decrease in oxygen saturation in the blood and directly affects people's lives. Detection of sleep apnea is crucial for assessing sleep quality. It is also an important parameter in the diagnosis of various other diseases (diabetes, chronic kidney disease, depression, and cardiological diseases). Recent studies show that detection of sleep apnea can be done via signal processing, especially EEG and ECG signals. However, the detection accuracy needs to be improved. In this paper, a ML model is used for the detection of sleep apnea using 19 static sensor data and 2 dynamic data (Sleep score and Arousal). The sensor data is recorded as a discrete signal and the sleep process is divided into 4.8 M segments. In this work, 19 different sensor data sets were recorded with polysomnography (PSG). These data sets have been used to perform sleep scoring. Then, arousal status marking is done. Model training was carried out with the feature vector consisting of 21 data obtained. Tests were performed with eight different machine learning techniques on a unique dataset consisting of 113 patients. After all, it was automatically determined whether people were diseased (a kind of apnea) or healthy. The proposed model had an average accuracy of 97.27%, while the recall, precision, and f-score values were 99.18%, 95.32%, and 97.20%, respectively. After all, the model that less feature engineering, less complex classification model, higher dataset usage, and higher classification performance has been revealed.

* Corresponding author at: Kayseri University, Department of Computer Engineering, Talas, Kayseri, Turkey.
E-mail address: recepsinanarslan@kayseri.edu.tr (R.S. Arslan).
https://doi.org/10.1016/j.bbe.2023.10.002
0168-8227/

1. Introduction

Sleep disorders refer to a wide range of conditions that affect the quality, duration, and pattern of a person's sleep. These disorders can have various causes, including medical conditions, psychological factors, lifestyle choices, or environmental factors. Insomnia, sleep apnea, narcolepsy, restless legs syndrome, parasomnias, circadian rhythm disorders are common sleep disorders [1]. Sleep apnea is estimated to affect approximately 936 million people globally, according to a study published in the Lancet Respiratory Medicine in 2019. However, it's important to note that this number may vary depending on the population studied and the diagnostic criteria used. Furthermore, it is estimated that up to 80% of moderate to severe cases of sleep apnea are undiagnosed. Many people may not be aware that they have the condition and remain untreated, which can lead to various health complications [2]. Sleep disorders are usually defined as cyclical stops in breathing (apneas) or a decrease in the size of the airflow (hypopneas), both of which cause brief awakenings and breaks in sleep. Both apnea types can be divided as obstructive or central. Obstructive means that upper airway blockage occurs, while central is related to the absence or reduction of neural input to upper airway muscles [3]. While obstructive sleep apnea has a strong association with cardiovascular diseases, central sleep apnea leads to stops in ventilation due to the neural drive of nerves [4]. Sometimes, apnea-like events may occur when 10 s or more of shallow breathing occurs, which may cause a 30–90% decrease in normal airflow. This event may end with a drop in oxyhemoglobin saturation (desaturation event) or electroencephalographic (EEG) arousal but does not look like apnea and can be described as hypopnea [5]. For computing the apnea-hypopnea index (AHI), sleep doctors group hypopnea, obstructive sleep apnea, and central apnea together. This index is the sum of apnea and hypopnea events divided by sleep time in minutes, and this value is multiplied by 60. The index values for normal breathing, mild, moderate, and severe are less than 5, between 5 and 15, between 15 and 30, and finally, greater or equal to 30 [6]. There are any other indexes that can be used for characterization of sleep disorders as Respiratory Disturbance Index (RDI) which is similar to the AHI but includes additional respiratory events such as respiratory effort-related arousals (RERAs). Oxygen Desaturation Index (ODI) which measures the frequency of drops in blood oxygen levels during sleep. Respiratory Arousal Index (RAI) which quantifies the frequency of respiratory-related arousals from sleep. Including both obstructive and central events [7]. These indices, along with the AHI, provide a comprehensive assessment of sleep apnea severity and its impact on various physiological parameters. Although AHI index is the most common method for evaluating the sleep disorders, it is affected from sleep posture as supine (back) sleeping, lateral (side) sleeping, prone (stomach) sleeping directly [8].

After giving brief information about apnea, polysomnography (PSG) is used for the detection of sleep disorders in a patient. This test is performed in sleep clinics to record the events of a patient by employing several sensors during the whole night [9]. Some of these sensors are pulse-oximetry, EEG, Electrooculogram (EOG), Electromyogram (EMG), Electrocardiogram (ECG), Respiratory monitor, Microphone, Airflow, Thoracic belt, and abdominal belt. Although PSG is the most common tool for detection of sleep apnea as well as scoring sleep stages, it can be considered the gold standard. It has drawbacks, such as long analysis times for sleep doctors or technicians, as well as lengthy training times for this staff [10,11]. Machine learning which is a subfield of artificial intelligence (AI) that focuses on the development of algorithms that can learn from and make predictions on data can effectively be used to eliminate these drawbacks.

In the past few years, machine learning based works developed to save a great amount of time for users during sleep stage scoring and apnea disease detection [12,13]. Machine learning methods have common limitations as requirement of large and high quality data, subjectivity of data evaluation by expert and computational complexity and resource requirements [14]. For these reasons, most of the researchers tried to eliminate them by expanding data set which is evaluated by experts in the related field and using cloud technologies. Machine learning methods can be used to estimate the probabilities of no, mild, moderate, and severe obstructive sleep apnea and such approaches may improve accurate initial OSA screening and help referring only the suspected moderate or severe OSA patients to sleep laboratories for the expensive tests.

Sleep apnea detection methods employ some breathing measurement results, oximetry, EEG, and ECG sensor readings [15]. These inputs can effectively be used for machine learning algorithms to guess the sleep apnea event detection. The machine learning process covers data collection, feature selection, algorithm choice, training, validation, and testing data set phases [16]. In the current state of the art, while some researchers prefer to use online available datasets for sleep apnea detection [17], some of them prefer to prepare their own datasets for detection [18]. In addition to the dataset preparation, some of the researchers choose small numbers or single-channel inputs, while others prefer to use multichannel PSG inputs [19]. Because sleep experts use all channels for event marking and sleep stage scoring, all channels are used in this work for a sleep apnea detection system.

The used dataset is annotated according to the rules defined by the AASM and used by the sleep doctor for marking apnea, hypopnea, and arousals. The sleep expert marks the related epochs if there is a peak signal drop that is greater than 90% captured by the oronasal thermal sensor with a duration of more than 10 s as apnea. If the peak signal drops by more than 30% for at least 10 s, these times are called hypopneas. If a patient shows efforts by flattening of the inspiratory portion of oronasal airflow with EEG arousal, these epochs are marked as arousal [20]. Our dataset contains all of these events with different numbers. For this reason, 113 patient recordings have been used for dataset creation, training, and testing phases.

A unique dataset containing 113 patient records was prepared for this paper. After preparation, the data preprocessing classifier is selected by running different methods such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), as well as tree-based algorithms such as Random Forest (RF), Decision Tree (DT), and Extra Tree (ET). All these classifiers have been effectively used in both sleep stage scoring and apnea disorder detection [21–24].

Most valuable contributions of this study is a model proposal for the use of a multi-channel feature vector in learning models for the detection of sleep apnea disorders (Apnea-Hypopnea or Healthy) with high performance. The data imbalance regarding the absence of disease in the data set was effectively resolved by the resampling method. The study was carried out with the dataset originally collected for this study. A new and unique sleep data has been added to the literature. Every patient has a record of around 4.8 million records, so a large dataset is created from 113 patients and 500 million records for increasing accuracy and reliability.

This study has been prepared to classification of sleep disorders (Apnea-Hypopnea or Healthy) with the original dataset, feature vector generation structure and machine learning model.

With the proposed model in this study, it is aimed to distinguish patients with sleep apnea and healthy individuals with a feature set using PSG recording. The developed model is aimed to have a structure that uses resources efficiently while working with high performance. Therefore, unlike all studies in the literature, discrete signal and machine learning were used instead of continuous signal-based feature extraction or image-based classification. In this way, an efficient model has been created that does not require deep feature engineering studies (zero crossing, slope sign change, svd, Wilson aptitude etc.) or complex classification models (R-CNN, VGG, EfficientNet or custom CNN). <u>As a result, the model has less feature engineering, less complex classification model, higher dataset usage, and higher classification performance.</u>

In section 2, original and current studies on the detection of sleep apnea are mentioned in detail. In section 3, the structure of the model proposed in this study is presented. In section 4, the results of the tests are shown comparatively and the results are discussed. In the last section, a general evaluation of the study was made and suggestions for the future were made.

2. Related work

Before explaining the related work in the field of sleep apnea classification, brief information about machine learning will be given. Machine learning is a technique that uses mathematical modelling to detect abnormal patterns in a dataset with a trained model on this dataset. According to the learning technique, it can be categorized as supervised, unsupervised, or reinforcement learning [25]. When we look at the current state of the art, while some of the works focus on the use of minimum number of sensors, the others use maximum number of PSG inputs to increase the accuracy for detecting sleep apnea syndrome as ours do. Proposed work use 19 inputs of PSG device to increase the accuracy for the detection of sleep apnea with different methods as clinician do. Because the proposed model employs the sleep stage as an input for detection of sleep apnea syndrome. The following studies are presented to summarize the last five years of work for clear explanation.

