A review on automated sleep study

Mehran Yazdi[*,1,2] , Mahdi Samaee[2] and Daniel Massicotte[1]

[1]Laboratory of Signal and System Integration, Department of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, Trois-Rivières, Canada
[2]Signal and Image Processing Laboratory, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

Mehran.Yazdi@uqtr.ca, mahdisamaee@gmail.com and Daniel.Massicotte@uqtr.ca

*Corresponding author

Abstract

In recent years, research on automated sleep analysis has witnessed significant growth, reflecting advancements in understanding sleep patterns and their impact on overall health. This review synthesizes findings from an exhaustive analysis of 87 papers, systematically retrieved from prominent databases such as Google Scholar, PubMed, IEEE Xplore and ScienceDirect. The selection criteria prioritized studies focusing on methods employed, signal modalities utilized, and machine learning algorithms applied in automated sleep analysis. The overarching goal was to critically evaluate the strengths and weaknesses of the proposed methods, shedding light on the current landscape and future directions in sleep research. An in-depth exploration of the reviewed literature revealed a diverse range of methodologies and machine learning approaches employed in automated sleep studies. Notably, K-Nearest Neighbors (KNN), Ensemble Learning Methods, and Support Vector Machine (SVM) emerged as versatile and potent classifiers, exhibiting high accuracies in various applications. However, challenges such as performance variability and computational demands were observed, necessitating judicious classifier selection based on dataset intricacies. Additionally, the integration of traditional feature extraction methods with deep structures and the combination of different deep neural networks were identified as promising strategies to enhance diagnostic accuracy in sleep-related studies. The reviewed literature emphasized the need for adaptive classifiers, cross-modality integration, and collaborative efforts to drive the field toward more accurate, robust, and accessible sleep-related diagnostic solutions. This comprehensive review serves as a solid foundation for researchers and practitioners, providing an organized synthesis of the current state of knowledge in automated sleep analysis. By highlighting the strengths and challenges of various methodologies, this review aims to guide future research toward more effective and nuanced approaches to sleep diagnostics.

Keywords: sleep study, sleep stage, sleep disorder, deep learning, sleep scoring system, feature extraction, machine learning

1. Introduction

Sleep is a natural state of rest for the body during which the brain undergoes a series of complex processes that help refresh body functions. Adequate sleep is essential for maintaining physical, mental, and emotional health. During sleep, the body repairs damaged tissues, consolidates memories, and regulates hormone levels [1, 2]. Lack of sleep or sleep disorders can lead to a variety of health problems, including obesity, diabetes, cardiovascular disease, depression, anxiety, and impaired cognitive function [3, 4]. Sleep apnea, insomnia, and restless leg syndrome are examples of prevalent sleep disorders that can exert a substantial influence on an individual's overall well-being and functional capacity. Sleep studies are frequently deemed indispensable for the precise identification and efficacious management of sleep disorders. Sleep studies involve monitoring an individual's brain waves, heart rate, breathing, and other physiological functions during sleep in order to identify any abnormalities or underlying medical conditions. Precisely identifying and effectively managing sleep disorders can lead to a notable enhancement in one's holistic health and well-being, as well as augment their life satisfaction and functional outcomes.

Within this review article, we present a comprehensive overview of diverse facets pertaining to automated sleep studies. Although some related review publications exist, their focus has been mostly on a narrow aspect of sleep study such as sleep scoring [5], sleep stage classification [6] or on a category of classification methods[7]. Here, we focus on introducing all aspects of sleep analysis such as the modalities to capturing sleep data, preprocessing techniques, classification methods, evaluation metrics and available benchmark datasets. We also bring a statistical insight analysis of these different aspects. We finally discuss the remained challenges, possible solutions and future trends. In the following section, the sleep study is briefly explained.

1.1. Sleep study

In the past, sleep studies were typically analyzed manually by experts, which was a time-consuming and subjective process. However, with advances in technology and machine learning algorithms, automated computer-based analysis has become a more common approach in sleep studies. Conventional sleep studies encompass the monitoring of a plethora of physiological signals during the sleep period, which may encompass electroencephalogram, electromyogram, electrooculogram, respiratory signals, and other metrics indicative of sleep behavior and sleep architecture. These signals are then manually analyzed by experts to identify sleep stages, breathing patterns, and other abnormalities. However, this procedure involves a significant amount of time and can be quite time-consuming, and the results are subjective and can vary depending on the expertise of the analyst. To address these issues, researchers have developed automated computer-based analysis methods that use machine learning algorithms to analyze sleep data. These algorithms can quickly and accurately identify sleep stages, breathing patterns and abnormalities, making sleep disorder diagnosis faster and more objective. Polysomnography (PSG) is widely recognized as one of the predominant methods for the analysis of sleep studies [8]. PSG, a comprehensive test used to diagnose various sleep disorders, involves the simultaneous monitoring of multiple physiological parameters while a patient sleeps. This includes the recording of brain waves (electroencephalography), eye movements (electrooculography), muscle activity (electromyography), heart rate, breathing patterns, and blood oxygen levels [9]. The automated PSG

scoring process utilizes machine learning algorithms to analyze these recorded signals, aiming to identify various parameters such as sleep stages, breathing patterns, and abnormalities. Studies have demonstrated that automated PSG scoring can achieve accuracy comparable to manual scoring, while offering the additional benefits of reduced time and cost in conducting sleep studies [10]. Recent advancements in sleep research have spurred exploration beyond traditional PSG modalities. Researchers are increasingly investigating alternative signal modalities, such as actigraphy data, photoplethysmogram (PPG), radar sensors data, and audio recordings [11, 12]. These emerging modalities offer the potential to enhance our understanding of sleep patterns and disorders, providing valuable insights that may complement or extend the capabilities of traditional PSG analysis.

Numerous methods have been proposed for automating sleep studies, ranging from traditional machine learning algorithms to state-of-the-art deep neural networks. This paper will provide a comprehensive review of these algorithms and will also provide an overview of related research papers.

1.2. Sleep disorders

Some common sleep disorders are as follows:

- **Insomnia:** Insomnia is a prevalent sleep disorder characterized by impaired sleep onset, maintenance, or both, leading to manifestations such as trouble initiating sleep, frequent nocturnal awakenings, early morning awakenings, and diurnal lethargy or irritability [13].
- **Sleep Apnea:** Sleep apnea is a sleep disorder that is typified by recurrent interruptions in breathing during sleep. Symptoms of sleep apnea include loud snoring, abrupt awakenings due to gasping or choking sensations, episodes of breathing cessation, and excessive daytime sleepiness [13].
- **Restless Legs Syndrome (RLS):** RLS is a type of sleep disturbance that causes an uncontrollable impulse to move the legs, especially when at rest or inactive. Indications of RLS comprise of discomfort or pain in the legs, a strong craving to move them, and trouble initiating or maintaining sleep [13].
- **Narcolepsy:** Narcolepsy is a sleep disorder that is distinguished by an excessive inclination to sleep during the daytime, as well as sudden episodes of muscle weakness, and hallucinations during sleep or wakefulness [13].
- **Parasomnias:** Parasomnias are a category of sleep disorders that entail anomalous behaviors during sleep, such as somnambulism, night terrors, and sleep-related eating disorder. Symptoms of parasomnias include abnormal behaviors during sleep, confusion upon waking, and a lack of awareness of the behavior [13].

1.3. Signal modalities for the sleep study

Common signal modalities used in the field of sleep studies are as follows:

- **Electroencephalogram (EEG):** Scalp electroencephalography is a non-intrusive technique that gauges the electrical activity of the brain by means of electrodes that are strategically positioned on the scalp. EEG is a widely used tool in sleep studies to monitor

the neural oscillations and electrical patterns of the brain during sleep and identify the different stages of sleep. Spectral analysis of EEG signals aids in the identification of specific frequency bands, including alpha, beta, and delta waves associated with different stages of sleep [14].

- **Electromyogram (EMG):** Electromyography is a non-intrusive procedure that gauges the electrical impulses of muscles, employing either surface electrodes or needle electrodes. EMG is commonly used in sleep studies to monitor muscle activity during sleep and to identify periods of muscle atonia, which are characteristic of rapid eye movement (REM) sleep. EMG signals are analyzed to identify muscle activity and to distinguish between different stages of sleep or diagnosis of sleep disorder [14].
- **Electrooculogram (EOG):** Electrooculography is a non-intrusive technique that gauges the electrical impulses of the muscles encompassing the eyes using surface electrodes. EOG is commonly used in sleep studies to monitor eye movements during sleep and to identify periods of REM sleep [14].
- **Respiratory Signals:** Respiratory signals, encompassing indicators such as nasal airflow, movements of the chest and abdomen, and oxygen saturation, are commonly used in sleep studies to monitor breathing during sleep and to diagnose sleep-related breathing disorders, such as sleep apnea. Respiratory signals are analyzed to detect abnormalities in breathing patterns and to identify periods of apnea or hypopnea [14].
- **Actigraphy data:** Actigraphy is a non-intrusive method that measures physical activity using an accelerometer attached to the wrist or ankle. Actigraphy is commonly used in sleep studies to monitor sleep-wake patterns and to diagnose sleep disorders, such as circadian rhythm disorders [14].
- **Photoplethysmogram (PPG):** Photoplethysmography is a non-intrusive method of monitoring blood volume changes, which operates by shining light onto the skin and detecting the resulting variations using a photodetector. PPG is commonly used in sleep studies to monitor heart rate and to detect periods of apnea or hypopnea [14].
- **Radar Sensors:** Radar sensors measure movements and breathing patterns using radio waves. Radar sensors are widely used in sleep studies to track breathing patterns and aid in the diagnosis of sleep-disordered conditions, including but not limited to sleep apnea [14].
- **Audio:** Audio sensors record sound using microphones. Audio signals are commonly used in sleep studies to monitor snoring and to diagnose sleep-related breathing disorders, such as sleep apnea [14].

The subsequent sections of the paper are organized as follows: Section 2 discusses various preprocessing, feature extraction and feature selection methods, Section 3 explains commonly used classification methods and different evaluation criteria, Section 4 describes different benchmark datasets, Section 5 presents a discussion, Section 6 explains challenges and future trends and a conclusion is presented in Section 7.

2. Methods

2.1. Literature Selection

The selection of literature for this study employed a meticulous process aimed at capturing the most relevant and comprehensive insights into the usage of signal processing and machine learning techniques in sleep-related studies. Initial exploration encompassed online databases, including Google Scholar, PubMed, IEEE Xplore and ScienceDirect. The inclusion criteria prioritized papers that demonstrated a clear and meaningful application of signal processing and machine learning methodologies in the context of sleep analysis. Additionally, papers that exhibited proximity in their subject matter underwent a meticulous comparison, with preference given to the more comprehensive and exhaustive studies. This approach ensured that the final selection encompassed a diverse array of methodologies, providing a nuanced understanding of the state-of-the-art in automated sleep analysis. To prioritize accessibility and ensure linguistic consistency, the literature search was confined to papers available in English. After the initial screening, a total of 87 papers were ultimately selected for detailed inclusion, providing a solid foundation for the subsequent comprehensive review. The thorough and selective literature selection process guarantees that the synthesized insights encapsulate the forefront of research in sleep-related studies, emphasizing the intricate integration of signal processing and machine learning techniques.

2.2. Evaluation criteria

In order to evaluate and compare various algorithms, quantitative, fair, and consistent metrics should be used. Commonly used metrics are cited in Table 1.

Table 1. Criteria for evaluating a method's performance

Metric	Description
Accuracy	Accuracy measures the percentage of correct predictions made by the model. $Accuracy = \frac{True\ Positive\ +\ True\ Negative}{True\ Positive\ +\ True\ Negative\ +\ False\ Positive\ +\ False\ Negative}$
Precision	Precision measures the proportion of true positives among all instances predicted as positive. $Precision = {True\ Positive}/{Predicted\ Positive}$
Recall (Sensitivity)	Recall measures the proportion of true positives among all actual instances of the positive class. $Recall = {True\ Positive}/{Real\ Positive}$
Specificity	Specificity is the proportion of true negatives that the model correctly predicts. $Specificity = {True\ Negative}/{True\ Negative + False\ Positive}$
F1-Score	F1-Score is the harmonic mean of precision and recall and is often used as a balanced measure between the two. $F1 - score = {2 \times Precision \times Recall}/{Precision + Recall}$
Kappa coefficient	Kappa coefficient measures the agreement between the predicted and actual classes, taking into account the possibility of agreement by chance. $kappa = \frac{Observed\ agreement - Expected\ agreement}{1 - Expected\ agreement}$
AUC	The performance curve of a classification model is generated by plotting its true positive rate against the false positive rate at various thresholds. AUC represents the "Area Under that Curve".