Li et al., developed a method which combines DNN with machine learning classifiers to detect sleep apnea by using single lead ECG channel. Their dataset size is 35 and 85% accuracy obtained for per segment classification [26]. Steenkiste et al., presented an algorithm based sleep apnea detection system which automatically extracts features and detect sleep apnea events in respiratory signals. The algorithm is evaluated on the Sleep-Heart-Health-Study-1 dataset and obtained an accuracy of 80% by employing PSG signals and long short-term memory (LSTM) techniques [27]. Erdenebayar et al. [28], conducted experiments by using LSTM classifier to extract features on own dataset which have sleep apnea events. Their proposal reach to 98.5% accuracy for apnea, hypopnea events. Mencar et al. [29], guessed the AHI index collected data about demographic characteristics, spirometry values, gas exchange (PaO2, PaCO2) and symptoms of 313 patients other than PSG inputs and developed a machine learning based model. Their model achieved 44.7 % accuracy. Tuncer et al., proposed a deep learning based decision support system for AHI index calculation by using CNN and pulse transition time signals [30]. Their developed model achieved 92.78% accuracy. Feng et al., proposed an unsupervised feature learning model for detection of sleep apnea by using single lead ECG [31]. They added time dependent to the hidden Markov model to increase the performance of their proposal. Chang et al., developed a sleep apnea detection system based on a one dimensional deep convolutional network model by using single lead ECG [32]. They employed ECG readings of MIT physioNet database occupied from 35 recording of subjects. Their proposal occupied from 1D deep CNN model for the detection of apnea events only using 1D ECG signals as input. They extracted 10 classification layers and 4 classification layers in their work which result with an accuracy of maximum 97.1%. ElMoaqet et al proposed a gaussian mixture model (GMM) for detection of sleep apnea from single oronasal airflow record [33].Their proposal is occupied from rule based classifier which is focused on the inter-breath intervals and breath amplitudes. 96 subjects record have been used in this work and they obtained 88.5% accuracy in the field of apnea detection. Mukherjee et al. [34], proposed an Ensemble model for sleep apnea detection which consists majority voting, sum rule and Choquet integral based fuzzy fusion and trainable ensemble using Multi-Layer Perceptron (MLP). Their experiments conducted on the benchmark PhysioNet Apnea-ECG Database which has a detection accuracy of 85.58%. Ramesh et al.[35], proposed a machine learning based sleep apnea detection model which uses waist-to-height ratio, waist circumference, neck circumference, bodymass index, lipid accumulation product, excessive daytime sleepiness, daily snoring frequency and snoring volume. They employed 1479 recording of wisconsin Sleep Cohort dataset. After hyper parameter evaluation and hybrid tuning they obtained 68.06 accuracy in their work. Bricout et al. [36], proposed a system which uses thoracic accelerometer and an abdominal accelerometer to propose an airflow estimation that enables estimation of AHI and they can be used for AHI estimation and sleep apnea detection. The motivation of that study the use of alternative signals such as Accelerometry-Derived Respiratory index may reduce the failure rate of home PSG as nasal signals are not working. They developed a machine learning based model on 28 subjects and obtained 89% accuracy. In that work. Arslan et al., conducted a sleep stage scoring system by using PSG device inputs and machine learning algorithms [37]. Their proposed model employs machine learning algorithms on private database and tree based algorithms obtained 95.25% accuracy. Yang et al., developed a a novel deep neural network named 1D squeeze-and-excitation (SE) residual group network (1D-SEResGNet) for detection of sleep apnea [38]. They applied their model on apnea ECG dataset and had an accuracy of 90.3%. Garcai et al., developed a 2D CNN for detection of sleep apnea in children using airflow and oximetry [39]. They applied their model on 2612 pediatric subjects to have accurate estimation of the severity and enhanced identification of mild and moderate cases other than algorithm outperformed other feature-engineering or single-signal approaches. Arslan et al., developed another sleep stage scoring system by using deep learning models in [40]. Their model reached to 91,6% average accuracy on fifty patients by using all PSG inputs. Han and Oh developed an application of various machine learning techniques to predict obstructive sleep apnea syndrome severity in [41]. They employed 4014 patients' data together with supervised and unsupervised learning method as hierarchical agglomerative clustering, K-means, bisecting K-means algorithm, Gaussian mixture model, and feature engineering. Classification work in their work made by XGBoost, LightGBM, CatBoost, and Random Forest for prediction of AHI index and apnea severity which reached to 88% accuracy. Hu et al., proposed another ECG based Semi-Supervised Learning model for Low-cost personalized apnea detection model by using 60 subject recordings [42]. They utilized the unsupervised deep learning model CNN based auto encoder for detection of sleep apnea events. Experimental results showed improvements with an accuracy of 86.3%. Maniaci et al., developed a machine learning based model for detection of severity of OSA by employing 498 patient's data. Their proposed method uses sleep stage scoring as an input in addition to the 8 channel of PSG. By this way this work is similar to this proposed work with smaller accuracy of 86% [43]. Han and Oh, proposed a clustering method by employing hierarchical agglomerative clustering, K-means, bisecting K-means algorithm, Gaussian mixture model, and feature engineering with 4014 patient data of NYX corporation [41]. Their proposed model do not uses PSG inputs to develop an alternative to them with considerably higher accuracy which is 91%. Since our proposal want to develop a model for clinician does, it differs from this work by applying PSG and sleep stage scores. Tyagi and Agrawal developed a deep belief network based automatic sleep apnea detection system by employing ECG signal of apnea ECG database with an accuracy of 89.9 [44]. Cheng et al. proposed another solution for detection of sleep apnea from bioelectric signal mainly rely on blood oxygen level (SpO2), ECG signal and airflow, with few EEG signal by using multiple instance learning model with an accuracy of 78.2% [45]. Pang et al. aimed to examine faster and less complicated machine learning models by using brain tensor imaging (DTI) to classify OSA from healthy controls with an accuracy of 77% [46]. Strumph et al. evaluated the performance of Belun Ring with second-generation deep learning algorithms in OSA detection and sleep stage classification obstructive sleep apnea (OSA) detection, OSA severity categorization, and sleep stage classification with and minimum accuracy of 85% by using SHHS and Yeh et al. datasets [47]. Abdulla et al., developed a sleep stage classification system by using EEG signals and genetic algorithm based short time fourier transform system with and accuracy of 96% on two publicly available datasets [48]. Hemrajani et al., developed a wearable device and hybrid model for detection of sleep apnea by using Physionet apnea dataset. Their device uses ECG and oxygen saturation signals of detection of sleep apnea with an accuracy of approximately 95% [49]. Huo et al., proposed a machine learning based questionnaire for classifying OSA risk by evaluating risk factor subtypes [50]. Their proposed model is too different from ours it is given for explaining current state of the art. Wang et al., developed a system which uses a photoplethysmography (PPG) optical sensors for getting pulse rate and oxygen saturation. Their model based on multiscale entropy and random forest algorithms for diagnosing sleep apnea with an accuracy of 85% [51].

As briefly given in here 1, this work differs from previous works by classifying sleep disorders as sleep apnea, hypopnea, and arousals by using its own dataset occupied by 113 patients. When the presented work is compared to other machine learning-based sleep apnea classifiers, it moves forward by having more accurate values in three-stage classifiers. Another important point is that this high accuracy value is obtained by using all 19 channels of the PSG device rather than using EEG or ECG channels. This point is important since the sleep doctor uses all channels for marking the field as apnea, hypopnea, and arousals.

3. Methodology

In this study, a unique methodology is proposed that will automatically detect sleep apnea and calculate the AHI index. With the proposed model, it is aimed to achieve high classification performance and to develop a decision support system that can detect sleep disorders. The model scheme designed for this purpose is given in Fig. 1.

The collection and association of patient records is the first step of the proposed model. To do so, a total of 19 static sensor data belonging to EEG, EMG, ECG, Eye, SpO2, Chest Effort, Thermistor, and Body Position were collected for each patient, while sleep scoring and arousal status, which indicate the person's sleep-wake status, and disease level, A total of 6 features consisting of events are recorded. While sleep scoring and arousal status are used as features among 6 events, 4 disease types are used as classes. For this use, it is necessary to combine two groups of data and apply some preprocessing steps. After all, a feature vector containing 21 features that can be used for modeling and a dataset with 4

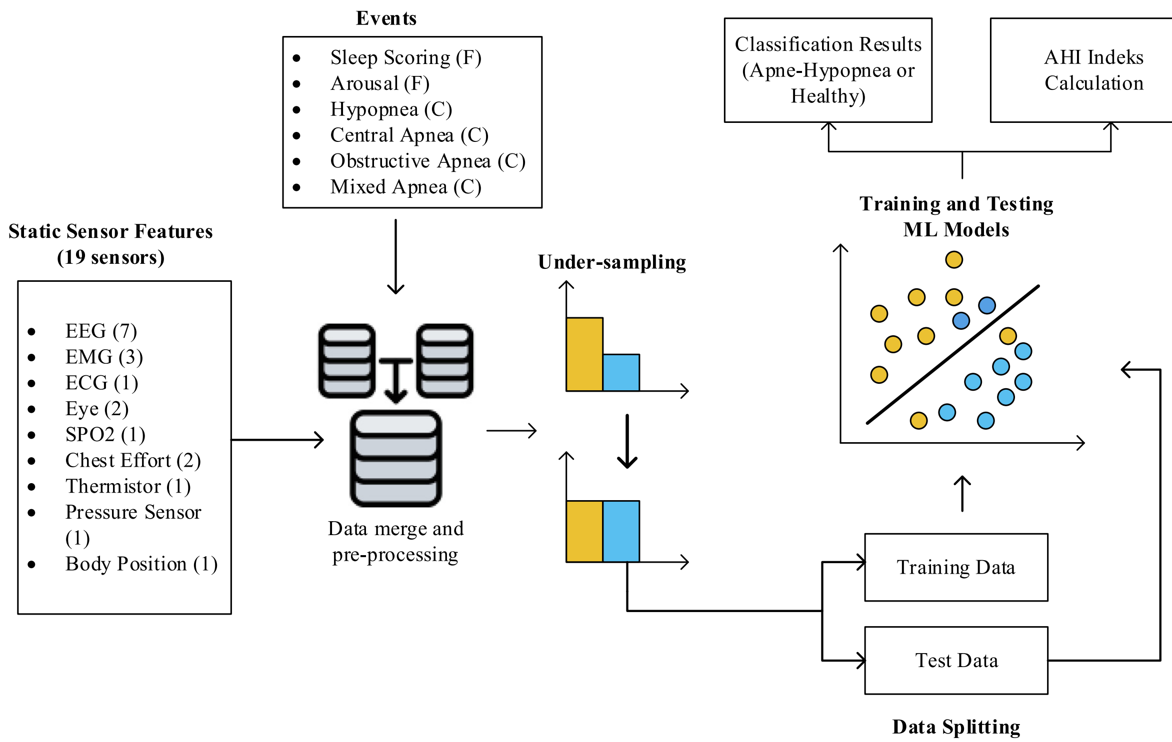

Fig. 1 – End-to-End Sleep Apnea Detection Methodology.

Table 1 – Sampling frequencies and multiplication factors of 19 sensor data.

No	Sensor Name	Sampling frequency	Multiplication factor	The amount of data obtained during the 1-hour period
1	Body Position	1	x200	720 k
2	SPO2	1	x200	720 k
3	Pressure Sensor	100	x2	720 k
4	Thermistor	100	x2	720 k
5	Chest Effort (Channel 1)	100	x2	720 k
6	Chest Effort (Channel 2)	100	x2	720 k
7	Leg EMG (Channel 1)	200	x1	720 k
8	Leg EMG (Channel 2)	200	x1	720 k
9	Eye (Channel 1)	200	x1	720 k
10	Eye (Channel 2)	200	x1	720 k
11	Chin EMG	200	x1	720 k
12	EKG-ECG	200	x1	720 k
13	EEG (Channel 1)	200	x1	720 k
14	EEG (Channel 2)	200	x1	720 k
15	EEG (Channel 3)	200	x1	720 k
16	EEG (Channel 4)	200	x1	720 k
17	EEG (Channel 5)	200	x1	720 k
18	EEG (Channel 6)	200	x1	720 k
19	EEG (Channel 7)	200	x1	720 k

classes are obtained. All these procedures were repeated in 113 patients.

When the data obtained was examined, it was seen that the moments of apnea-hypopnea during the sleep process were very few compared to the moments when they were healthy, even in the most severe patients. This situation affects the classification negatively. For this reason, under sampling was performed according to the “majority” class, and the data was balanced within both classes (Apnea-Hypopnea or Healthy). The balanced data was randomly allocated for training and testing in the next stage and was used in the training and testing steps of eight different machine learning models. As a result, while determining whether the person has apnea or not, the AHI index calculation is made automatically using 4.8 million sleep records, and the apnea level of the people is determined as 4 groups.

Thus, a model that can be used as a decision support system that performs all the operations performed by a sleep specialist doctor has been revealed. In the end, the goal is to be able to mark and sort the sleep-time episodes with and without illness with a high success rate.

3.1. Data merge and pre-processing

Two different data files with an average of 4.8 M records per patient were used in this study. A large amount of data emerges for a total of 113 patients. The software given below as pseudo-code has been prepared to combine 19 features with data from static sensors, where data at different frequencies is collected, with 2 features and 4 classes of data collected as events. Thus, by combining the data, producing a feature vector containing 21 features, combining 4 classes and using them in the AHI index calculation, and marking them as healthy in the records, other than these were performed automatically. In order to be used in the learning model, the AHI indexes of the existing patients were also calculated before the training at this stage.

Sensor data is collected from individuals over an average of 8 h of sleep a night. Each sensor makes measurements at different frequencies (1 Hz, 100 Hz and 200 Hz), but the data collected at lower frequencies is multiplied with a multiplication factor, thus equalizing the amount of all sensor data. As a result, 5.76 M pieces of data will be collected at the end of 8 h of night sleep, 200 pieces per second. However, due to the fact that sleep times are different for each individual, only healthy measurement periods are taken into account, and some other physical difficulties, there is an average of 4.8 M data for each individual, although it varies from person to person. In other words, usable data was collected for a period of approximately 6.5 h. Table 1 shows the sampling frequencies of the sensor data and the multiplication factors used to equalize the amount of data for each sensor.

Algorithm 1. Combination of two record files
input: *ssr*: 2 dimensional array (static sensor records), *er* = 2 dimensional array (event records)
output: cr: 2 dimensional array (combined record file)
l = record file line number (data amount)
ss = static sensor property count
sc = status record property count
for n ← 1 to l **do**
for m ← 1 to ss **do**cr
(n,m) = ssr(n,m)
for p ← 1 to sc **do**
if(er(n,p) == 1 || er(n,p) == 2 || (er(n,p) == 3 || (er(n,p) == 4) **then**cr
(n,m + 1) = “Apnea-Hypopnea”
else thencr
(n,m + 1) = “Healthy”
end for
end for
return cr

A very complex data collection process has to be carried out due to the fact that the data is taken through sleep recording software. For this reason, it is necessary to carry out this pre-processing stage in order to use the data modeling together. In addition, this whole process should be repeated on an individual basis. Because the disease states of people are quite different from each other. Under-sampling as a means to mitigate the class imbalance concern present in our dataset. Ensuring equal effectiveness in identifying instances from both the ’healthy’ and ’diseased’ patient classes was of particular significance in our context. The process of under-sampling involved decreasing the magnitude of the majority class in order to achieve equilibrium with the minority class within the dataset. The approach involved the random elimination of instances from the majority class until a balance was achieved, resulting in both classes being of comparable size. The process of under-sampling was meticulously controlled to ensure that the resultant dataset retained a representative sample of the population. The performance of the model was monitored during the process to ascertain that the reduction in samples from the majority class did not have any adverse impact on the model’s performance. The findings indicate that the ExtraTree classifier model maintained a high level of precision, thereby validating the efficacy of the under-sampling approach employed in the study. The aforementioned methodology enabled us to effectively sustain the model’s performance while guaranteeing an equitable portrayal of the two categories within the dataset. It is our contention that the aforementioned methodology played a pivotal role in the triumph of our model, as evidenced by its exceptional precision across various disease stages.

3.2. Dataset

Sleep recordings of 113 different people were used, which were taken from the Sleep Center in University. An Ethics committee decision was taken to obtain the records. All records have 19 static sensors data belonging to 113 people, and sleep scoring features marked by the sleep doctor, arousal marking indicating the state of being asleep-awake, and 4 types of sleep disease (hypopnea, apnea (mixed, central, and obstructive) data. Static data is also marked by the sleep doctor. A number of preprocessing processes were carried out to use the data together. The table showing the total amount of data per person is as presented in Appendix Table A1.

In the data set, there are 19 sensor data measured with a PSG device, and 2 event data marked by an expert. These two together form the input vector. The output vector in the dataset contains four different types of sleep diseases and this information is marked by the expert. The structure of the dataset is illustrated in Fig. 2.

The best results are obtained when all 19 data from PSG records are used [9]. These data are also used for sleep scoring and determination of arousal events. The sleep specialist decides which sleep stage the patient is in and whether the arousal event has occurred by looking at 19 static features. These two events, together with the 19 static features, are used to determine whether the person they belong to is healthy or has one of the 4 classes of sleep disorders. Detection of disease by looking at only 19 features made machine learning work with lower performance. When the 2 events were used together with these 19 features, the success increased. Therefore, in this study, a total of 21 types of data consisting of 19 features and 2 events were selected.

To calculate the AHI index, it is necessary to determine how many of the patients' data indicate a sleep disorder. In the dataset, there are approximately 4.8 M records from each of the 113 patients, and how many of these records indicated a disease and how many indicated a healthy individual were determined as shown in Appendix Table A1. As seen in Appendix Table A1, for example, 4,747,800 of the data of the patient with PID number 1 were marked as healthy and 16,200 as disease (hypopnea or any type of apnea). Some patients have a count of 0 marks for disease, which shows that people in that condition do not have any sleep sickness. In some patients, disease marking was much higher than the average, which states that the AHI index is high and shows signs of sleeping sickness for most of the sleep time for these patients.

Among 113 patients, eight did not have any signs of sleep sickness. Apart from them, when the AHI index calculation is used according to AASM standards, 72 people have mild sleep apnea, 12 people have moderate and 21 people have severe sleep apnea. It was prepared with a balanced dataset containing all types of patients, and it was used in the tests.

3.3. *Evaluation metrics*

Standard metrics were used to measure the performance of the proposed model, and the formulation structures are given in Equation 1–4.

$$Acc. = \frac{TrueP + TrueN}{TrueP + TrueN + FalseP + False\,N} \tag{1}$$

$$Pre. = \frac{TrueP}{TrueP + FalseP} \tag{2}$$

$$Rec. = \frac{TrueP}{TrueP + False\,N} \tag{3}$$

$$F - score = 2x\frac{Pre. + Rec.}{Pre. + Rec.} \tag{4}$$

As a result of the calculation, it is to observe how accurately the diseases can be detected in 3 different classes. Precision and Recall values were also included in the calculation, since the sample imbalance between the classes did not allow us to evaluate directly with Accuracy. As a result, it was possible to compare the proposed model with similar models.

3.4. *AHI calculation*

The Apnea-Hypopnea Index (AHI) is used to evaluate both obstructive apnea severity and model/device classification/ detection reliability. Evaluation is based on comparison of the coefficient within a fixed reference range [52].