Note: In the context of sleep studies, whether addressing a two-class problem like sleep apnea detection or a multi-class problem like sleep stage classification, essential metrics include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In a two-class scenario, such as sleep apnea detection, TP represents instances where the model correctly identifies patients with sleep apnea, TN signifies accurate identifications of patients without sleep

apnea, FP denotes instances of the model incorrectly indicating the presence of sleep apnea, and FN corresponds to the model failing to detect sleep apnea when it is present. In sleep stage classification, a multi-class problem, these metrics are calculated for each class separately. TP denotes instances where the model correctly identifies a segment of data as belonging to a specific sleep stage, FN occurs when the model fails to classify a segment into the correct sleep stage, and FP signifies misclassification, where the model incorrectly assigns a sleep stage label to a segment that does not belong to that stage. However, since multi-class problems lack a concept of "true negatives" for each specific class, TN is not directly applicable in this context. Final metrics, such as accuracy and Macro-F1 score (MF1), are calculated by averaging the individual class metrics. To be precise, With the True Positives (TPi), False Positives (FPi) and False Negatives (FNi) for each class i, we can define the overall accuracy (ACC) and macro-F1 ($MF1$) as follows.

$$ACC = \frac{\sum_{i=1}^{K} TP_i}{M}, \qquad MF1 = \frac{1}{K}\sum_{i=1}^{K} \frac{2*Precision_i*Recall_i}{Precision_i+Recall_i}, \quad Precision_i = \frac{TP_i}{TP_i+FP_i}, \qquad Recall_i = \frac{TP_i}{TP_i+FN_i}$$

Here, *M* represents the overall sample count, while *K* denotes the quantity of classes or sleep stages.

2.3. Benchmark datasets

Common datasets used for sleep studies are as follows:

- **Physionet Apnea-ECG database [15]:** This is a collection of ECG recordings used to study sleep apnea and consists of 70 records.
- **UCD[15]:** This dataset contains 25 sets of overnight polysomnograms collected from participants with symptoms of sleep-disordered breathing.
- **EDF [15]:** This dataset consists of polysomnograms and corresponding annotations exhibiting various sleep stages. The expanded version includes a total of 197 sets of recordings.
- **CCSHS [16, 17]:** This is a compilation of polysomnographic recordings taken overnight from children between the ages of 8 and 11 years.
- **ISRUC [18]:** This is a dataset consisting of polysomnograms obtained from adult subjects, categorized into three subgroups, with a total of 118 recordings.
- **MASS [19]:** This is a compilation of polysomnograms gathered from a group of 200 individuals.
- **SHHS [20]:** This dataset comprises polysomnograms collected from approximately 6000 participants, with the primary objective of examining the correlation between sleep-disordered breathing and cardiovascular disease.
- **MrOS [16, 21]:** The dataset comprises polysomnograms obtained from 2,911 males aged 65 years or above, primarily intended to research the correlation between sleep disorders and vascular disease, mortality, falls, and fractures.
- **MESA [16, 22]:** This dataset comprises polysomnograms as well as data from wrist-worn actigraphy, collected with the primary objective of examining the variations in sleep

patterns and sleep disorders among different genders and ethnic groups and their correlation with subclinical markers of atherosclerosis

- **STAGES [16]:** This dataset includes various recordings from 1,500 adolescent and adult participants, such as polysomnography, actigraphy, and facial scans.
- **WSC [16, 23]:** This dataset consists of raw physiological data obtained from overnight polysomnography, with the primary focus being the investigation of sleep disorders, specifically sleep apnea.
- **CFS [16, 24]:** This is a collection of polysomnograms extracted from 2,284 individuals belonging to 361 families. The main focus of the related study is sleep apnea.
- **HomePAP [16, 25]:** This dataset contains polysomnograms from 373 participants who were susceptible to sleep apnea.

3. Results

In the preliminary phase of this extensive review, we initially identified a total of 1250 papers. Subsequently, through the elimination of duplicates and a meticulous examination of titles and abstracts, the pool was refined to 210 potential articles. Following a detailed examination of methods, 87 papers emerged as the focal point of our in-depth analysis. To enhance the organization and systematic presentation of the literature, we adopted a structured approach based on the signal processing techniques and machine learning methods employed in the selected papers. This categorization aimed to elucidate the diverse methodologies prevalent in the field of automated sleep analysis, providing a framework for a more coherent and detailed exploration. Our focused subgroup analysis not only facilitated a nuanced examination of the literature but also allowed for the arrangement of papers based on their signal processing techniques and machine learning methods. This categorization ensures a comprehensive coverage of the varied methodologies employed in automated sleep studies, guiding the subsequent discussion towards a more in-depth understanding of the landscape.

3.1. Signal processing

Before injecting to a classifier, the quality of the data should be first improved, most of the relevant information should be then extracted and finally an optimal set of features that allows an acceptable performance should be selected.

3.1.1. Filtering

In order to accurately extract features from sleep data, they should be undergone some subsequent preparations including noise and artefact removal. Hereby, filtering techniques can be effectively applied to remove undesired spectrum bands from sleep signals. Some of the common filters are as follows:

- **Notch filter:** Basically, we can use this filter to remove interference frequencies of power lines (50 or 60 Hz).

- **Linear Adaptive filters:** The most used filters in this category are Least Mean Square (LMS), Recursive Least Square (SRL) and Kalman filter. These filters are efficient in noise cancellation where the signal characteristics change rapidly.
- **Non-linear Adaptive filters:** The objective of using Non-linear adaptive filters is to overcome the limitations of linear filters in better modeling of coherence of signals. Volterra LMS, Fuzzy and Neural networks are among these filters.

Details of other filters can be found in [26].

3.1.2. Feature extraction

Feature extraction plays a critical role in both signal processing and machine learning. It involves the use of signal processing to extract relevant features from the raw data such as underlying repetitive patterns and information. This step is crucial since accurately extracting relevant features from these signals enables the development of robust and efficient machine learning models that can reliably classify and predict outcomes based on the signal data. Within this section, an outline of feature extraction methodologies is presented, encompassing techniques and algorithms employed in time, frequency and time-frequency domain analysis.

3.1.2.1. Temporal features

Temporal features pertain to the attributes of a signal that are related to the time domain, such as its amplitude, periodicity and duration. These features are particularly important in applications where the temporal dynamics of the signal are critical, such as in electroencephalogram and electrocardiogram.

There are several methods for extracting temporal features from a signal, including:

- **Statistical Features:** The statistical features of a signal encompass various measures, such as its median, mean, variance, standard deviation, and skewness. These features provide information on the signal's central tendency, variability, and symmetry. Statistical features are easy to compute and interpret, making them widely used in machine learning applications.
- **Autocorrelation:** Autocorrelation is a metric that assesses the degree of similarity between a signal and its time-delayed version. Autocorrelation features capture the periodicity and regularity of the signal.
- **Waveform Features:** Waveform features include zero-crossing rate, slope, and energy of the signal. The zero-crossing rate quantifies the average frequency at which a signal alternates between positive and negative values. Slope measures the rate of signal's amplitude change, and power measures the signal's global strength.
- Temporal features are advantageous because they are easy to interpret, however, they have limitations. They may not capture the frequency content of the signal accurately, and they may be sensitive to noise and artifacts, which can affect their reliability and accuracy.

In [27] the technique for sleep staging incorporated the application of features related to the time domain and the similarity of structural graphs. The proposed approach entails segmenting the

electroencephalogram signal into smaller fragments, followed by the extraction of diverse time domain features. The study reports a 95.93% average accuracy of sleep stage classification.

3.1.2.2. Spectral features

Spectral features pertain to the attributes of a signal that are related to the frequency domain, such as its power spectral density (PSD) and spectrogram. These attributes hold significant value in applications where the signal's frequency content plays a crucial role.

There are several methods for extracting spectral features from a signal, including:

- **Fourier Transform:** Fourier transform is a non-parametric mathematical operation that resolves a signal into its individual frequency constituents [28]. An efficient computation method for the Fourier Transform is the Fast Fourier Transform (FFT) algorithm, which provides a computationally expedient solution.
- **Periodogram:** The periodogram is a spectral estimation method that attenuates any added noise to the signal and computes its power spectral density (PSD). The periodogram provides a spectral representation of the signal, capturing the frequency content of the signal [29]. The periodogram is easy to compute and has a simple interpretation.
- **Welch's Method:** Welch's method is a spectral estimation method that computes the PSD of a signal using a modified periodogram approach. Welch's method divides the signal into overlapping segments, windows each segment with a Hann window function, and then computes the periodogram of the signal [30]. Welch's method provides a spectral representation of the signal, capturing the frequency content of the signal with improved accuracy and reduced variance compared to the periodogram.

Spectral features are advantageous as they allow for the precise capture of a signal's frequency content, thereby enhancing the efficacy of signal processing and machine learning models. In addition, spectral features can be used to separate different sources of a mixed signal. Furthermore, providing valuable insights into the underlying characteristics of the signal. Furthermore, spectral features can provide useful information for understanding the physical properties of the signal, such as the dominant frequencies or the presence of harmonics. However, spectral features also have limitations. They may not capture the temporal dynamics of the signal, which can be important in some applications. Additionally, spectral features are sensitive to noise and artifacts in the signal, which can affect their accuracy and reliability.

3.1.2.3. Time-frequency features

Time-frequency features are generated by a type of transformation of the signal that embodies both the temporal dynamics and frequency content of the signal, providing a comprehensive representation of the signal's behavior. These features are particularly important in applications where both time and frequency information are critical, such as biomedical signal processing.

There are several methods for extracting time-frequency features from a signal, including:

- **Short-Time Fourier Transform (STFT):** STFT is a mathematical technique that computes the Fourier Transform of a signal in short time windows. STFT provides a time-

frequency representation of the signal, capturing both the temporal dynamics and the frequency content of the signal [31].

- **Continuous Wavelet Transform (CWT):** CWT is a mathematical technique that computes the Wavelet Transform of a signal at all possible scales. CWT provides a time-frequency representation of the signal, capturing both the temporal dynamics and the frequency content of the signal [32].
- **Time-Frequency Distribution (TFD):** The TFD is a mathematical approach that generates a hybrid time-frequency illustration of the signal. By simultaneously capturing the signal's temporal dynamics and frequency content, the TFD produces a comprehensive representation of the signal that encompasses both domains [33].

Time-frequency features offer several benefits, such as their precise capture of the signal's time-varying and spectral characteristics, which can significantly enhance the accuracy of machine learning models. Time-frequency features can also provide insights into the underlying characteristics of the signal, such as the presence of transient events or non-stationary behavior. However, time-frequency features also have limitations. They may require additional computational resources compared to spectral or temporal features alone, and they may be sensitive to noise and artifacts in the signal, which can affect their accuracy and reliability. Moreover, identifying and optimizing time-frequency features can be a complex task due to the abundance of time-frequency representations available, and the selection of the appropriate representation can significantly impact the performance of machine learning models.

[34] utilizes a wavelet-based method to extract time-frequency features from electroencephalogram signals. The reported average accuracy and F1 score of sleep stage classification are 78% and 84%, respectively.

3.1.3. Feature selection

The process of feature selection is of utmost importance in signal processing, requiring the identification of a pertinent subset of features from a larger pool of candidate features. The feature selection techniques are employed to decrease the dimension of the feature space, to increase the precision of machine learning models, and to improve the comprehensibility of the findings. Outlined below are some of the most prevalent techniques employed that decrease the dimension of the feature space.

- **Filter methods**: Filter methods involve evaluating and selecting the top-ranked features based on statistical measures, such as correlation or mutual information. Filter methods are characterized by their computational efficiency and can be applied to large datasets, but they may not consider the interaction between features [35, 36].
- **Wrapper methods:** Wrapper methods assess the efficacy of machine learning models using feature subsets and choose the subset that yields the highest performance. Wrapper methods take into account the interplay between features and can enhance the precision of machine learning models, but they are computationally expensive [35, 36].
- **Embedded methods**: Embedded methods incorporate feature selection into the machine learning algorithm itself, selecting the most relevant features during the model training

process. Embedded methods are computationally efficient and can improve the accuracy of machine learning models, but they may not be suitable for all machine learning algorithms [35, 36].

Other feature selection methods include hybrid methods that combine filter, wrapper, and embedded methods, as well as model-based methods that use statistical models to identify relevant features. The choice of feature selection method depends on the specific application and the characteristics of the dataset.

In [37] , an algorithm is presented that analyzes polysomnography signals by extracting a range of temporal and spectral features. The algorithm then employs various filter methods to select the best features. In [38], a wrapper-based feature selection method is employed to extract the key features from electrocardiogram and oxygen saturation signals in both time and frequency domains, for the detection of sleep apnea. The study reports an accuracy of 86.90%.

3.2. Classifiers

Here, the methods used in automatic sleep studies are categorized into two main classes: (i) methods that use traditional machine learning algorithms and (ii) methods that use powerful deep learning methods. While traditional machine learning methods need to receive features as input to act as a classifier, deep learning methods can either act as a classifier alone based on the given features or perform both feature extraction and classification together. In this section, these algorithms and studies that have utilized these algorithms are presented. Prior to delving into this topic, the evaluation criteria for the performance of these algorithms will be briefly discussed.