The AHI index, is generally accepted as a gold standard in apnea detection, that term is the value that expresses the degree of respiratory disorder during sleep, taking into account the number of apneas and hypopneas for an average hour interval during sleep. It is a general and accepted index and has four qualitative subranges according to medical guidelines [53]:

The degree of sleep apnea:

- Normal, if AHI is less than or equal to 5 per hour
- Mild, if AHI is between 5 and 15 per hour
- Moderate, if AHI is between 15 and 30 per hour
- Severe, if greater than or equal to 30 per hour

The AHI index calculation will be made for the accuracy of the models and devices for the detection of apnea and checked by comparing them with the reference. For this reason, the AHI index calculation was made according to the classification results obtained in this study, and the results were confirmed.

3.4.1. *Extremely randomized trees (Extra Trees) model*

Tree-based ensemble learning methods are one of the most widely used supervised learning approaches for supervised classification and regression problems. Ensemble learning models, on the other hand, may outperform single models, but whether this construct produces good results depends on how many models are combined in the model prediction. The best performance of tree-based ensemble learning methods is achieved when core learners are independent of each other. This is achieved by using or randomizing very different training algorithms for each decision tree [54].

Extra Tree is a random trees algorithm that works in a classical top-down growth style and aims to make predictions by combining the results of a collection of decision trees without pruning. It creates a forest similar to other tree-based clusters, but cares about randomization to reduce bias that negatively impacts results. It is quite similar to a random forest, except that the nodes in the extra trees are split by choosing completely random breakpoints during the tree initialization phase and the original learning instance is used instead of the preload [55].

When this tree structure is used, it is necessary to select the basic parameters of the model. Accordingly, the minimum number of instances = 2 for a node split and the number of forests in the tree is set to the default value of 100. Since it is faster than entropy, "gini" is determined as the criterion. The maximum depth of the tree is allowed to be extended to reach all the leaves and is not restricted. No preload is used so all data is used to build each tree.

The use of tree-based classifiers in modelling enables the capture of intricate feature relationships owing to their non-linearity. This property is particularly useful for addressing the feature interactions. The potential utility of this approach is noteworthy in the domain of sleep apnea, as the interplay among diverse physiological signals could provide valuable insights into the existence or gravity of the ailment. Robustness to outliers is a desirable property in statistical analysis. It has been observed that tree-based methods exhibit greater robustness to outliers in the data than other methods. The inclusion of outliers in a dataset may enhance the efficacy

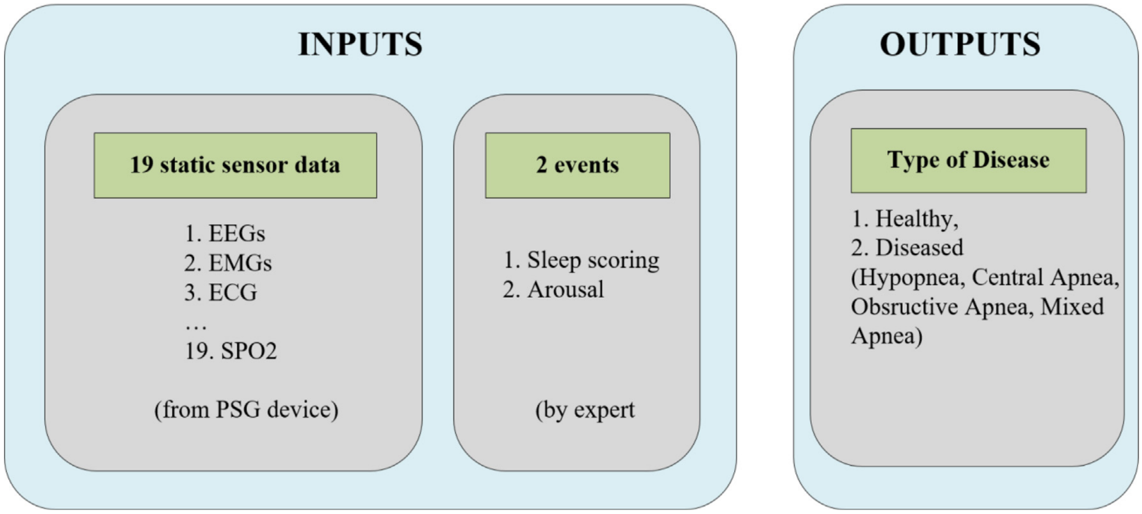

Fig. 2 – Dataset structure.

of tree-based methodologies. The management of missing values can be effectively addressed using tree-based classifiers, obviating the need for imputation methods that may introduce bias. Missing values in the dataset may have been contributing factors. Tree-based classifiers, such as Random Forests and Extra Trees, can conduct automatic feature selection and can effectively indicate the significance of features. This may imply that they exhibited an enhanced capacity to concentrate on the most pertinent characteristics for forecasting sleep apnea.

3.5. *Experimental results and their analysis*

It aims to detect sleep disorders automatically by using the sleep records of 113 people in this work. The data of 113 people was tested separately for 8 different classifiers for this machine learning-based approach, and the average classification results obtained are shown in Table 2. The results are given separately for 3 disease classes according to the AHI index as mild moderate and severe. After that, the average value is shown in the last column. Since apnea refers to the

Table 2 – Average Classification Results for all patients.

Classifier	Patient Level	Accuracy	Recall	Precision	F-Score	Average Accuracy
K-Nearest Neighbors (KNN)	Mild	83.06%	91.61%	72.87%	81.08%	83.38%
	Moderate	83.21%	91.21%	73.64%	81.38%	
	Severe	84.57%	92.07%	75.71%	82.88%	
Gradient Boosting (GB)	Mild	90.47%	92.99%	87.58%	90.18%	84.48%
	Moderate	81.56%	82.96%	79.54%	81.17%	
	Severe	80.64%	83.01%	77.15%	79.91%	
Ada Boosting(AB)	Mild	86.57%	88.30%	84.41%	86.29%	83.98%
	Moderate	78.39%	79.42%	76.72%	78.02%	
	Severe	78.32%	80.55%	74.78%	77.51%	
Logistic Regression(LR)	Mild	66.84%	69.14%	63.87%	65.74%	64.33%
	Moderate	57.40%	58.38%	54.59%	55.79%	
	Severe	59.79%	61.48%	57.38%	58.94%	
Linear Discriminant Analysis(LDA)	Mild	66.27%	68.10%	64.19%	65.57%	66.31%
	Moderate	65.93%	68.57%	64.54%	65.38%	
	Severe	66.68%	68.65%	65.55%	66.48%	
Random Forest (RF)	Mild	97.61%	99.06%	96.13%	97.57%	96.95%
	Moderate	95.41%	97.33%	93.38%	95.31%	
	Severe	95.61%	98.06%	93.06%	95.49%	
Extra Tree(ET)	Mild	97.88%	99.54%	96.21%	97.84%	97.27%
	Moderate	96.00%	98.05%	93.86%	95.90%	
	Severe	95.91%	98.62%	93.13%	95.78%	
Decision Tree(DT)	Mild	95.05%	95.67%	94.39%	95.02%	93.92%
	Moderate	91.39%	92.09%	90.56%	91.32%	
	Severe	91.52%	92.07%	90.89%	91.47%	

complete cessation of airflow for a minimum duration of 10 s during sleep and hypopnea is defined as a partial reduction in airflow accompanied by a drop in blood oxygen levels during sleep, lasting for at least 10 s. These disorders are evaluated together for calculation of AHI index by sleep experts which is also one of the aims of this work. Accordingly, high classification success was achieved with ExtraTree, Random Forest, and Decision Tree classifiers. Accuracy values were 97.27%, 96.95%, and 93.92%, respectively. Recall, Precision and F-score values were also similar to accuracy. The values obtained are one of the highest performances in the literature according to the studies conducted for the detection of apnea. In addition, since there are patients with different disease levels in the dataset, it was also examined whether the proposed model produced different results according to the disease level. Accordingly, 97.88%, 96.00%, and 95.91% values were obtained for Mild, Moderate, and Severe classes for the ExtraTree classifier. These results prove that the model does not show much variability in terms of performance according to the disease level and can produce successful results for all patients.

It is very critical to obtain similar results at different disease levels. Because this situation causes accurate results in the AHI index calculation. If there was a significant change in the performance at different levels, we would not be able to predict the performance of the proposed model since we could not know the condition of the incoming patient. If the patient is mild, the average performance for mild is valid, while the moderate average performance for moderate will need to be considered. In this case, it will not be possible to measure the average performance of the model. Since balanced results are obtained for all levels in the model proposed in this study, it is possible to calculate the average ACC and evaluate it as the performance of the model.

The results obtained on a patient basis for a dual classification as having or not having any disease are examined in more detail below. The tests were repeated for people in 3 different groups and at different disease levels according to the AHI index, and the ExtraTree classifier, which had the highest performance on average, was used as a classifier. When the complexity matrices given in Fig. 3 are examined, the increase in the number of samples tested due to the increase in the AHI index draws attention. The reason of it is can be accepted as undersampling according to the majority class, but the increase in the test sample has no effect on the model. Apart from that, it is seen that the FP and FN values at all three levels are quite low numerically. This low status is balanced in both classes (healthy or diseased). The results show that the proposed model is successful in detecting sleep apnea.

The reason for the numerical increase in FP and FN values depending on the increase in the AHI index is due to the fact that the tests are performed on a patient basis and the number of training and test samples changes. The average accuracy value was 97.88%, 95.99%, and 95.91% for mild, moderate, and severe patients, respectively. The numerical increase does not directly affect the accuracy value, and the average Acc values are equivalent within the three disease grades. This proposed model proves to be successful for all disease levels.

Similarly, k-fold cross validation was performed by choosing k = 10 to prove the accuracy of the results obtained in the complexity matrix for 3 different disease levels. Thus, the tests were performed with a balance between low computational cost and low bias, and they were aimed at measuring the average performance of the model for all sleep sickness levels. The obtained results are shown with a box-plot graph in Fig. 4.