3.2.1. Traditional Machine Learning Methods

The upcoming subsections will provide an explanation of traditional classification methods and related studies. Commonly used classifiers are the following:

- **Naive Bayes classifier:** The Naive Bayes classifier is a probabilistic technique that applies Bayes' theorem to predict the class probability of a given input belonging to [39]. The investigation conducted in [40] adopts a systematic methodology, employing Empirical Mode Decomposition (EMD) to analyze ECG signals and extract statistical features from intrinsic mode functions (IMFs) for sleep apnea detection. The dataset used is the physionet apnea-ECG database. While this technique offers a detailed temporal-frequency representation of the ECG signal, its exclusive reliance on ECG data raises questions about the discriminative capacity of the selected features and the efficacy of the Naive Bayes Classifier, resulting in a reported accuracy of 39.47%. To enhance classification performance, further exploration into sophisticated signal processing techniques and feature selection methods is warranted. In contrast, the approach taken in [41] embraces a multi-modal strategy by integrating ECG and SpO2 signals, achieving a commendable accuracy of 88.86% for sleep apnea detection from the physionet apnea-ECG database. The strengths lie in the utilization of complementary physiological data and the preservation of temporal information through a feature-level fusion scheme. However, the complexity introduced by integrating different modalities raises concerns about potential

overfitting, necessitating careful consideration of challenges associated with synchronizing heterogeneous data streams. The application of advanced statistical modeling or regularization techniques may mitigate these challenges and improve the model's generalizability. The study presented in [42] concentrates on real-time sleep apnea diagnosis using wearable devices, analyzing heart rate and movement signals. While continuous monitoring's practicality is a significant advantage, the reliance on simple statistical features like variance and mean may limit the classifier's discriminative power. Additionally, real-time implementation challenges, such as power consumption and data transmission, necessitate further investigation. Exploring advanced feature extraction methods and signal processing techniques may enhance the model's accuracy and efficiency. Within the realm of sleep stage classification, two studies, [43] and [44], employ the Naive Bayes Classifier, each applying distinctive methodologies for signal processing, feature extraction, and classification accuracy. In [43] , the focus is on EEG signal analysis, with signal preprocessing conducted through wavelet decomposition using the Maximal Overlap Discrete Wavelet Transform (MODWT) from the Sleep-EDF database. The feature extraction process involves employing 5-second epochs with non-overlapping temporal windows, extracting a comprehensive set of features, including the estimation of signal power in eight brain rhythms, cross-frequency coupling, and a total of 120 features per epoch. Impressively, the Naive Bayes Classifier achieves an accuracy of 94.4%, underscoring the efficacy of utilizing EEG signals processed through advanced wavelet decomposition techniques and the richness of the feature set. On the other hand, [44] takes a unique approach by utilizing only one-lead ECG for automatic sleep staging from the Stroke Volume Data Bank dataset. Initial preprocessing involves calculating Heart Rate Variability (HRV) signals from denoised ECG signals, with subsequent extraction of HRV and respiratory amplitude variability (RAV) signals. Features are derived from both time and frequency domains, and while the main classification method employs a Long Short-Term Memory (LSTM) network, a Naive Bayes Classifier is also integrated for comparative analysis. The Naive Bayes Classifier achieves an accuracy of 67.6%, while the LSTM network outperforms with an accuracy of 84.07%. This study highlights the utilization of ECG signals for sleep staging and explores the potential of a Naive Bayes Classifier alongside more complex models. However, the comparatively lower accuracy suggests the need for further optimization or exploration of alternative classifiers.

- **k-Nearest Neighbors:** The k-Nearest Neighbors classifier is a classification algorithm that determines the class a new input belongs to by examining the most frequently occurring class among its k-nearest neighboring points in the training dataset [45]. In [46], a Piezo-Electric (PE) system captures comprehensive body motion, rib cage movements, and cardioballistic effects, employing a sophisticated Apnea/Hypopnea detection algorithm with KNN. Achieving a sensitivity of 88% and specificity of 89%, the study excels in signal analysis's comprehensiveness, extracting multiple physiological parameters (see Table 2). Yet, the algorithm's intricacy may pose challenges in computational efficiency. Nasal airflow signals take center stage in [47], where KNN attains an 80.1% accuracy. Leveraging the Teager Energy Operator for signal evaluation, potential apnea episodes are identified based

on energy level changes. While KNN performs commendably, the study's inclusion of SVM and linear regression underscores the importance of employing multiple classifiers for robust performance. In [48], a wavelet-based transform decomposes electrocardiogram (ECG) signals, with KNN achieving a 77.57% accuracy, outperforming SVM but trailing behind other classifiers. Emphasizing the significance of frequency sub-band analysis, the study highlights trade-offs between accuracy and model complexity associated with classifier choice. [41], previously discussed for Naive Bayes Classifier, also integrates KNN for sleep apnea detection, achieving an accuracy of 80.96%. While competitive, it lags behind the Naive Bayes Classifier in the same study, signaling the necessity for tailored classifier selection based on the dataset and problem characteristics. Utilizing smartphone sensors,[49] employs KNN and Classification and Regression Trees (CART) for sleep apnea classification, achieving remarkable accuracies of 97.7% and 98.2%, respectively. Diverse sensor data and comprehensive feature extraction contribute to the high accuracy, showcasing the potential of smartphone-based approaches. In [50], ultra-wideband physiological sensing technology embedded in a mattress incorporates KNN, SVM, decision tree, and adaptive boost, with KNN exhibiting the best performance. The study underscores sensor technology's importance in sleep apnea detection and posits KNN as well-suited for this application. For ECG-based sleep apnea detection in [51], SVM slightly outperforms KNN, employing QRS complexes detection, heart rate variability (HRV) parameter analysis, and wavelet-based methods for signal preprocessing. The study underscores the significance of comprehensive signal processing techniques. In [42], KNN, alongside Naive Bayes Classifier, achieves an 80.96% accuracy (see Table 2). While performing well, the classifier choice should align with specific dataset characteristics. In summary, selecting sleep apnea detection methods involves a nuanced trade-off between signal processing complexity, feature richness, and classifier choice. Smartphone-based ([49]) and mattress-based systems ([50]) demonstrate high accuracies with KNN, influenced by physiological signals and signal analysis depth. Context-specific classifier superiority ([41], [51], [42]) stresses the need for tailoring the approach to unique dataset and physiological signal characteristics. Further research could explore hybrid approaches, optimizing both accuracy and computational efficiency. In [52], employing Doppler radar technology for sleep stage classification, KNN's application is justified by its adeptness in handling complex, multidimensional data. The non-contact approach aligns seamlessly with KNN's adaptability, and the comprehensive feature set benefits from its ability to discern patterns in high-dimensional spaces. The accuracy of 86.6% surpasses traditional classifiers, emphasizing KNN's efficacy. In [53], focused on Periodic Limb Movement Disorder (PLMD) detection through accelerometer and gyroscope data, KNN emerges as the most effective classifier. Its inherent simplicity aligns well with straightforward features from movement data. KNN excels by leveraging spatial and temporal dependencies in movement patterns associated with PLMD, showcasing its prowess in discerning complex relationships. The accuracy of 96.92% solidifies KNN's suitability for this specific movement disorder detection task. For more details see Table 2.

In summary, KNN's effectiveness in these studies stems from its adaptability to diverse data types, ability to handle complex, multidimensional data, and aptitude for discerning patterns in high-

dimensional spaces. Its simplicity makes it particularly well-suited for scenarios where straightforward features are extracted, enabling excellence in applications beyond traditional sleep apnea detection.

Table 2. Sleep studies based on k-Nearest Neighbors

Paper	Year	Signal	Case of study	Dataset	Evaluation criteria
[40]	2015	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=69.72%
[46]	2016	Piezo-electric sensor	Sleep Apnea Detection	Subjects recruited by researchers	Sensitivity=88% Specificity=89% positive predictive value = 81% negative predictive value = 93%
[47]	2016	Nasal respiratory airflow signals	Prediction of Sleep Apnea	Subjects recruited by researchers	Accuracy=80.1% Sensitivity=79.5% Specificity=80.60%
[48]	2017	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=77.57%
[41]	2017	ECG, SPO2	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=80.96 %
[49]	2017	Data from sensors of smart phones	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=97.7%
[52]	2018	CW Doppler Radar	Sleep Stage Estimation	Subjects recruited by researchers	Accuracy=86.6%
[50]	2019	UWB signals	Sleep Apnea Detection	Subjects recruited by researchers	Sensitivity=71.5% Specificity=98.02%
[51]	2020	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=94.7%
[42]	2020	Respiration, SpO2, heartrate, 3-ACC signals	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=95%
[54]	2015	EEG	Narcolepsy detection	Subjects recruited by researchers	Sensitivity=32.90% Specificity=97.50% positive predictive value =62.20% negative predictive value = 91.90%
[53]	2017	Motion signal	Periodic limb movement disorder	Subjects recruited by researchers	Accuracy=96.62%

Ensemble Learning-Based Classifiers: Ensemble Learning-Based Classifiers are a type of machine learning algorithm that combine multiple base classifiers to improve the overall predictive performance [55]. In [56], the application of the bagging method with an ADtree classifier for real-time sleep apnea diagnosis exemplifies the prowess of ensemble methods in managing diverse physiological data. The approach, adept at processing SpO2 signals on a mobile device, accentuates the adaptability and parallel processing capabilities inherent in ensemble techniques, resulting in an impressive accuracy of 95.96% (see Table 3). Nevertheless, caution is warranted, considering potential computational demands and the specter of overfitting if not judiciously tuned. In [57], a real-time sleep apnea detection system leans on basic statistics and an extended definition of the oxygen desaturation index (ODI), employing the bagging approach for classification. With a notable accuracy of 83.26%, the method surpasses other classifiers, showcasing the efficiency in computation time—an admirable strength in mitigating potential weaknesses associated with more computationally intensive methods like SVM and MLP. A return

to [40] reveals the integration of Ensemble Learning-Based Classifiers for sleep apnea detection, achieving an accuracy of 80.07% (see Table 3). The ensemble approach fortifies the model's resilience in handling complex physiological data, though potential vulnerabilities may surface in scenarios prioritizing computational efficiency due to the method's potentially higher computational demands. [58], employing the random forest classifier, emerges as a standout in achieving an accuracy of 95.5% in sleep apnea detection. The ensemble method excels in processing signals from multiple regions for respiration analysis, showcasing prowess in capturing complex relationships within diverse datasets. Yet, potential pitfalls, such as the need for meticulous parameter tuning and susceptibility to overfitting, linger on the horizon. Journeying to [48], where Ensemble Learning-Based Classifiers elevate accuracy to 82.27%, the approach's strength lies in its improvement over individual classifiers like KNN. This ensemble methodology effectively captures intricate patterns within respiration signals. However, potential drawbacks may include the requisite for a diverse set of base classifiers and careful consideration of ensemble size. In [59], the AdaBoost classifier extracts spectral and nonlinear features for sleep apnea detection, securing an accuracy of 92.9% (see Table 3). While the ensemble approach deftly leverages diverse features, demonstrating adaptability, potential vulnerabilities may surface in scenarios with imbalanced datasets where AdaBoost could prove sensitive to outliers. Conversely, [50], previously discussed in the context of KNN, incorporates Ensemble Learning-Based Classifiers, yet falls short of KNN's performance. This comparison underscores the critical importance of selecting a classifier aligned with dataset characteristics, cautioning against introducing unnecessary complexity without a proportional boost in performance. Likewise, [51], previously discussed with KNN, embraces Ensemble Learning-Based Classifiers. The nuanced interplay between individual classifiers and ensemble methods underscores the importance of understanding specific requirements in sleep apnea detection applications. Comparative performance underscores the necessity of selecting methods attuned to dataset characteristics and physiological signal nuances. In conclusion, Ensemble Learning-Based Classifiers unveil remarkable strengths in managing diverse physiological signals, unraveling complex relationships, and fortifying robustness in sleep apnea detection. The judicious choice between methods necessitates a delicate balance, weighing factors like computational efficiency, dataset characteristics, and the potential trade-offs between heightened accuracy and additional complexity introduced by ensemble methods. Each study paints a vivid picture of the adaptability and efficacy of ensemble techniques, reinforcing the crucial role of thoughtful method selection in navigating the intricacies of sleep apnea detection. Transitioning to sleep stage classification in [60], an analysis of electroencephalogram (EEG) data incorporates three distinct feature selection methods, employing the bootstrap aggregating (bagging) algorithm for classification, resulting in a 75% accuracy. The strength lies in the use of diverse feature selection methods, providing a comprehensive understanding of EEG signals. However, the achieved accuracy may indicate potential limitations in capturing the complexity of sleep stage transitions. [61] and [62] delve into Empirical Mode Decomposition (EMD) for sleep stage classification using EEG signals. Bagging and Adaptive Boosting serve as classifiers, showcasing the versatility of ensemble methods. Impressive accuracies of 90.69% and 97.73% emphasize the efficacy of the EMD-ensemble approach (see Table 3). The strength lies in the ability of EMD to capture localized time-frequency estimations, bolstered by ensemble methods, ensuring robust classification. However, potential weaknesses