As can be seen in the results, the heights of the boxes are quite low. This shows that the minimum, median, and maximum classification success of the model are very close to each other. When training and testing are performed for randomly allocated samples, the closeness of these values shows that the model is not affected by the sample selection and always produces consistent results.

When the results for mild, moderate, and severe patients were compared, it was seen that the Extra Tree and Random Forest algorithms produced values very close to 1 and that all three levels were close to each other. There was another demonstration proving that the proposed model was not affected by the level of disease. Also, the median value found is quite high, which is a good sign when compared to other studies.

The ROC Curve shows the performance of our proposed model for all classification thresholds and is given in Fig. 5 for each of the 3 disease levels. The AUC gives the total value

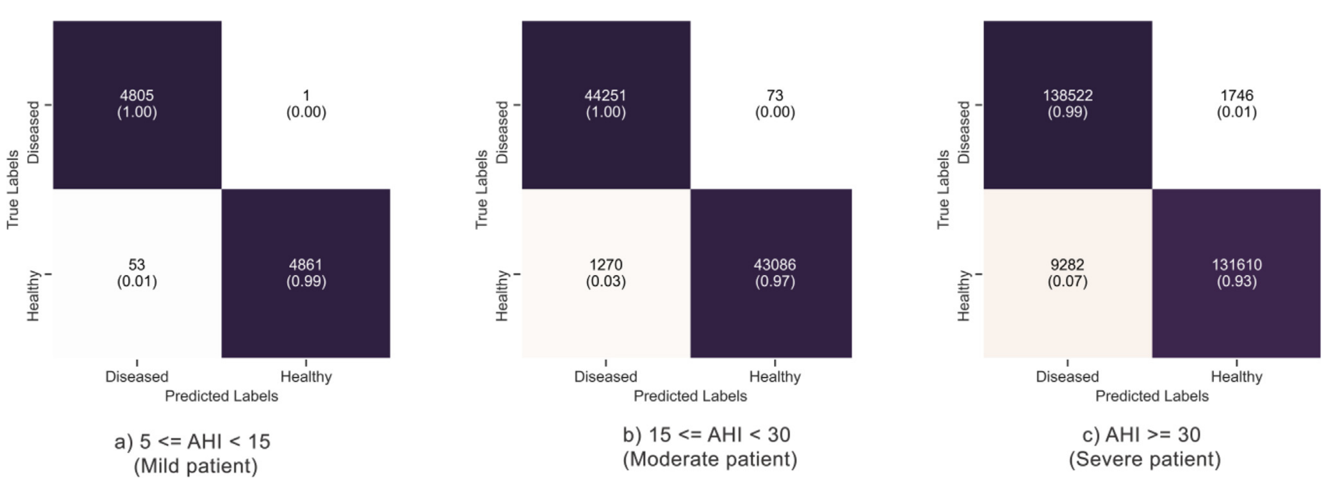

Fig. 3 – Confusion matrices for 3 disease levels in sleep apnea.

for all thresholds and is quite high as can be seen in the graph. This indicates that the proposed model shows high success in both classes.

In order to present proposed model as a decision support system by the doctor, it should be run in a client–server architecture. Because each patient's sleep record is approximately 400 MB, it is not possible to process it on mobile devices. For this reason, the performance of the classifier and the short training and testing times are also very important. The training and test times is shown in Fig. 6. Accordingly, when compared according to training time, LDA was the fastest, while Gradient Boosting was the longest classifier with training time. Test times, on the other hand, were the fastest in LDA. On the other hand, values close to each other were obtained on average in other classifiers. KNN, on the other hand, has much longer training and testing times than all other classifiers. For this reason, it has been removed from the chart.

Throughout the study, the tests were repeated with eight different ML algorithms, and the results is given. Accordingly, when selecting the best classifier for the detection of sleep apnea, it is expected to have both accuracy values and high AUC values, as well as low training and test times whenever possible. When these criteria are applied, the Tree based classifiers stand out. As a result, the decision support model showed high success in all 3 levels of sleep apnea with its wide feature vector and structure. In Table 3, training and testing times per sample and AUC Score values are given for each classifier. Accordingly, the analysis time per application is 0.22 ms for the Extra Tree classifier, where the highest performance is achieved. Besides, the under- sampling-time per sample is 2.32x e^{-6} ms as briefly shown in Table 3.

4. Discussion

The main objective of our research was to utilize machine-learning algorithms to improve the identification and categorization of sleep apnea across different degrees of severity. The objective of our study was to develop a resilient model that could produce precise outcomes regardless of the severity of the ailment. The outcomes suggest that our endeavor has been fruitful, as our model has exhibited superior performance in comparison to various prior investigations conducted in the same domain. In order to relate this work's

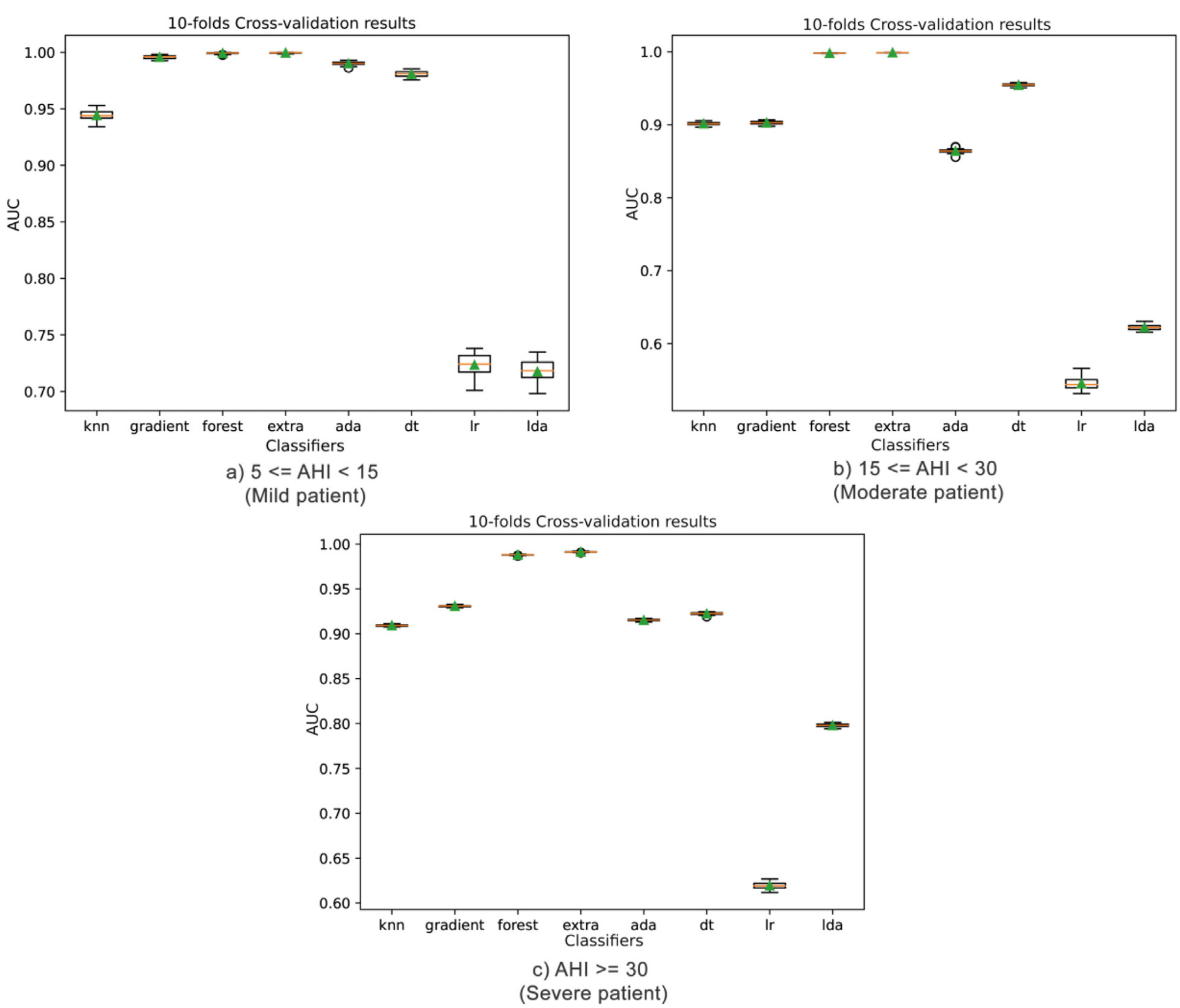

Fig. 4 – 10-folds Cross-validation results for 3 disease levels.

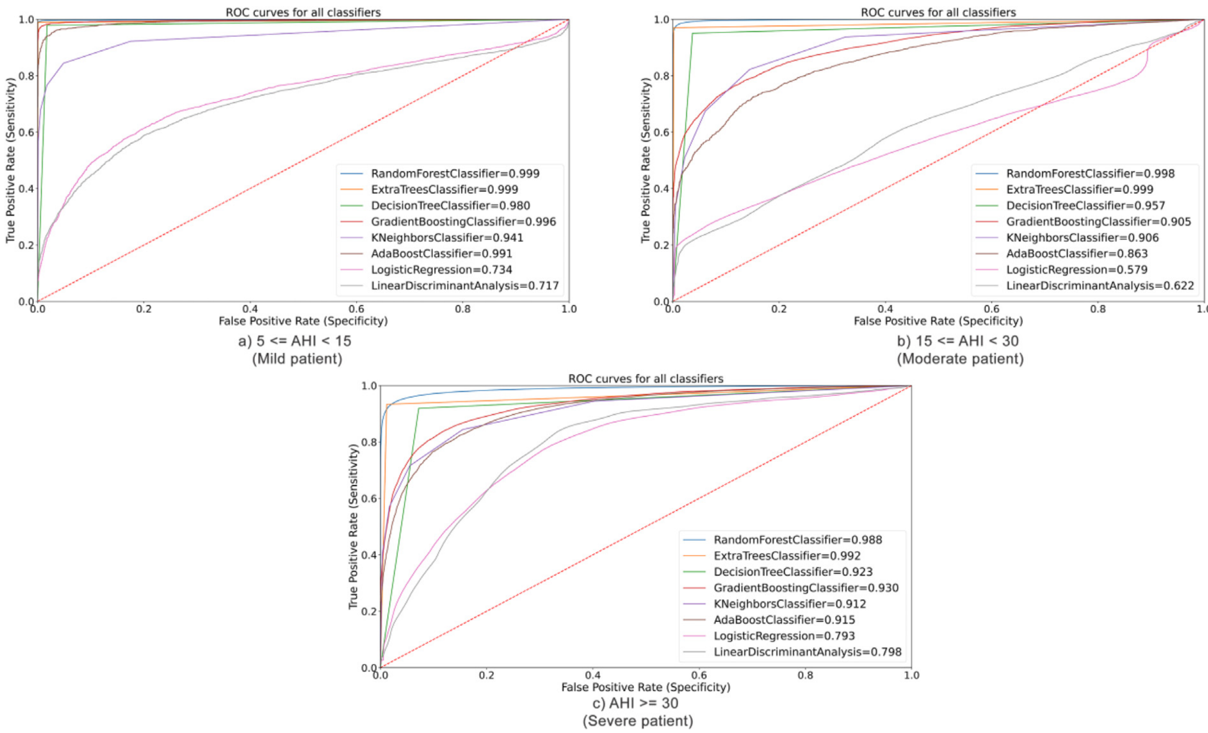

Fig. 5 – ROC curve diagrams for 3 disease levels in sleep apnea.

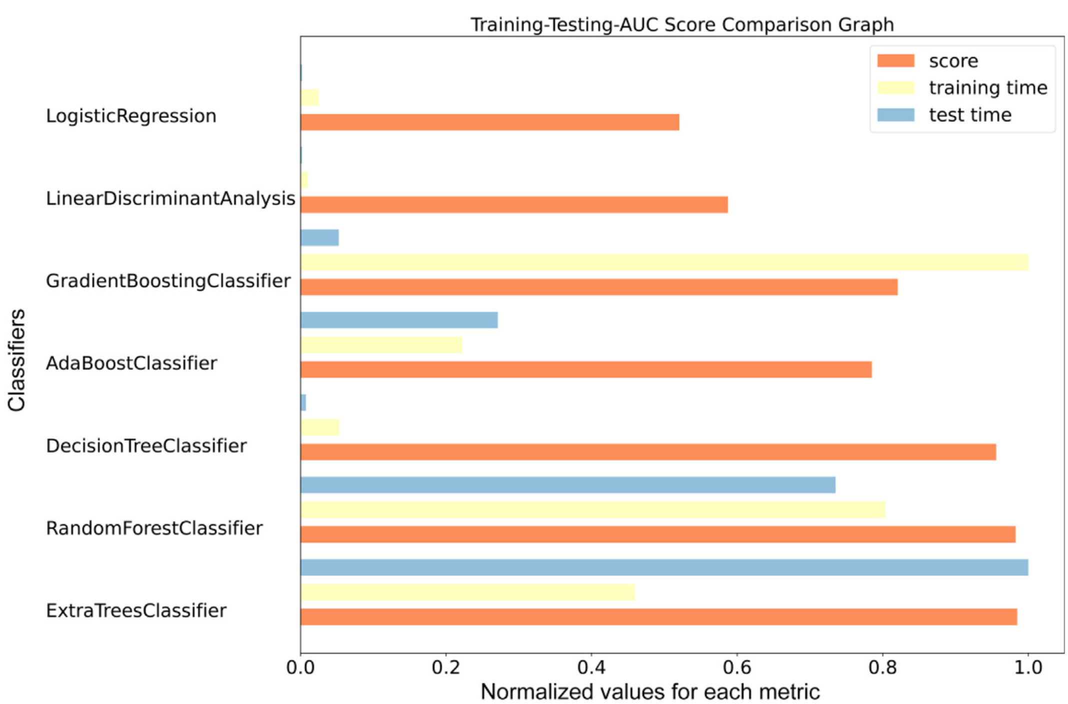

Fig. 6 – Training-Testing diagram on average.

results to individually cited results of other authors, a comparison table in Table 4 is given as below.

If this work's strong sides to be compared with similar works, all PSG inputs have been used for AHI calculation as clinician does. This work aimed to develop a decision support system for sleep clinicians who uses all PSG inputs. However, when we look at the other works, they generally employ single or small number of inputs for apnea detection. The other significant methodological strengths of our research was the use of k-fold cross-validation (k = 10). This approach enhanced the reliability of our findings by mitigating potential bias and variance. The box plots illustrated in Fig. 4 indicate that our model exhibits resilience to sample selection bias, as evidenced by the low heights of the plots. This finding

Table 3 – Training and Testing Time and AUC Score for each classifier.

Classifier	TrainingTime/Sample (milliseconds)	TestingTime/Sample (milliseconds)	Score(AUC)(Average)
Extra Tree (ET)	0.222720011	0.022929734	98.47
Random Forest (RF)	0.389264738	0.016858053	98.25
Decision Tree (DT)	0.025885713	0.000170531	95.58
Ada Boosting (AB)	0.107562814	0.006213415	78.50
Gradient Boosting (GB)	0.484505827	0.001203832	82.05
Linear Discriminant Analysis (LDA)	0.004964817	4.53855E-05	58.73
Logistic Regression (LR)	0.012074033	4.58401E-05	52.45

underscores the model's ability to maintain high levels of performance across diverse patient samples, a critical attribute for practical applications that involve patients with varying characteristics.

In addition, the outcomes of our study suggest that the model we developed, which employs the extra-tree and random forest algorithms, exhibits outstanding performance across various levels of disease severity. The clinical significance of our model is highlighted by the consistent values observed across all levels of severity, which is particularly relevant in real-world scenarios where patient conditions can vary greatly. The elevated median value serves to bolster our confidence in the model's comprehensive efficacy.

In contrast to numerous other investigations, our study placed significant emphasis on the pragmatic application of our model within an authentic clinical environment. The model was developed to function within a client–server architecture, with a focus on optimizing model performance and minimizing training and testing durations due to the substantial data size associated with each patient's sleep record. Tree-based classifiers provide an attractive trade-off between elevated precision, high AUC metrics, and reduced training and testing durations, rendering them especially appropriate for the identification of sleep apnea. Subsequent research endeavors in this particular domain ought to concentrate on the aforementioned classifiers.

Our study exhibits a higher level of accuracy in the binary classification of sleep apnea as compared to previous research (as shown in Table 4). Better performance values were obtained than all of the current studies in the literature. This signifies noteworthy progress in this domain, with an accuracy rate of 97.27%. The present study employed a diverse and distinctive dataset comprising 113 patients to evaluate the efficacy of our model. This approach has contributed to the increased generalizability of our findings, which is a notable departure from the methodology adopted in several prior investigations.

The methodology employed in our study for detecting sleep apnea involved the use of a binary classifier on a unique dataset comprising 113 patients. This dataset was characterized by 19 static and two dynamic features. The distinctive dataset not only presents originality to the discipline but also expands the groundwork for our theoretical structure. Notwithstanding the high precision attained, it is imperative to recognize the constraints of the current investigation. One of the prevalent obstacles in sleep apnea research that employs machine learning techniques is the susceptibility of the Apnea-Hypopnea Index (AHI) to various environmental factors. In order to address this issue, a diverse sample size of 113 patients was used to gather data.

In addition to the continuous endeavors, novel techniques for detecting sleep apnea are being investigated by scholars, including the utilization of remote sensors and radar systems [57–59]. These developments signify a positive advancement in this particular area and reduces the negative impact of environmental effects which are affecting the AHI index value calculations.

5. Conclusion

In conclusion, the current study has provided a precise decision-support model for detecting and classifying nocturnal sleep apnea in patients. By using 19 static sensor data points and 2 state data points as attributes in a novel way, this model significantly advances the detection of sleep apnea.

In the study, a special dataset comprised of data from 113 patients was used to train and assess machine learning models. In order to address issues with imbalanced data, undersampling techniques were put into practise. The Apnea-Hypopnea Index (AHI) was calculated automatically, and the study's success rate in diagnosing the patients' sleep apnea conditions was a noteworthy 97.27%.

Despite the high level of precision attained, it is important to acknowledge the limitations of our study, including the vulnerability to environmental factors that affect AHI. Data collected from a large and diverse patient population was used to address this restriction. To confirm the viability of our model, additional validation and testing on various datasets are necessary.

In the future, we hope to expand the sample size by adding more patients and improve the model through optimization processes to boost efficacy rates. Regarding the incorporation of additional static sensor data, the classification performance of the model was assessed. In addition, we want to classify sleep apnea into three distinct categories: mixed, obstructive, and central. With the help of this methodology, we will be able to determine the patient's pathological condition more precisely, in addition to calculating the AHI index.

Table 4 – Similar works related to sleep apnea detection.