may include increased computational demands. [63] proposes a sleep stage classification method utilizing Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. Adaboost is employed as a classifier, achieving an accuracy of 89.52%. The method's strength lies in the adaptive noise reduction capabilities of the decomposition process, contributing to accurate feature extraction. However, potential weaknesses may arise in scenarios where noise levels vary significantly across different datasets. In [64], a diverse set of statistical and spectral features, including Bootstrapped Aggregating, achieves an accuracy of 86.53%. The inclusion of both statistical and spectral features provides a comprehensive representation of EEG signals. Conversely, a variant using only spectral features and boosted decision trees yields a lower accuracy of 82.03%. This comparison emphasizes the importance of feature selection in achieving optimal classification results. [65] introduces a method for sleep stage classification using EEG signals, incorporating a tunable Q-factor wavelet transform for signal decomposition and subsequent feature extraction. Random Forest is employed as a classifier, achieving an impressive accuracy of 97.5%. The strength lies in the precision of the wavelet transform, enhancing the discriminative power of extracted features. However, potential weaknesses may include the computational demands associated with the precise decomposition process. [66] focuses on EEG signals, extracting diverse time, frequency, and nonlinear features for sleep stage classification. Random Forest serves as the classifier, yielding an accuracy of 85.95%. The comprehensive feature set provides a nuanced representation of EEG signals, while the ensemble approach enhances the robustness of classification. Potential weaknesses may include challenges in interpretability due to the complexity of the feature set. In [67], a method for sleep stage classification based on EEG and EOG signals employs a Random Forest classifier, achieving an accuracy of 92.60%. The diverse set of time and frequency features captures nuanced patterns in EEG and EOG signals, demonstrating the efficacy of ensemble methods. However, potential weaknesses may include increased complexity in feature interpretation. [68] introduces a two-layer stacked ensemble model for sleep stage classification using EEG signals. The model combines Random Forest and LightGBM, achieving an accuracy of 91.20%. The incorporation of both subject-specific and signal-derived features contributes to a holistic understanding. The ensemble model's strength lies in the balanced integration of two classifiers, but potential weaknesses may include increased model complexity. In [69], a method for detecting REM sleep behavior disorder utilizes features from EEG, EOG, and EMG signals. Random Forest serves as the classifier, achieving an accuracy of 96%. The diversity of time, frequency, and nonlinear features captures intricate patterns associated with REM sleep behavior disorder. The strength lies in the comprehensive feature set, but potential weaknesses may include challenges in handling multi-modal signals. [70] proposes a method for cyclic alternating pattern (CAP) identification, utilizing EEG signals and wavelet decomposition. Hjorth parameters and wavelet entropy are calculated, and an ensemble of bagged trees (EBagT) serves as the classifier, achieving an accuracy of 78%. The strength lies in the adaptability of wavelet decomposition for capturing patterns in EEG signals. However, the achieved accuracy may indicate potential challenges in distinguishing CAP patterns. In summary, the Ensemble Learning-Based Classifiers showcased in these studies demonstrate versatility and effectiveness in capturing complex patterns within sleep stage classification. The choice of classifier depends on the characteristics of the dataset, the interpretability of features, and the computational demands associated with specific methods. Each study underscores the importance

of thoughtful method selection and feature extraction in optimizing accuracy and robustness in sleep stage classification. Table 3 has resumed all above mentioned methods.

- Table 3. Sleep studies based on Ensemble Learning Based Classifiers

Paper	Year	Signal	Case of study	Dataset	Evaluation criteria
[56]	2010	SPO2	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=93% AUC=98.5% Sensitivity=92.35% Specificity=93.52%
[57]	2012	ECG, SPO2	Sleep Apnea Detection	UCD database	Accuracy=83.26% Sensitivity=79.05% Specificity=84.62%
[60]	2014	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=75%
[61]	2016	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=90.69%
[62]	2016	EEG	Automatic sleep scoring	Sleep EDF database	Accuracy=97.73%
[63]	2015	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=89.52%
[64]	2015	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=86.53%
[71]	2015	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=82.03%
[40]	2015	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=80.07%
[58]	2015	Abdominal, chest and nasal respiratory signals	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=95.50% F-score = 0.95 Kappa=0.93
[65]	2016	EEG	automatic sleep staging	Sleep EDF database	Accuracy=97.50%
[6]	2016	EEG	Automatic Sleep Staging	Sleep EDF Database	Sensitivity=89.06% Specificity=98.61% Accuracy=93.13%
[72]	2016	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=91.36%
[73]	2017	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=96.55%
[74]	2017	EEG	Automatic Sleep Staging	Sleep EDF database	Accuracy=83.49%
[48]	2017	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=82.27%
[66]	2018	EEG	Automatic Sleep Stage Scoring	CCSHS	Accuracy=85.95% Kappa coefficient=0.8046
[67]	2018	EEG, EOG	automated sleep staging	ISRUC sleep database	Accuracy=92.6% Kappa coefficient=0.856
[59]	2019	SPO2	Estimate Sleep Apnea Severity	Subjects recruited by researchers	Accuracy=92.9% Kappa coefficient=0.479
[50]	2019	UWB signals	Sleep Apnea Detection	Subjects recruited by researchers	Sensitivity=55.18% Specificity=95.63%
[68]	2020	EEG	Automatic Sleep Staging	Sleep-EDF database and Sleep-EDF Expanded database	Accuracy=91.2% weight F1-score=0.916 Kappa coefficient=0.864 Sensitivity=72.52%
[75]	2020	Demographic information, EEG	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=91.30% Precision=88.51% Recall=95.06% F1-score=91.67% AUC=91.28%
[51]	2020	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=96.1%

[69]	2019	PSG	REM behavior disorder	MASS, CAP, Subjects recruited by researchers	Accuracy=96 % Sensitivity= 83% Specificity=98% F-score =81%
[70]	2021	EEG	cyclic alternating pattern identification	CAP sleep database	Accuracy=78% F-score = 0.775 Kappa coefficient=0.56

Discriminant Analysis: Discriminant Analysis classifier is a statistical method used for predicting the class membership of a sample based on a set of predictor variables [76]. [40] employed Discriminant Analysis for sleep apnea detection from physionet apnea ECG database, yielding an accuracy of 64.60%. This performance, while indicative of the simplicity and interpretability of Discriminant Analysis, falls short when compared to more intricate models such as KNN and ensemble learning methods, suggesting the need for a more exhaustive examination of classification methodologies. Secondly, [77] introduced a radar panel positioned beneath a mattress for sleep apnea detection. In this study subjects were recruited by researchers. The two-phase feature extraction process showcased a commendable endeavor in capturing changes in amplitude and frequency during the initial phase, featuring metrics such as The Mean Absolute Deviation, Number of times the signal crosses the mean, The variance of the time duration, and the inter-quartile range. The subsequent phase involved processing envelope differences for each epoch. Despite the technological innovation, the achieved accuracy of 70% underscores potential avenues for improvement, warranting a deeper exploration of advanced signal processing techniques or alternative classifiers. Thirdly, [78] proposed a method combining cardiopulmonary coupling features with heart rate variability-based features, yielding a notable accuracy of 89.80% for sleep apnea detection from physionet apnea ECG database. While the high accuracy signifies the potential of feature combination, a rigorous analysis is warranted to quantify the specific contributions of each feature and assess potential computational complexities. Conversely, [46], discussed in previous sections, demonstrated a diminution in accuracy with Discriminant Analysis classification compared to KNN and ensemble learning methods, underscoring the imperative for a methodological paradigm shift. In paper [48], the focus was on sleep apnea detection using ECG signals. The case of study involved utilizing the Physionet Apnea-ECG database as the dataset. The evaluation criteria for the study were based on accuracy, with a reported result of 39.95%. In a different study, [79] strategically integrated three signals—heart rate, breathing effort, and oxygen saturation—leveraging power spectral density features and achieving an accuracy of 87% for Sleep Apnea Detection. In this study, subjects were recruited by researchers. The triumph of multi-signal integration not only underscores its potential but beckons further scientific exploration. Lastly, [51], discussed earlier, achieved a remarkable accuracy of 98.70% with Discriminant Analysis classification. Elucidating the intricacies contributing to this success is paramount for deriving scientific principles governing effective sleep apnea detection.

Extreme Learning Machine (ELM): Extreme Learning Machine classifier is a type of feedforward neural network that uses a single hidden layer with randomly generated weights and biases to learn the mapping between input and output data [80]. Explored earlier [40] , the incorporation

of Extreme Learning Machine (ELM) yielded remarkable results, boasting an impressive accuracy of 83.77%. This surpassed the performance of alternative classifiers investigated in the study, affirming the prowess of ELM as a discerning force in unraveling the intricate patterns associated with sleep apnea. Conversely, [48], scrutinized in the same context, integrated ELM but achieved a diminished accuracy of 53.02%, positioning it as the least precise among the classifiers under evaluation. This lower accuracy suggests potential challenges in capturing the nuanced patterns of sleep apnea within the specific dataset, pointing to considerations regarding the effectiveness of ELM in certain contexts. In essence, the incorporation of ELM as a classifier for sleep apnea detection resulted in divergent outcomes. The findings of [40], showcasing a robust accuracy of 83.77%, stand in contrast to the less accurate performance of [48] at 53.02% , revealing potential limitations in its efficacy for deciphering sleep apnea patterns within the unique dataset. The selection of this approach may hinge on the nuanced characteristics and complexities inherent in the dataset under consideration.

Support Vector Machine (SVM): Support Vector Machine classifier is a supervised machine learning algorithm that separates data into classes by finding an optimal hyperplane in a high-dimensional space using a kernel function [81]. In the case of [41] (see Table 4), where SVM achieved the highest accuracy of 96.64%, the classifier demonstrated its prowess in effectively discerning intricate sleep apnea patterns. This remarkable performance suggests that SVM excels in scenarios where the dataset exhibits diverse and complex features related to sleep disorders. Conversely, challenges arise in studies like [50], where SVM yielded a sensitivity of 51.07%. This points to potential limitations in capturing specific types of sleep apnea events, indicating the need for a careful consideration of the dataset's characteristics and the nuances of sleep patterns when employing SVM. Innovative applications, as seen in [51] and [82], showcase SVM's adaptability to unconventional data sources such as blood oxygen levels and smart pillow data. Achieving impressive accuracies of 98.70% and 93.52%, respectively, these studies underscore SVM's versatility, making it a valuable choice for incorporating novel modalities into sleep apnea detection systems. Furthermore, in studies like [83] and [84], where SVM was applied to multimodal signal processing, accuracies of 80% and 80.90% were achieved, respectively. Although not the highest accuracies, these results highlight SVM's potential for integrating different physiological signals, showcasing its ability to contribute to comprehensive sleep apnea detection approaches. In [85], the success of SVM in achieving an accuracy of 92.93% underscores its strength in handling complex feature spaces derived from time-frequency analysis of electroencephalogram signals. This signifies SVM's proficiency in decoding intricate patterns crucial for sleep stage classification, particularly in studies employing advanced signal processing techniques. Similarly, in [86], where SVM yielded an accuracy of 87.50% for sleep stage classification using EEG signals and visibility graph mapping, its ability to navigate unconventional signal representations is evident. The method's success emphasizes SVM's flexibility, making it suitable for tasks involving diverse data modalities. The multimodal approach of [87], integrating EEG, EOG, and EMG signals, showcases SVM's capability in achieving an accuracy of 88%. SVM proves effective in handling a comprehensive set of time and frequency domain features, demonstrating its robustness in scenarios where multiple physiological signals contribute to sleep stage classification. However,

[88] presents a unique scenario where SVM's accuracy of 79.70% is comparatively lower than other classifiers in the study. This highlights the importance of judicious classifier selection based on dataset characteristics and the nature of extracted features, underscoring that SVM's performance may vary based on specific study requirements. In [89], the intricate feature extraction steps applied to EEG signals result in SVM achieving an accuracy of 91.80%. This highlights SVM's efficiency in handling complex preprocessing methods, making it a suitable choice for studies requiring in-depth signal analysis and sophisticated feature extraction. For [90], utilizing EOG signals for sleep stage classification, SVM achieves an accuracy of 83.10%. The success in this study points to SVM's adaptability to different signal modalities and reinforces its potential for applications beyond traditional EEG-based studies. A distinctive application is observed in [91], where SVM achieves 80% accuracy in detecting psychophysiological insomnia based on functional magnetic resonance imaging. This showcases SVM's versatility in handling diverse data sources, even in unconventional sleep-related studies. In conclusion, SVM emerges as a versatile and potent classifier in sleep-related studies, excelling in scenarios demanding advanced signal processing, intricate feature extraction, and adaptability to diverse physiological signal modalities. While its performance is generally robust, careful consideration of study-specific requirements is crucial for optimizing SVM's effectiveness in different sleep-related contexts. The comparison of the methods based on SVM is brought in Table 4.

Table 4. Sleep studies based on SVM

Paper	Year	Signal	Case of study	Dataset	Evaluation criteria
[47]	2016	Nasal respiratory airflow signals	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=87.6% Sensitivity=91.3% Specificity=77.2%
[48]	2017	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=75%
[41]	2017	ECG, SPO2	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=96.64%
[50]	2019	UWB signals	Sleep Apnea Detection	Subjects recruited by researchers	Sensitivity=51.07% Specificity=99.3%
[51]	2020	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	accuracy=98.7%
[92]	2009	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database UCD	Accuracy=92.85% Kappa coefficient=0.85
[85]	2013	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=92.93%
[82]	2013	SPO2	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=93.52% Sensitivity=95.71% Specificity=89.47%
[86]	2014	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=87.5% Kappa coefficient=0.81
[87]	2015	EEG, EOG, EMG	Automatic Sleep Staging	PSG data collected at DyCog Lab of Lyon Neuroscience Research Center	Accuracy=88% Sensitivity=78% Specificity=92%
[83]	2017	ECG, RES	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=80% Sensitivity=60% Specificity=100%
[88]	2018	EEG	Automatic Sleep Staging	MASS	Accuracy=79.70% F-score =75.01%
[89]	2018	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=91.8% Kappa coefficient=0.87
[90]	2018	EOG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=86.81%
[84]	2019	SPO2, PPG	Sleep Apnea Detection	Subjects recruited by researchers	AUC=80.9% Sensitivity=72.7% Specificity=89.1%
[93]	2020	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=96%

					Kappa coefficient=0.87
[91]	2021	fMRI	insomnia detection	Subjects recruited by researchers	Accuracy=80% Specificity=81% Precision = 79% Recall= 79%
[94]	2018	EEG	insomnia detection	Subjects recruited by researchers	Sensitivity=88% Specificity=91%

3.2.2. Deep Learning Methods

Deep networks can function as either a standalone classifier or as a combined feature extractor and classifier, depending on the specific architecture and training methodology used.