Ref., Year	Dataset and size	Purpose of work	Engineering approach & machine learning classifier	Accuracy
[26],2018	Dr. Thomas Penzel of Philip-University, 35 subjects	Detection of sleep apnea by using deep learning	They applied deep learning method to extract features, and the decision fusion method for improving sensitivity	84.7%
[27],2018	Publicly available, 2100	Detection of apnea events and its severity	LSTM	80%
[28], 2019	Own dataset, 69 subjects	Detection of sleep apnea events by using CNN	a CNN network has been created and LSTM classifier applied on it.	98.5%
[29],2019	Own dataset, 313 subjects	Detection of sleep apnea events without using PSG inputs	8 different machine learning based algorithms applied	44.7%
[30],2019	PhsioNet dataset, 1200 subjects	Binary classification of patient and healthy subjects by using pulse transition time signals	CNN models have been used for feature extraction after that SVM and KNN applied for classification	92.78%
[31], 2020	PhysioNet dataset, 32 subjects	Detection of sleep apnea severity by using unsupervised model.	A new time dependent cost sensitive classification model by combining hidden Markov model and metacost algorithm is proposed.	86.2%
[32], 2020	MIT physioNet, 35 readings	Detection of sleep apnea events	A CNN based model proposed which emloys ECG channel for detecting sleep apnea events.	97.1%
[33], 2020	Own dataset, 96 subjects	Detection of sleep apnea events	A Rule based method is proposed for detection of apnea events	88.5%
[34], 2021	PhysioNet Apnea-ECG database, 187 subjects	Detection of sleep apnea events.	Ensemble model developed for detection of sleep apnea	85.58%
[35],2021	Wisconsin Sleep Cohort dataset, 1479 recording	Detection of sleep apnea without using PSG inputs	Hybrid model developed on machine learning	68.06%
[36], 2021	Own dataset, 28 subjects	Detection of sleep apnea by using thoracic accelerometer and an abdominal accelerometer	Machine learning model is proposed.	89%
[37],2022	Own dataset, 50 subjects	Sleep stage scoring by using machine learning	Machine learning techniques applied on private dataset,	95.25%
[38],2022	Apnea-ECG, 35 subjects	Sleep apnea detection via single lead ECG	1D-SEResGNet is developed	90.3%
[39], 2022	Childhood Adenotonsillectomy Trial, 1638	Airflow and oximetry signals employed for detection of sleep apnea in children	2d-CNN model has been developed	94.4%
[40],2023	Private dataset, 50 patients	All PSG inputs employed for high accuracy scoring.	Deep neural network	91.6%
[41],2023	Different datasets, 4014	Prediction of sleep apnea without using PSG inputs	Clustering technique has been developed and applied to different datasets	88%
[42],2023	Different datasets used, total 60 subjects	Sleep apnea detection by using single lead ECG	supervised learning model developed on CNN	86.3%
[43],2023	Private dataset, total 498 subjects	Machine learning	Machine learning identification of sleep apnea	86%
[41],2023	4014 questionary data	Machine learning based model for predicting sleep apnea	Machine learning model is developed according to the questionaries	91%
[44],2023	35 subjects, Physionet ECG dataset	Prediction of sleep apnea by using single channel ECG	Deep belief network based model developed for predicting sleep apnea	89.11%
[45],2023	121 Subjects, different university datasets	Detection of sleep apnea by using small number of PSG inputs	Multiple instance learning model (MIL) is proposed for detection of sleep apnea	78.2%
[46], 2023	59 subjects, University of California Medical Center	Detection of sleep apnea by using brain tensor imaging	Machine learning based model is developed according brain tensor images	77%
[47],2023	5804 + 50 subjects, 2 open datasets	a device is developed for sleep apnea detection	Sensor readings form ring evaluated by deep learning based model for detection of sleep apnea	85%
[48],2023	205 subjects from 2 open datasets	Multi channel spectrum has been used for automatic sleep stage classification	Multi channel inputs fed to genetic algorithm based model for sleep stage classification	96%
[49],2023	70 subjects from apnea ECG database	Detection of sleep apnea by using proposed device	ECG and oxygen sensor values read by device and hybrid model evaluates sleep apnea risk	95%
[50],2023	2357 questionaries used	Detection of sleep apnea by using questionaries	Machine learning model is developed to assess OSA screening	78% auroc
[51],2023	10 subjects, private database	Detection of sleep apnea by using proposed PPG device	Pulse rate and oxygen saturation level is used for machine learning assisted system to detect sleep apnea diagnose	85%
[56],2023	50 patient	Classification of sleep apnea as Apnea, Hypopnea and Normal	DNN-based two tier model	95.74%
This work	**Own dataset (113 patients)**	**Binary classifier for sleep apnea**	**19 static features and 2 events used with machine learning classifiers**	**97.27%**

As a result, the current study provides a strong foundation for further research in this particular area of inquiry. It opens up possibilities for improved and precise sleep apnea identification by pushing the boundaries of conventional techniques.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A Appendix

See Table A1.

REFERENCES

[1] Guilleminault C, Sullivan SS, Huang YS. Sleep-disordered breathing, orofacial growth, and prevention of obstructive sleep apnea. Sleep Med Clin 2019;14(1):13–20.

[2] Dheda K, Gumbo T, Maartens G, Dooley KE, Murray M, Furin J, et al. The Lancet Respiratory Medicine Commission: 2019 update: epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable tuberculosis. Lancet Respir Med 2019;7(9):820–6.

[3] Nassi TE. Algorithms for automated scoring of respiratory events in sleep. University of Twente; 2021 [Master's thesis].

[4] Giannoni A, Gentile F, Sciarrone P, Borrelli C, Pasero G, Mirizzi G, et al. Upright Cheyne-Stokes respiration in patients with heart failure. J Am Coll Cardiol 2020;75(23):2934–46.

Table A1 – Table of data amount per patient.

PID	Healthy	Disease (Hypopnea or Apnea)	PID	Healthy	Disease (Hypopnea or Apnea)	PID	Healthy	Disease (Hypopnea or Apnea)
1	4,747,800	16,200	39	4,762,600	9800	77	4,743,800	28,600
2	4,739,400	24,600	40	4,452,000	24,000	78	4,387,800	100,200
3	4,586,000	16,000	41	4,350,000	54,000	79	4,650,000	0
4	4,176,000	0	42	4,381,400	4600	80	4,457,400	24,600
5	4,530,000	30,000	43	4,538,000	4000	81	4,455,600	86,400
6	4,500,800	59,200	44	3,515,200	870,800	82	4,553,400	54,600
7	4,177,400	208,600	45	4,394,400	33,600	83	4,679,200	93,200
8	4,071,600	446,400	46	4,762,600	9800	84	4,750,800	21,600
9	4,536,000	0	47	4,625,000	19,000	85	4,440,000	6000
10	4,489,000	143,000	48	4,157,000	169,000	86	4,552,000	14,000
11	3,978,400	443,600	49	4,462,400	7600	87	4,746,400	26,000
12	4,630,600	7400	50	4,012,400	361,600	88	4,724,000	4000
13	4,340,000	28,000	51	4,322,000	22,000	89	4,763,400	9000
14	4,505,600	42,400	52	4,508,200	147,800	90	4,250,600	75,400
15	4,344,000	6000	53	4,456,000	56,000	91	4,466,400	306,000
16	4,277,200	246,800	54	4,478,000	22,000	92	4,636,000	26,000
17	4,526,000	148,000	55	4,456,000	8000	93	4,370,000	226,000
18	4,114,200	289,800	56	3,628,000	914,000	94	4,223,600	432,400
19	4,627,200	46,800	57	4,544,600	27,400	95	4,426,000	50,000
20	4,532,400	63,600	58	4,265,200	288,800	96	4,588,000	8000
21	4,018,800	649,200	59	4,344,200	287,800	97	4,702,000	14,000
22	4,348,000	338,000	60	4,530,000	0	98	3,935,400	468,600
23	4,458,000	24,000	61	3,672,001	0	99	4,374,000	18,000
24	4,239,400	182,600	62	4,118,000	346,000	100	3,787,200	700,800
25	4,287,800	56,200	63	4,474,600	43,400	101	4,532,000	10,000
26	4,196,200	345,800	64	4,318,000	26,000	102	4,276,800	97,200
27	4,694,800	15,200	65	4,476,000	0	103	4,490,000	28,000
28	4,762,400	10,000	66	4,446,000	0	104	4,477,200	10,800
29	4,594,800	177,600	67	4,450,200	322,200	105	4,738,000	34,400
30	3,992,800	333,200	68	4,528,000	14,000	106	3,845,000	709,000
31	4,488,000	60,000	69	4,374,000	24,000	107	4,698,400	74,000
32	4,772,400	0	70	4,455,200	62,800	108	3,747,000	3000
33	4,472,400	15,600	71	4,553,600	90,400	109	4,588,600	127,400
34	4,746,400	26,000	72	4,729,600	16,400	110	4,427,000	253,000
35	4,558,000	20,000	73	4,518,800	65,200	111	4,402,000	50,000
36	4,396,000	152,000	74	4,654,000	118,400	112	4,253,400	186,600
37	4,758,400	14,000	75	4,250,200	123,800	113	4,285,200	82,800
38	4,722,000	30,000	76	4,492,000	8000	**AVG**	**4,423,143**	**121,926**