Frequently used deep neural networks for sleep study are as follows:

Deep belief net (DBN): A DBN is a type of generative neural network that is composed of multiple layers of hidden units trained in an unsupervised manner to extract high-level features from input data and generate new samples from the learned distribution [95] . In [96], a groundbreaking methodology is presented for sleep apnea detection, leveraging blood oxygen saturation signals. The signals undergo meticulous preprocessing before being channeled into a Deep Belief Net neural network. The network's initial layers undergo unsupervised training, facilitating the extraction of intricate features. As the model progresses, a supervised learning approach is employed, fine-tuning the network based on the acquired features. The result is a staggering accuracy of 97.67% on Physionet Apnea-ECG database, underscoring the prowess of Deep Belief Nets in deciphering complex physiological signals for highly accurate sleep apnea detection. This study also utilized the UCD database and achieved an accuracy of 85.26%. Similarly, [97] delves into the domain of automatic sleep staging, focusing on electrooculogram signals transformed into the frequency domain. The data is intelligently segmented into smaller batches, optimizing the input for a Deep Belief Net endowed with three hidden layers. This strategic architectural choice accommodates the intricate nuances of sleep staging. The achieved accuracy of 83.30% (on Sleep EDF Database) demonstrates the efficacy of Deep Belief Nets even in tasks requiring the interpretation of multifaceted signal data. The strength of Deep Belief Nets in these studies lies not only in their ability to discern patterns within the data but also in their hierarchical learning approach. The unsupervised pretraining allows the network to autonomously identify underlying structures, while the subsequent supervised fine-tuning refines the model for precise and nuanced classification. This dual-phase learning paradigm contributes to the robustness and adaptability of Deep Belief Nets in decoding complex physiological signals related to sleep disorders. However, it's crucial to note that the application of Deep Belief Nets is not without challenges. The architectural complexity and the need for substantial labeled data for supervised learning can pose hurdles. Additionally, the interpretability of the learned features in the hidden layers may be a consideration in certain applications. In essence, the studies employing Deep Belief Nets showcase their unparalleled ability to tackle intricate tasks in sleep-related research. The impressive accuracies achieved in sleep apnea detection and automatic sleep staging attest to the capacity of Deep Belief Nets to navigate the complexities of physiological signals, marking a significant stride in the pursuit of advanced sleep disorder diagnostics.

Recurrent Neural Network (RNN): Recurrent Neural Network (RNN) is a distinctive form of neural network structure designed to effectively process sequential data. It achieves this by integrating feedback connections, which facilitate the transfer of information from one sequence step to another. This characteristic empowers the network to capture the relationships and patterns that evolve over time, enabling it to comprehend temporal dependencies and context within the data [98]. [99] offers a multifaceted approach, incorporating heartbeat, respiratory, and movement signals for Automatic Sleep Staging. The comparative exploration of a 1-dimensional convolutional neural network (CNN) and an LSTM-based recurrent neural network introduces an insightful dimension. While the CNN outperformed the LSTM in this specific context, achieving a higher accuracy, it's crucial to dissect the factors contributing to this outcome. CNNs excel in capturing spatial dependencies within data, and the nature of physiological signals might align more naturally with this architecture. The LSTM's relatively lower accuracy (51%) unveils its sensitivity to the intricate temporal dependencies of the sleep-related signals under consideration. In this study subjects were recruited by researchers. In contrast, [100] navigates the domain of predicting sleep apnea occurrences using electrocardiogram segmented samples. The exceptional accuracy of 94.95% achieved by the LSTM network underscores its proficiency in learning and exploiting temporal dependencies within the data. However, the study doesn't delve into potential challenges or limitations faced during training, which could be valuable insights for future implementations. In this study Subjects were recruited by researchers. [101] delves into automatic sleep staging with multiple LSTM networks handling EEG, EOG, and EMG signals. The achieved accuracy of 83.07% is commendable, considering the complexity of the physiological signals involved. The study's strength lies in the comprehensive exploration of LSTM structures, allowing for a deeper understanding of how variations in network architecture impact performance. However, a detailed examination of computational complexity or training time would enhance the analysis. The PhysioNet dataset is used in this study. [102] adopts an intriguing perspective by proposing a method for sleep apnea detection using LSTM-RNN with Instantaneous Heart Rate and oxygen saturation signals. The Physionet Apnea-ECG dataset is used. The achieved accuracy of 92.1% speaks to the adaptability of recurrent networks to signals obtainable from wearable devices. The separate analysis of each signal type unveils the nuanced strengths of each, with oxygen saturation demonstrating higher accuracy. However, a critical examination of potential challenges, such as signal noise or varying physiological conditions, could enrich the discussion. In paper [103], the research focused on Sleep apnea severity detection using Electrocardiogram signals. The case of study involved utilizing the Physionet Apnea-ECG database. The evaluation criteria for the study included accuracy, with a reported result of 99%, and an F1-score of 0.99. In summary, the efficacy of LSTM networks in sleep-related tasks is evident, but their performance is contingent on multiple factors. The choice between CNNs and LSTMs demands a meticulous consideration of signal characteristics, and understanding the intricacies of each network type is crucial for informed model selection. Future studies could benefit from a more explicit discussion of challenges faced during training, potential overfitting concerns, and a deeper exploration of the interpretability of learned features, contributing to a more holistic understanding of the methodologies employed.

- **Convolutional neural network (CNN):** CNN utilizes convolutional layers to extract and learn local features from input data and pooling layers to reduce the dimensionality of the feature maps [104]. In [105], the use of CNNs for classifying electrocardiogram signals stands out. The network structure, comprising multiple 1-dimensional convolutional layers, demonstrates an accuracy of 88.23%. The success of this model suggests that CNNs are adept at capturing features related to heart rate variability and R-wave intervals. However, it would be beneficial to delve into potential challenges faced during training, such as class imbalance or noise in the signals. The work presented in [106] introduces a multifaceted CNN architecture, employing parallel paths and multiple convolutional layers for diverse purposes. Achieving an accuracy of 86.6%, the study showcases the effectiveness of CNNs in handling complex signal dependencies. Yet, a more detailed exploration of computational requirements and potential trade-offs in accuracy versus model complexity would augment the analysis. [107] adopts a unique approach, constructing visibility graphs from electroencephalogram signals, and employing parallel CNNs for both raw signals and graphs. The accuracy of 87.21% underscores the viability of this method. However, understanding the interpretability of features extracted from visibility graphs and the rationale behind employing parallel networks would provide valuable insights. Moving to Automatic Sleep Staging, [108] (see Table 5) applies CNNs to raw EEG signals, achieving an accuracy of 74%. While this accuracy might be considered moderate, it prompts a critical evaluation of the network architecture and the inherent challenges of mapping raw EEG signals to sleep stages. In [109], a novel model, FDCCNN, is proposed for Automatic Sleep Staging, addressing imbalanced datasets. With an impressive accuracy of 92%, the study highlights the potential of specialized CNN architectures. However, a more in-depth discussion of the model's interpretability and generalizability would fortify the analysis. [110]delves into spectral analysis of EEG signals, transforming sleep patterns into interpretable images for CNN input. The VGGNet-based CNN achieves an accuracy of 86%, showcasing the efficacy of this novel approach. Further exploration of the interpretability of the images and potential correlations with sleep stages could deepen the understanding. [111] presents a CNN for Automatic Sleep Staging with raw EEG signals, featuring 12 convolutional and 2 fully connected layers. The accuracy of 87% positions it favorably among the discussed studies, and a nuanced examination of the layer architecture and its impact on performance would enhance the analysis. [112] ventures into sleep apnea detection using 3D facial scans converted to 2D images and transfer learning with VGGFace. Although the accuracy of 67.42% might be perceived as moderate, the innovative application of CNNs to facial scans warrants discussion on the feasibility and generalizability of this approach. [113] employs CNNs for automatic sleep staging using EEG, EOG, and EMG signals, achieving an accuracy of 82.3%. The utilization of time-frequency images and the CNN architecture could be explored in greater detail to discern the factors influencing performance. In [114], a DeConvolutional Neural Network is proposed for Automatic Sleep Staging in child patients, achieving an accuracy of 90.89%. The unique architecture warrants scrutiny, especially regarding its ability to capture latent features deemed unimportant in the convolutional block. [115] applies CNNs to ECG signals for Automatic Sleep Staging, featuring both local and dilated convolutional blocks. The

accuracy of 77% prompts consideration of the balance between local and long-range feature extraction, shedding light on the network's adaptability to physiological signals. [116]introduces a smart belt for detecting chest movements, converting 1D signals to 2D images, and employing CNNs for sleep apnea detection with an impressive accuracy of 94%. The wearable nature of this method invites a discussion on its practicality and potential integration into daily life. In conclusion, the analysis of these CNN-based studies reveals the manifold applications of this architecture in sleep-related tasks. However, a more nuanced discussion of network interpretability, potential challenges faced during training, and trade-offs between model complexity and accuracy would contribute to a more comprehensive understanding of their strengths and limitations in diverse contexts. The performance of the CNN based methods is brought in Table 5.

Table 5. Sleep studies based on convolutional neural network

Paper	Year	Signal	Case of study	Dataset	Evaluation criteria
[108]	2016	EEG	Automatic Sleep Staging	Subjects recruited by researchers	Accuracy=74% Sensitivity=74% Precision=91% F-score = 81%
[109]	2017	EEG	Automatic Sleep Staging	UCD and MIT-BIH	Accuracy=92% Kappa coefficient=0.84
[110]	2017	EEG	Automatic Sleep Staging	Sleep EDF Database	Precision=93% Sensitivity=78% F1-score=84% Accuracy=86%
[111]	2018	EEG	Automatic Sleep Staging	SHHS	Accuracy=87 % Kappa coefficient=0.81
[112]	2018	Facial images	Sleep Apnea Prediction	Subjects recruited by researchers	Accuracy=67.42%
[113]	2018	EEG, EOG, EMG	Automatic Sleep Staging	Sleep EDF	Accuracy=82.3%
				MASS	Accuracy=83.6%
[99]	2020	heartbeat signal, respiratory signal, and movement signal	Automatic Sleep Staging	PSG records created at Charite Clinic in Berlin	Accuracy=55% F-score =44% Precision=48% Recall= 47%
[114]	2020	PSG	Sleep stage classification for child patients	SDCP	Accuracy=84.27% , F-score =72.51%
				Sleep EDF	Accuracy=90.89%
[115]	2020	ECG	Automatic Sleep Staging	SHHS	Accuracy=77% Kappa coefficient=0.66
[105]	2020	ECG	Sleep Apnea Detection	Physionet Apnea-ECG database	Accuracy=88.23% AUC=0.9453
[116]	2020	Respiratory signals from accelerometer and pressure transducer	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=94%
[117]	2021	EEG	Automatic Sleep Staging	SHHS1	Accuracy=88.1% F-score = 74.5% Precision=80.40% Recall= 73.90%
				Sleep EDF Database	Accuracy=85.3% F-score = 70.80% Precision=75.10% Recall= 71%
[107]	2021	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=87.21% Kappa coefficient=0.80
[118]	2021	EEG	Automatic Sleep Staging	UCD	Accuracy=77.2%
				Sleep EDF Database	Accuracy=83.4%

[106]	2021	EEG	Automatic Sleep Staging	Sleep EDF Database	Accuracy=86.6% MF1=80.9%
[119]	2022	EEG,EOG	Automatic Sleep Staging	ISRUC sleep database	Accuracy=88.85% Kappa coefficient=0.83

- **Mixture of multiple deep networks:** In recent years, there has been increasing interest in combining multiple deep neural networks to improve the accuracy and robustness of machine learning systems, where each network is tasked with a specific aspect of the analysis, leading to more efficient and effective processing of complex data. Most of these networks leverage a combination of convolutional and recurrent architectures, often using various specialized variants of each to achieve optimal performance for specific tasks. Table 6 shows the proposed methods. [120] (see Table 6) proposes a network architecture with two integral components. The first component employs parallel convolutional networks for structural feature extraction from raw electroencephalogram signals, while the second component utilizes bidirectional long short-term memory networks to capture temporal dependencies. Achieving an accuracy of 86.2%, this method highlights the synergy between spatial and temporal information, but computational demands and potential overfitting require scrutiny. [88] introduces a hybrid approach, combining a multi-layer perceptron with a long short-term memory network for classification. Utilizing spectral features from the electroencephalogram signal, the method attains an accuracy of 81.65%. This amalgamation of networks offers a compromise between feature learning and temporal dependencies, though further exploration into the impact of spectral representation is warranted. [121] proposes a comprehensive method for insomnia detection based on EEG signals, incorporating a pre-training module with convolutional and fully connected layers, and a subsequent insomnia detection module with multiple LSTM networks. The intricate design achieves an impressive accuracy of 90.90%. The multi-module architecture suggests a nuanced understanding of feature extraction and classification, though potential complexity and interpretability challenges should be considered. In [122], a portable device for sleep apnea detection combines a microphone and accelerometer to capture respiratory sounds and movements. Utilizing a deep neural network with convolutional and LSTM layers, the method achieves an accuracy of 84%. The integration of diverse sensor data showcases adaptability, but potential limitations in data compatibility and generalizability should be explored. Other studies, including [123-125], echo the combination of convolutional and recurrent networks for sleep-related tasks, achieving favorable results. While these methods provide promising outcomes, a comprehensive evaluation considering computational demands, interpretability, and generalizability across diverse datasets is crucial for a holistic understanding of their applicability. More details of the mentioned methods has been brought in Table 6.