[5] Malhotra A, Ayappa I, Ayas N, Collop N, Kirsch D, Mcardle N, et al. Metrics of sleep apnea severity: beyond the apnea-hypopnea index. Sleep 2021;44(7):zsab030.
[6] Hilmisson H, Lange N, Duntley SP. Sleep apnea detection: accuracy of using automated ECG analysis compared to manually scored polysomnography (apnea hypopnea index). Sleep Breath 2019;23(1):125–33.
[7] Matsumoto T, Chin K. Prevalence of sleep disturbances: sleep disordered breathing, short sleep duration, and non-restorative sleep. Respir Investig 2019;7(3):227–37.
[8] Ravesloot MJL, Vonk PE, Maurer JT, Oksenberg A, De Vries N. Standardized framework to report on the role of sleeping position in sleep apnea patients. Sleep Breath 2021;25:1717–28.
[9] Arslan RS, Ulutas H, Köksal AS, Bakır M, Çiftçi B. Tree-based machine learning techniques for automated human sleep stage classification. Traitement du Signal 2023;40(4):1385–400. https://doi.org/10.18280/ts.400408.
[10] Sateia MJ. International classification of sleep disorders. Chest 2014;146(5):1387–94.
[11] Bozkurt F, Uçar MK, Bozkurt MR, Bilgin C. Detection of abnormal respiratory events with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea. Irbm 2020;41(5):241–51.
[12] Mousavi S, Afghah F, Acharya UR. SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach. PLoS One 2019;14(5):e0216456.
[13] Fiorillo L, Puiatti A, Papandrea M, Ratti PL, Favaro P, Roth C, et al. Automated sleep scoring: a review of the latest approaches. Sleep Med Rev 2019;48 101204.
[14] Budhiraja R, Thomas R, Kim M, Redline S. The role of big data in the management of sleep-disordered breathing. Sleep Med Clin 2016;11(2):241–55.
[15] Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: a comprehensive updated review. Sensors 2021;21(11):3758.
[16] Stephen M. Machine learning: an algorithmic perspective. 2nd ed. Routledge & CRC Press. https://www.routledge.com/Machine-Learning-An-Algorithmic-Perspective-Second-Edition/Marsland/p/book/9781466583283 [visited on 12/04/2021].
[17] Xu S, Faust O, Silvia S, Chakraborty S, Barua PD, Loh HW, et al. A review of automated sleep disorder detection. Comput Biol Med 2022;106100.
[18] Mostafa SS, Mendonca F, Ravelo-Garcia AG, Juliá-Serdá GG, Morgado-Dias F. Multi-objective hyperparameter optimization of convolutional neural network for obstructive sleep apnea detection. IEEE Access 2020;8:129586–99.
[19] Salari N, Hosseinian-Far A, Mohammadi M, Ghasemi H, Khazaie H, Daneshkhah A. Ahmadi A Detection of sleep apnea using Machine learning algorithms based on ECG Signals: a comprehensive systematic review. Expert Syst Appl 2020;187 115950.
[20] Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM scoring manual updates for 2017 (version 2.4). J Clin Sleep Med 2017;13(5):665–6.
[21] Zhongxu Z, Fengxia W, Xuan Y, Li Z, Chang H, Jing X, Changzhi L, Hong H. Accurate contactless sleep apnea detection framework with signal processing and machine learning methods. Methods 2022:167–78.
[22] Cao K, Lv X. Multi-task feature fusion network for Obstructive Sleep Apnea detection using single-lead ECG signal. Measurement 2022;02 111787.
[23] Tsouti V, Kanaris AI, Tsoutis K, Chatzandroulis S. Development of an automated system for obstructive sleep apnea treatment based on machine learning and breath effort monitoring. Microelectron Eng 2020;231 111376.
[24] Satapathy SK, Loganathan D. A study of human sleep stage classification based on dual channels of EEG signal using machine learning techniques. SN Computer Sci 2021;2(3):1–16.
[25] Alpaydin E. Machine learning. MIT Press; 2021.
[26] Li K, Pan W, Li Y, Jiang Q, Liu G. A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal. Neurocomputing 2018;294:94–101.
[27] Van Steenkiste T, Groenendaal W, Deschrijver D, Dhaene T. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. IEEE J Biomed Health Inform 2018;23(6):2354–64.
[28] Erdenebayar U, Kim YJ, Park JU, Joo EY, Lee KJ. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. Comput Methods Programs Biomed 2019;180 105001.
[29] Mencar C, Gallo C, Mantero M, Tarsia P, Carpagnano GE, Foschino Barbaro MP, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. Health Informatics J 2020;26(1):298–317.
[30] Tuncer SA, Akılotu B, Toraman S. A deep learning-based decision support system for diagnosis of OSAS using PTT signals. Med Hypotheses 2019;127:15–22.
[31] Feng K, Qin H, Wu S, Pan W, Liu G. A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. IEEE Trans Instrum Meas 2020;70:1–12.
[32] Chang HY, Yeh CY, Lee CT, Lin CC. A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram. Sensors 2020;20(15):4157.
[33] ElMoaqet H, Kim J, Tilbury D, Ramachandran SK, Ryalat M, Chu CH. Gaussian mixture models for detecting sleep apnea events using single oronasal airflow record. Appl Sci 2020;10(21):7889.
[34] Mukherjee D, Dhar K, Schwenker F, Sarkar R. Ensemble of deep learning models for sleep apnea detection: an experimental study. Sensors 2021;21(16):5425.
[35] Ramesh J, Keeran N, Sagahyroon A, Aloul F. Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning no. 11. In: Healthcare. MDPI; 2021. p. 1450.
[36] Bricout A, Fontecave-Jallon J, Pépin JL, Guméry PY. Accelerometry-derived respiratory index estimating apnea-hypopnea index for sleep apnea screening. Comput Methods Programs Biomed 2021;207 106209.
[37] Arslan RS, Ulutaş H, Köksal AS, Bakır M, Çiftçi B. Automated sleep scoring system using multi-channel data and machine learning. Comput Biol Med 2022;146 105653.
[38] Yang Q, Zou L, Wei K, Liu G. Obstructive sleep apnea detection from single-lead electrocardiogram signals using one-dimensional squeeze-and-excitation residual group network. Comput Biol Med 2022;140 105124.
[39] Jiménez-García J, García M, Gutiérrez-Tobal GC, Kheirandish-Gozal L, Vaquerizo-Villar F, Álvarez D, et al. A 2D convolutional neural network to detect sleep apnea in children using airflow and oximetry. Comput Biol Med 2022;147 105784.
[40] Arslan RS, Ulutas H, Köksal AS, Bakir M, Çiftçi B. Sensitive deep learning application on sleep stage scoring by using all PSG data. Neural Comput & Applic 2023;35(10):7495–508.
[41] Han H, Oh J. Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity. Sci Rep 2023;13(1):6379.
[42] Hu S, Liu J, Yang C, Wang A, Li K, Liu W. Semi-supervised learning for low-cost personalized obstructive

sleep apnea detection using unsupervised deep learning and single-lead electrocardiogram. IEEE J Biomed Health Inform 2023.
[43] Maniaci A, Riela PM, Iannella G, Lechien JR, La Mantia I, De Vincentiis M, et al. Machine learning identification of obstructive sleep apnea severity through the patient clinical features: a retrospective study. Life 2023;13(3):702.
[44] Tyagi PK, Agrawal D. Automatic detection of sleep apnea from single-lead ECG signal using enhanced-deep belief network model. Biomed Signal Process Control 2023;80 104401.
[45] Cheng L, Luo S, Li B, Liu R, Zhang Y, Zhang H. Multiple-instance learning for EEG based OSA event detection. Biomed Signal Process Control 2023;80 104358.
[46] Pang B, Doshi S, Roy B, Lai M, Ehlert L, Aysola RS, et al. Machine learning approach for obstructive sleep apnea screening using brain diffusion tensor imaging. J Sleep Res 2023;32(1):e13729.
[47] Strumpf Z, Gu W, Tsai CW, Chen PL, Yeh E, Leung L, et al. Belun Ring (Belun Sleep System BLS-100): deep learning-facilitated wearable enables obstructive sleep apnea detection, apnea severity categorization, and sleep stage classification in patients suspected of obstructive sleep apnea. Sleep. Health 2023.
[48] Abdulla S, Diykh M, Siuly S, Ali M. An intelligent model involving multi-channels spectrum patterns based features for automatic sleep stage classification. Int J Med Inf 2023;171 105001.
[49] Hemrajani P, Dhaka VS, Rani G, Shukla P, Bavirisetti DP. Efficient deep learning based hybrid model to detect obstructive sleep apnea. Sensors 2023;23(10):4692.
[50] Huo J, Quan SF, Roveda J, Li A. BASH-GN: a new machine learning–derived questionnaire for screening obstructive sleep apnea. Sleep Breath 2023;27(2):449–57.
[51] Wang S, Xuan W, Chen D, Gu Y, Liu F, Chen J, et al. Machine learning assisted wearable wireless device for sleep apnea syndrome diagnosis. Biosensors 2023;13(4):483.
[52] Mlynczak M, Valdez TA, Kukwa W. Comparing sleep studies in terms of the apnea-hypopnea index using the dedicated Shiny web application. Biomed Signal Process Control 2021;68:1–7.
[53] Ruehland WR, Rochford PD, O'Donoghue FJ, Pierce RJ, Singh P, Thornton AT. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. Sleep 2009;32(2):150–7.
[54] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on Ensemble Learning. Front Comp Sci 2019;14(2):241–58.
[55] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.; 2022.
[56] Arslan RS. Sleep disorder and apnea events detection framework with high performance using two-tier learning model design. PeerJ Computer Science 2023;9:e1554. https://doi.org/10.7717/peerj-cs.1554.
[57] Zhou Y, Shu D, Xu H, Qiu Y, Zhou P, Ruan W, et al. Validation of novel automatic ultra-wideband radar for sleep apnea detection. J Thorac Dis 2020;12(4):1286.
[58] Lai DKH, Yu ZH, Leung TYN, Lim HJ, Tam AYC, So BPH, et al. Vision transformers (ViT) for blanket-penetrating sleep posture recognition using a triple ultra-wideband (UWB) radar system. Sensors 2023;23(5):2475.
[59] Lee WH, Kim SH, Na JY, Lim YH, Cho SH, Cho SH, et al. Non-contact sleep/wake monitoring using impulse-radio ultra wideband radar in neonates. Front Pediatr 2021;9 782623.