Table 6. Sleep studies based on mixed neural network

Paper	Year	Method	Signal	Case of study	Dataset	Evaluation criteria
[120]	2017	CNN + Bi-LSTM	EEG	Automatic Sleep Staging	Sleep EDF Dataset MASS Dataset	Accuracy=86.2% F1-score=81.7%

[121]	2021	CNN + RNN	EEG	insomnia detection	Subjects recruited by researchers	Accuracy=90.90% MF1=90.90%
[88]	2018	MLP + LSTM	EEG	Automatic Sleep Staging	MASS dataset	Accuracy=81.65% MF1=76.59%
[122]	2020	CNN + LSTM	Tracheal movements	sleep apnea severity estimation	Subjects recruited by researchers	Accuracy=84% Sensitivity=81% Specificity=87%
[123]	2022	CNN, LSTM	EEG	Automatic Sleep Staging	Subjects recruited by researchers	Accuracy=79.7% Kappa=0.72
[124]	2022	CNN-RLM, BiGRU-TDM	ECG	Sleep Apnea Detection	PhysioNet ECG Sleep Apnea	Accuracy=91.1% Sensitivity=88.9% Specificity=92.4%
[125]	2022	Bi-LSTM, TCN	PPG	Sleep Apnea Detection	Subjects recruited by researchers	Accuracy=93.39% Sensitivity=98.98% Specificity=94.37%
[126]	2023	CNN, LSTM	EEG,EOG	Automatic Sleep Staging	Sleep EDF Dataset	Accuracy=86.30% Sensitivity=80.20% Specificity=96.40% Kappa=0.81 MF1=80.60%

4. Discussion

The field of sleep-related studies has witnessed a diverse array of methodologies, each with its strengths and challenges. As we navigate through the effectiveness of various classifiers and deep learning architectures, there are notable opportunities for improvement and exploration to advance the understanding and diagnosis of sleep disorders.

4.1. Classifiers in Sleep-Related Studies

4.1.1. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) stands out for its simplicity, adaptability, and commendable accuracy in various sleep-related tasks. Its effectiveness is exemplified in applications such as sleep apnea detection using smartphone sensors [49], where it achieves impressive accuracy rates of 97.7% and 98.2%. However, the method's performance varies across studies and datasets, outperforming SVM in certain applications [48] while facing challenges in maintaining superiority across diverse datasets and feature extractions [51]. The choice between classifiers should be made judiciously, considering the intricacies of the dataset and the nature of the extracted features.

4.1.2. Ensemble Learning Methods

Ensemble learning methods, particularly Bagging and Random Forest, demonstrate prowess in managing diverse physiological data and achieving exceptional accuracy in sleep apnea detection. The bagging method coupled with an ADtree classifier [56] showcases adaptability and parallel processing capabilities inherent in ensemble techniques, delivering an impressive accuracy of 95.96%. However, caution is warranted due to potential computational demands and the looming specter of overfitting, as observed in instances such as [56] and [58]. The comparative prowess of ensemble methods over individual classifiers, as emphasized in [48], positions them as efficient tools for capturing intricate patterns within respiration signals. Yet, their computational efficiency, as

seen in [57], may come at the cost of slightly diminished accuracy (83.26%). The evolving paradigm towards ensemble learning in sleep-related studies necessitates ongoing exploration to optimize ensemble size dynamically based on the dataset's complexity.

4.1.3. Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) demonstrates divergent outcomes in sleep apnea detection. Explored in [40], ELM yields remarkable results with an accuracy of 83.77%, surpassing alternative classifiers. However, in [48], integrating ELM results in a diminished accuracy of 53.02%, indicating potential challenges in capturing nuanced patterns within the specific dataset. Careful consideration of dataset characteristics and optimization strategies is essential when employing ELM.

4.1.4. Discriminant Analysis

Discriminant Analysis, employed in various sleep-related studies, showcases both successes and challenges. For instance, [77] achieves a 70% accuracy in sleep apnea detection, while [51] attains an impressive accuracy of 98.70%. Discrepancies such as those observed in [48], where Discriminant Analysis lags behind KNN and ensemble learning methods, underscore the need for a methodological paradigm shift in classifier selection.

4.1.5. Support Vector Machine (SVM)

Support Vector Machine (SVM) emerges as a versatile and potent classifier, excelling in scenarios demanding advanced signal processing, intricate feature extraction, and adaptability to diverse physiological signal modalities. Despite challenges observed in [50], where SVM yielded a sensitivity of 51.07%, innovative applications, as seen in [51] and [82], showcase SVM's adaptability to unconventional data sources with accuracies of 98.70% and 93.52%, respectively.

4.1.6. Naive Bayes Classifier

Naive Bayes classifiers, known for their simplicity and efficiency, emerge as valuable tools in sleep-related studies. Their ease of implementation and effectiveness in handling high-dimensional data make them suitable for quick initial assessments and baseline modeling. Despite their simplicity, Naive Bayes classifiers demonstrate strengths when the assumption of independence among features holds reasonably well. The model's efficiency becomes evident in scenarios where computational resources are a consideration. However, the reliance on the assumption of feature independence poses challenges, especially in complex datasets where dependencies among features are prevalent. In the intricate landscape of physiological signals, these dependencies may hinder performance. Recommendations for future research involve exploring extensions or adaptations of the Naive Bayes model to handle feature dependencies. Combining the strengths of Naive Bayes with other classifiers or ensemble methods could provide a more robust approach, particularly in scenarios with complex and interdependent features. While the Naive Bayes classifier contributes to the array of sleep study methodologies, researchers must judiciously consider its assumptions and suitability to the specific characteristics of the dataset to obtain meaningful results.

4.2. Deep Learning Architectures in Sleep-Related Studies

4.2.1 Deep Belief Nets

Deep Belief Nets exhibit remarkable capabilities in sleep-related studies, showcasing impressive accuracies in sleep apnea detection [96] and automatic sleep staging [97]. The strength of Deep Belief Nets lies in their hierarchical learning approach, with unsupervised training facilitating the extraction of intricate features, followed by supervised fine-tuning for precise classification. Challenges such as architectural complexity and the need for substantial labeled data remain considerations. Despite these challenges, the application of Deep Belief Nets marks a significant stride in advanced sleep disorder diagnostics.

4.2.2. Recurrent Neural Networks (RNN)

The efficacy of Recurrent Neural Networks (RNN) in sleep-related tasks is evident, with studies achieving notable accuracies in automatic sleep staging [101], sleep apnea prediction [102], and handling multiple physiological signals [99]. However, the performance of RNN is contingent on multiple factors, and the choice between RNNs and other architectures demands a meticulous consideration of signal characteristics. The comparative exploration of a 1-dimensional CNN and an LSTM-based RNN in [99] highlights the sensitivity of RNNs to intricate temporal dependencies, as the CNN outperforms the LSTM in this specific context. The LSTM's relatively lower accuracy in [99] unveils its sensitivity to the intricate temporal dependencies of sleep-related signals. Conversely, in [100], the LSTM network achieves an exceptional accuracy of 94.95% for predicting sleep apnea occurrences using electrocardiogram segmented samples.

4.2.3. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) exhibit versatility in various sleep-related applications, such as classifying electrocardiogram signals [105], automatic sleep staging [109], and sleep apnea detection [116]. The success of CNNs in these studies emphasizes their adaptability to different signal modalities and complex dependencies within physiological signals. Challenges such as potential class imbalance or noise in signals must be carefully considered, as seen in the varied accuracies achieved across studies. The exploration of novel CNN architectures, as seen in [109], showcases the potential for specialized models to address specific challenges in sleep-related studies. Proposed architectures combining structural feature extraction through CNNs and temporal dependency capture through LSTM networks [120] exhibit promising accuracy in sleep-related tasks.

4.2.4. Integration of Traditional Feature Extraction with Deep Structures

To further enhance the capabilities of sleep-related studies, a promising avenue is the integration of traditional feature extraction methods with deep neural networks. Traditional methods, such as signal processing techniques and handcrafted feature extraction, have long proven effective in capturing specific characteristics of sleep-related signals. Combining these traditional methods with deep learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), could yield a synergistic approach. For instance, utilizing CNNs to automatically learn hierarchical features from raw signals and then extracting handcrafted features from the learned representations may enhance the interpretability and performance of the models.

4.2.5. Combining Different Deep Neural Networks

Moreover, exploring the fusion of different deep neural networks presents an opportunity for comprehensive signal analysis. Combining the strengths of CNNs for spatial feature extraction and RNNs for capturing temporal dependencies, as demonstrated in [120], can lead to more robust models. Additionally, incorporating attention mechanisms within deep architectures can improve the networks' ability to focus on relevant information, enhancing performance in complex signal patterns.

4.2.6. State-of-the-Art Deep Networks

Several state-of-the-art deep networks hold potential for application in sleep-related studies. Transformer-based architectures, such as the Transformer and its variants (e.g., BERT, GPT), have demonstrated remarkable success in natural language processing tasks and sequential data analysis. Adapting these architectures to model the temporal dependencies within sleep-related signals could prove beneficial, especially in tasks like automatic sleep staging.

Graph Neural Networks (GNNs) represent another cutting-edge approach. GNNs excel in modeling relationships within graph-structured data, which can be valuable in scenarios where sleep-related signals exhibit complex interdependencies. Applying GNNs to capture the intricate connections between physiological signals may offer new insights and improve diagnostic accuracy.

Furthermore, the utilization of pre-trained models, such as those trained on large-scale image or audio datasets, can be explored for transfer learning. Fine-tuning these models on sleep-related data might leverage the learned features from diverse domains, potentially enhancing the performance of diagnostic models.

4.3. Proposed Roadmap for Future Research

The reviewed literature showcases a diverse array of methodologies for sleep-related studies, each with its strengths and challenges. To advance the field, a unified roadmap encompassing future directions based on the conclusions drawn from various methods is proposed:

Adaptive Classifiers: Developing classifiers that dynamically adjust hyperparameters based on dataset characteristics, ensuring optimal performance across diverse datasets and feature extractions.

Cross-Modality Integration: Exploring methodologies that integrate information from multiple physiological signal modalities, promoting a comprehensive understanding of sleep patterns and disorders.

Hybrid Models: Continuing to explore hybrid architectures that synergize the strengths of different models, potentially incorporating attention mechanisms or reinforcement learning to enhance feature extraction and temporal dependency modeling.

Interpretability and Generalizability: Prioritizing research into the interpretability of features extracted by advanced models, ensuring a deeper understanding of the learned representations. Assess the generalizability of models across diverse datasets and populations.

Collaborative Efforts and Standardized Datasets: Encouraging collaborative efforts across institutions to accumulate diverse datasets necessary for training complex models. Promote the use of standardized datasets to facilitate benchmarking and comparison across different methodologies.

Interdisciplinary Research: Fostering interdisciplinary research collaborations between experts in machine learning, signal processing, and sleep medicine to ensure a holistic approach to sleep-related diagnostic solutions.

In conclusion, the proposed roadmap emphasizes the need for adaptive classifiers that dynamically adjust to dataset characteristics, cross-modality integration to harness information from multiple sources, and continued exploration of hybrid models. Future studies should prioritize interpretability, scalability, and generalizability to ensure the practicality of advanced diagnostic tools for sleep disorders. Collaborative efforts, standardized datasets, and interdisciplinary research can collectively drive the field toward more accurate, robust, and accessible sleep-related diagnostic solutions. For a comprehensive overview, please refer to the detailed table summarizing each methodology's attributes, challenges, and recommendations presented below.

Table 7. Cybernetic Landscape of Sleep Study Methodologies: Strengths, Challenges, and Future Directions

Methodology	Strengths	Challenges	Recommendations and Future Directions
Naive bayes classifier	- Simplicity and ease of implementation - Efficiency with high-dimensional data	- Assumption of independence among features may not hold in complex datasets	- Suitable for quick initial assessments and as a baseline model. - Explore extensions and adaptations for handling dependencies in features.
K-Nearest Neighbors (KNN)	- Simplicity, adaptability - High accuracy (e.g., 97.7%-98.2%) [38]	- Performance variability across studies and datasets - Challenges in maintaining superiority [37]	- Carefully choose classifiers based on dataset intricacies and feature characteristics.
Ensemble Learning Methods	- Prowess in managing diverse data - Exceptional accuracy (e.g., 95.96%) [45]	- Computational demands - Overfitting risks [45][47] - Slightly diminished accuracy (83.26%) [51]	- Optimize ensemble size dynamically based on dataset complexity. - Address computational demands and overfitting risks.
Extreme Learning Machine (ELM)	- Remarkable results (e.g., 83.77%) [29]	- Diminished accuracy in specific datasets (e.g., 53.02%) [37]	- Consider dataset characteristics and optimization strategies when employing ELM. - Investigate methods to enhance accuracy in specific datasets.
Discriminant Analysis	- Successes (e.g., 98.70%) [42] - Radar panel application [66]	- Diminished accuracy compared to KNN and ensemble methods [37] - Varied performance across studies	- Advocate for a methodological paradigm shift in classifier selection. - Investigate factors contributing to varied performance.
Support Vector Machine (SVM)	- Versatile, adaptable to diverse physiological modalities - Impressive accuracy (e.g., 98.70%) [42]	- Sensitivity challenges in certain scenarios (e.g., 51.07%) [40] - Varied performance across studies	- Explore SVM adaptability to unconventional data sources. - Investigate factors contributing to varied performance.
Deep Belief Nets	- Hierarchical learning - Unsupervised training - Impressive accuracies [85][86]	- Architectural complexity - Need for substantial labeled data	- Investigate ways to overcome architectural complexity and data requirements. - Explore semi-supervised learning approaches for labeled data scarcity.

Recurrent Neural Networks (RNN)	- Efficacy in multiple tasks - Sensitivity to temporal dependencies	- Choice sensitivity to signal characteristics - Varied performance across studies	- Meticulously consider signal characteristics in choosing between RNNs and other architectures. - Investigate factors contributing to varied performance.
Convolutional Neural Networks (CNN)	- Versatile in various applications - Adaptability to different signal modalities [94][98][105]	- Challenges in potential class imbalance, noise in signals	- Address potential challenges such as class imbalance and noise. - Explore novel CNN architectures for specialized challenges [98][108].
Integration of Traditional Feature Extraction with Deep Structures	- Synergistic approach leveraging traditional feature extraction and deep learning [Recommendation in 4.2.4]	- Limited studies exploring integrated approaches	- Explore the combination of traditional methods with deep learning for enhanced interpretability and performance. - Conduct more research on integrated approaches.
Combining Different Deep Neural Networks	- Fusion for comprehensive signal analysis [Recommendation in 4.2.5]	- Limited studies exploring combined architectures	- Explore combining strengths of CNNs and RNNs for more robust models. - Incorporate attention mechanisms for improved focus. - Conduct more research on combined architectures.
State-of-the-Art Deep Networks	- Potential application in sleep-related studies [Recommendation in 4.2.6]	- Limited studies applying state-of-the-art networks	- Investigate the adaptation of Transformer-based architectures, Graph Neural Networks, and pre-trained models for improved diagnostic accuracy and insights. - Conduct studies applying state-of-the-art networks in sleep-related tasks.
Proposed Roadmap for Future Research	- Adaptive classifiers - Cross-modality integration - Hybrid models - Interpretability - Generalizability - Collaborative efforts - Standardized datasets - Interdisciplinary research	- -	- Emphasizes future directions to enhance sleep-related diagnostic tools.

4.4. Synthesis, Key Insights, and Methodological Guidelines

In consolidating the diverse landscape of automated sleep studies, this section aims to distill overarching insights, crystallize key findings, and delineate methodological guidelines that transcend individual methodologies. Our comprehensive survey not only enumerates the strengths and challenges of various classifiers and deep learning architectures but also aims to provide a robust framework for decision-making in the pursuit of advancing the field.

4.4.1. Key Findings and Methodological Best Practices

The exploration of classifiers underscores the pivotal role of adaptability in method selection. K-Nearest Neighbors (KNN) shines as an exemplar of simplicity and accuracy, particularly in scenarios where these attributes are paramount. Ensemble learning methods, notwithstanding computational demands, exhibit efficiency in handling diverse physiological data. The versatility of Support Vector Machines (SVM) and the efficiency of Naive Bayes classifiers are apparent, but their nuanced performance nuances emphasize the indispensability of discerning dataset intricacies. Within the realm of deep learning architectures, the hierarchical learning approach of Deep Belief Nets, coupled with the temporal modeling capabilities of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), emerges as promising. The integration of traditional feature extraction methods with deep structures not only augments interpretability but also signifies a synergistic approach to signal analysis. The fusion of different deep neural networks, signifies potential for enhanced model robustness. Data science and artificial

intelligence are still in its infancy. Several effective methods will emerge and could present interesting solutions to automatized sleep study, such as, advanced deep learning, domain adaptation and graph signal processing.

4.4.2. Methodological Guidelines for Future Research

Navigating the intricate landscape of automated sleep studies requires a judicious methodological approach. Researchers should adopt adaptive classifiers that dynamically adjust hyperparameters based on dataset characteristics, ensuring optimal performance across diverse datasets and feature extractions. The exploration of hybrid models, amalgamating the strengths of diverse architectures, warrants further attention. Researchers must meticulously consider the intricacies of temporal dependencies, spatial features, and computational efficiency in the selection between models like RNNs, CNNs, or their integration.

4.4.3. Addressing Challenges and Charting Future Directions

Despite significant strides, challenges persist. The interpretability of features extracted by advanced models stands as a foremost priority, fostering a deeper comprehension of learned representations. Rigorous assessment of model generalizability across diverse datasets and populations is paramount for real-world applicability. Future research endeavors should champion collaborative efforts, the establishment of standardized datasets, and interdisciplinary collaborations. An adaptive roadmap that encompasses dynamic classifiers, cross-modality integration, hybrid models, interpretability, scalability, and generalizability is proposed to steer the field towards the development of more accurate, robust, and accessible sleep-related diagnostic solutions. This synthesized perspective not only distills key findings but also offers nuanced methodological guidelines, equipping researchers and practitioners with a comprehensive framework to navigate the complexities inherent in automated sleep studies.

4.5. Signal modalities

In the studies reviewed, electroencephalogram was the most utilized signal for sleep analysis, likely due to its richness in sleep-related information. In contrast, Electrooculogram and fMRI were used less frequently, with the former providing limited information and the latter being difficult to acquire. As some signal modalities alone may not provide sufficient and comprehensive information, researchers often combine multiple signals. A chart showing the percentage of signal modalities used in the reviewed studies is presented in Figure 1.

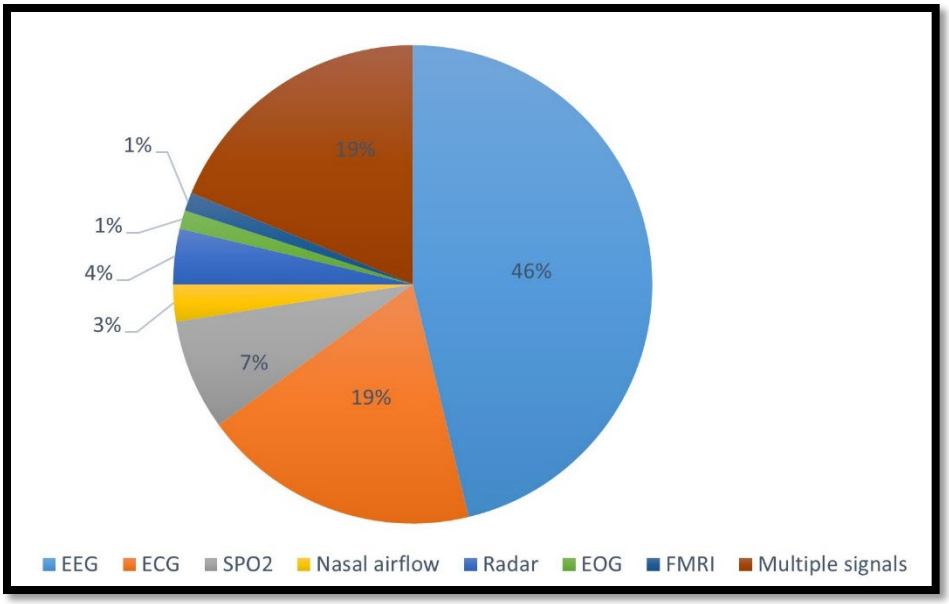

Fig. 1. Percentage of usage of each signal in the reviewed sleep studies.

4.6. Traditional Machine learning methods

In the studies reviewed, Ensemble learning-based classifiers along with SVM and KNN were the most commonly used machine learning methods for sleep analysis, while Restricted Boltzmann Machine, Extreme Learning Machine and Discriminant analysis were used less frequently. The popularity of SVM and KNN can be attributed to their respective advantages in dealing with high-dimensional data and time efficiently. Ensemble learning-based classifiers outperformed individual models in terms of predictive accuracy and stability, particularly in datasets with both linear and non-linear data types. The chart in Figure 2 provides a breakdown of the percentage use of each machine learning method in the reviewed studies.

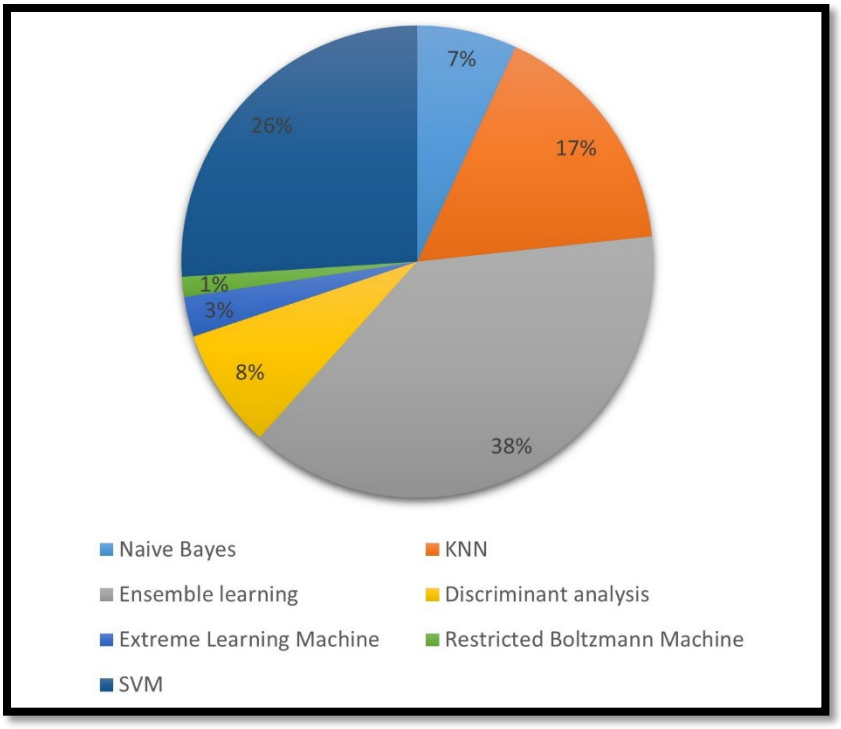

Fig. 2. Percentage of usage of each traditional machine learning method in the reviewed sleep studies.

4.7. Deep learning methods

In the studies reviewed, CNN and a mixture of multiple deep networks were commonly used for sleep analysis, while deep belief nets and RNNs were less frequently utilized as standalone techniques. A combined architecture of CNN and RNN was found to be particularly effective for sleep analysis. A chart in Figure 3 provides the percentage use of each deep learning method in the reviewed studies.

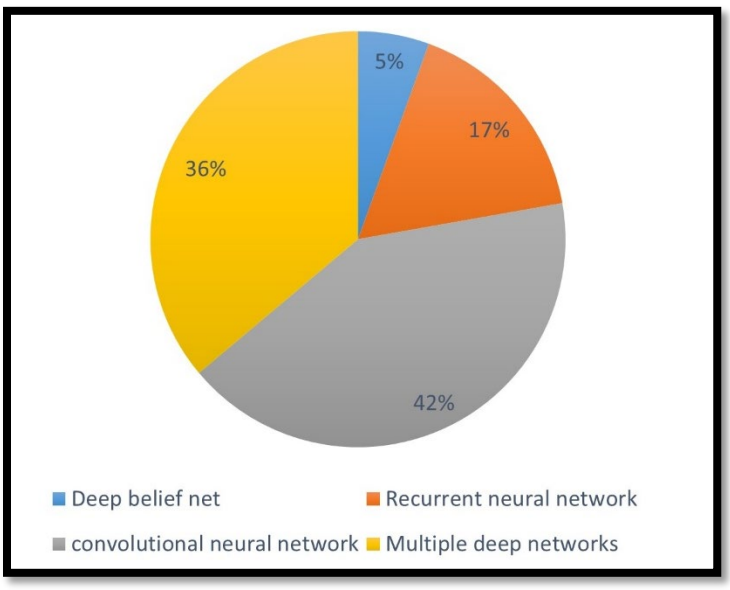

Fig. 3. Percentage of usage of each deep learning method in the reviewed sleep studies.

4.8. Datasets

In the studies reviewed, researchers predominantly used their own datasets for sleep analysis. However, among the available dataset, the Sleep-EDF Database was the most commonly used due to its comprehensiveness, including data from multiple participants and various signal modalities. A chart in Figure 4 illustrates the percentage utilization of different datasets in the reviewed studies.

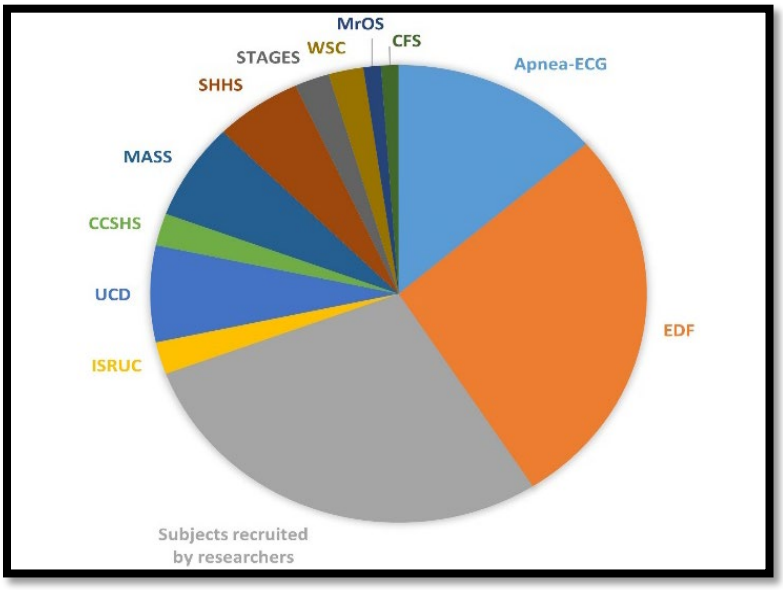

Fig. 4. Percentage of usage of different datasets in the reviewed sleep studies.

5. Challenges and future trends for the sleep study

Some challenges are as follows:

- The effective utilization of temporal and spectral features from multi-channel biological signals continues to pose a significant challenge in the field of signal processing.
- Due to variations in biological signals among individuals, it's crucial, yet difficult, to address these disparities in order to enhance the generalizability of deep neural networks.
- The interpretability of deep learning models remains a significant challenge.
- The high computational requirements of most deep learning architectures restrict their use to applications that can accommodate real-time and embedded processing.
- Some signal modalities have received less attention, resulting in a lack of recording of these types of signals.
- Numerous proposed techniques heavily depend on datasets from a handful of small-scale databases for model development, and it remains unclear whether these models can be effectively applied to new datasets.

- The vast majority of available datasets exhibit a notable class imbalance problem, whereby the number of instances of sleep stages is highly disproportionate.

Future research can focus on addressing the aforementioned challenges. Furthermore, there is a need for further investigation in the following areas:

- The popularity of user-friendly health monitoring devices, such as smartwatches, has surged in recent years. Future research efforts could focus on developing algorithms to analyze data from these devices and creating lightweight algorithms that can be implemented on them. The primary challenge for these devices is their limited memory and processing power.
- Although body movements and facial expressions can provide valuable insights for sleep studies, these areas have received relatively little attention compared to other sleep-related factors. Therefore, there is a need for further research in these areas. Various techniques, such as image processing and radar signal processing, can be employed to better capture and analyze this information.
- Based on the sleep study findings, it appears that exploring innovative options such as smart pillows and mattresses could be a promising direction.
- While deep networks have been extensively studied in the realm of sleep research, there remains an underexplored territory concerning the potential applications of cutting-edge models like generative autoencoders, reinforcement learning algorithms, and graph neural networks. These models offer unique capabilities, such as the generation of representative features (generative autoencoders), optimization of decision-making processes (reinforcement learning algorithms), and the ability to capture intricate data relationships (graph neural networks). Given their distinctive features, future investigations could delve into the application of these models in sleep-related tasks, aiming to leverage their specific strengths to enhance the understanding and analysis of sleep patterns and disorders.
- Exploring the statistics of transitional stages in sleep may yield valuable insights into sleep physiology. Therefore, future studies could focus on investigating sleep stage transitions.
- In terms of sleep analysis methods, the works can be focused on deep learning methods that could detect more than one kind of sleep disorder.

6. Conclusion

The present study provides an overview of research conducted in the area of automated sleep analysis, encompassing the automatic classification of sleep stages and disorders. In recent years, there has been a significant surge in research efforts within this domain, with the incorporation of various machine learning techniques and deep neural network architectures. Notably, deep neural networks and their combinations have gained substantial popularity in numerous recently published papers. Researchers have explored diverse types of signals, either individually or in combination, and employed various datasets. Some have utilized existing datasets, while others have curated their own. Over time, the accuracy of computer-based sleep stage and disorder classification has witnessed notable advancements, thereby offering a valuable tool for diagnosing and predicting neurological and physiological conditions.

Acknowledgment

This work was partly supported by the Natural Sciences and Engineering Research Council of Canada Grant and partly by the Research Chair on Signal and Intelligent High-Performance Systems.

Declarations

Funding: Partial financial support was received from the Natural Sciences and Engineering Research Council of Canada Grant and partly by the Research Chair on Signal and Intelligent High-Performance Systems.
Competing interests: The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Venter RE. Role of sleep in performance and recovery of athletes: a review article. South African Journal for Research in Sport, Physical Education and Recreation. 2012;34(1):167-84.
2. Fonseca AG, Genzel L. Sleep and academic performance: considering amount, quality and timing. Current Opinion in Behavioral Sciences. 2020;33:65-71.
3. Kalantar-Zadeh K, Lockwood MB, Rhee CM, Tantisattamo E, Andreoli S, Balducci A, et al. Patient-centred approaches for the management of unpleasant symptoms in kidney disease. Nature Reviews Nephrology. 2022;18(3):185-98.
4. Darchia N, Oniani N, Sakhelashvili I, Supatashvili M, Basishvili T, Eliozishvili M, et al. Relationship between sleep disorders and health related quality of life—results from the Georgia SOMNUS study. International journal of environmental research and public health. 2018;15(8):1588.
5. Alsolai H, Qureshi S, Iqbal SMZ, Vanichayobon S, Henesey LE, Lindley C, et al. A Systematic Review of Literature on Automated Sleep Scoring. IEEE Access. 2022.
6. Aboalayon KAI, Faezipour M, Almuhammadi WS, Moslehpour S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. Entropy. 2016;18(9):272.
7. Loh HW, Ooi CP, Vicnesh J, Oh SL, Faust O, Gertych A, et al. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). Applied Sciences. 2020;10(24):8963.
8. Markun LC, Sampat A. Clinician-focused overview and developments in polysomnography. Current sleep medicine reports. 2020;6:309-21.
9. Espinosa MA, Ponce P, Molina A, Borja V, Torres MG, Rojas M. Advancements in Home-Based Devices for Detecting Obstructive Sleep Apnea: A Comprehensive Study. Sensors. 2023;23(23):9512.
10. Liang S-F, Shih Y-H, Chen P-Y, Kuo C-E. Development of a human-computer collaborative sleep scoring system for polysomnography recordings. PloS one. 2019;14(7):e0218948.
11. Imtiaz SA. A systematic review of sensing technologies for wearable sleep staging. Sensors. 2021;21(5):1562.
12. Roebuck A, Monasterio V, Gederi E, Osipov M, Behar J, Malhotra A, et al. A review of signals used in sleep analysis. Physiological measurement. 2013;35(1):R1.
13. Panossian LA, Avidan AY. Review of sleep disorders. Medical Clinics of North America. 2009;93(2):407-25.

14. De Fazio R, Mattei V, Al-Naami B, De Vittorio M, Visconti P. Methodologies and Wearable Devices to Monitor Biophysical Parameters Related to Sleep Dysfunctions: An Overview. Micromachines. 2022;13(8):1335.
15. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation. 2000;101(23):e215-e20.
16. Zhang G-Q, Cui L, Mueller R, Tao S, Kim M, Rueschman M, et al. The National Sleep Research Resource: towards a sleep data commons. Journal of the American Medical Informatics Association. 2018;25(10):1351-8.
17. Rosen CL, Larkin EK, Kirchner HL, Emancipator JL, Bivins SF, Surovec SA, et al. Prevalence and risk factors for sleep-disordered breathing in 8-to 11-year-old children: association with race and prematurity. The Journal of pediatrics. 2003;142(4):383-9.
18. Khalighi S, Sousa T, Santos JM, Nunes U. ISRUC-Sleep: A comprehensive public dataset for sleep researchers. Computer methods and programs in biomedicine. 2016;124:180-92.
19. O'reilly C, Gosselin N, Carrier J, Nielsen T. Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. Journal of sleep research. 2014;23(6):628-35.
20. Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, et al. The sleep heart health study: design, rationale, and methods. Sleep. 1997;20(12):1077-85.
21. Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Ensrud KE, Stefanick ML, et al. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the osteoporotic fractures in men sleep study. Journal of the American Geriatrics Society. 2011;59(12):2217-25.
22. Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). Sleep. 2015;38(6):877-88.
23. Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, Hla KM. Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. WMJ: official publication of the State Medical Society of Wisconsin. 2009;108(5):246.
24. Redline S, Tishler PV, Tosteson TD, Williamson J, Kump K, Browner I, et al. The familial aggregation of obstructive sleep apnea. American journal of respiratory and critical care medicine. 1995;151(3):682-7.
25. Rosen CL, Auckley D, Benca R, Foldvary-Schaefer N, Iber C, Kapur V, et al. A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. Sleep. 2012;35(6):757-67.
26. Chatterjee S, Thakur RS, Yadav RN, Gupta L, Raghuvanshi DK. Review of noise removal techniques in ECG signals. IET Signal Processing. 2020;14(9):569-90.
27. Diykh M, Li Y, Wen P. EEG sleep stages classification based on time domain features and structural graph similarity. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2016;24(11):1159-68.
28. Bracewell RN, Bracewell RN. The Fourier transform and its applications: McGraw-Hill New York; 1986.
29. Akin M, Kiymik MK. Application of periodogram and AR spectral analysis to EEG signals. Journal of Medical Systems. 2000;24:247-56.
30. Solomon Jr OM. PSD computations using Welch's method. NASA STI/Recon Technical Report N. 1991;92:23584.
31. Sejdić E, Djurović I, Jiang J. Time–frequency feature representation using energy concentration: An overview of recent advances. Digital signal processing. 2009;19(1):153-83.

32. Sadowsky J. The continuous wavelet transform: A tool for signal investigation and understanding. Johns Hopkins APL Technical Digest. 1994;15:306-.
33. Umapathy K, Ghoraani B, Krishnan S. Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking. EURASIP Journal on Advances in Signal Processing. 2010;2010:1-28.
34. Tsinalis O, Matthews PM, Guo Y. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. Annals of biomedical engineering. 2016;44:1587-97.
35. Altae AA, Rad AE, Tati R. Comparative Study on Effective Feature Selection Methods. International Journal for Innovative Engineering & Management Research, Forthcoming. 2023.
36. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Frontiers in Bioinformatics. 2022;2:927312.
37. Alvárez-Estévez D, Sánchez-Marono N, Alonso-Betanzos A, Moret-Bonillo V. Filter-based Feature Selection for the detection of Arousals in Sleep Studies.
38. Ravelo-García AG, Kraemer JF, Navarro-Mesa JL, Hernández-Pérez E, Navarro-Esteva J, Juliá-Serdá G, et al. Oxygen saturation and RR intervals feature selection for sleep apnea detection. Entropy. 2015;17(5):2932-57.
39. Yang F-J, editor An implementation of naive bayes classifier. 2018 International conference on computational science and computational intelligence (CSCI); 2018: IEEE.
40. Hassan AR, editor A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram. 2015 International Conference on Electrical & Electronic Engineering (ICEEE); 2015: IEEE.
41. Memis G, Sert M, editors. Multimodal classification of obstructive sleep apnea using feature level fusion. 2017 IEEE 11th international conference on semantic computing (ICSC); 2017: IEEE.
42. Jeon Y, Heo K, Kang SJ, editors. Real-time sleep apnea diagnosis method using wearable device without external sensors. 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); 2020: IEEE.
43. Dimitriadis SI, Salis C, Linden D. A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates. Clinical Neurophysiology. 2018;129(4):815-28.
44. Wei Y, Qi X, Wang H, Liu Z, Wang G, Yan X. A multi-class automatic sleep staging method based on long short-term memory network using single-lead electrocardiogram signals. IEEE Access. 2019;7:85959-70.
45. Zhang Z. Introduction to machine learning: k-nearest neighbors. Annals of translational medicine. 2016;4(11).
46. Davidovich MLY, Karasik R, Tal A, Shinar Z, editors. Sleep apnea screening with a contact-free under-the-mattress sensor. 2016 computing in cardiology conference (CinC); 2016: IEEE.
47. Ozdemir G, Nasifoglu H, Erogul O, editors. A time-series approach to predict obstructive sleep apnea (OSA) Episodes. Proceedings of the 2nd world congress on electrical engineering and computer systems and science; 2016.
48. Hassan AR, Bashar SK, Bhuiyan MIH, editors. Computerized obstructive sleep apnea diagnosis from single-lead ECG signals using dual-tree complex wavelet transform. 2017 IEEE region 10 humanitarian technology conference (R10-HTC); 2017: IEEE.
49. Camcı B, Kahveci AY, Arnrich B, Ersoy C, editors. Sleep apnea detection via smart phones. 2017 25th Signal Processing and Communications Applications Conference (SIU); 2017: IEEE.
50. Wang C, Chan J-H, Fang S-H, Cheng H-T, Hsu Y-L, editors. Novel sleep apnea detection based on UWB artificial intelligence mattress. 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS); 2019: IEEE.

51. Ivanko K, Ivanushkina N, Rykhalska A, editors. Identifying episodes of sleep apnea in ECG by machine learning methods. 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO); 2020: IEEE.
52. Hong H, Zhang L, Gu C, Li Y, Zhou G, Zhu X. Noncontact sleep stage estimation using a CW Doppler radar. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2018;8(2):260-70.
53. Kye S, Moon J, Lee T, Lee S, Lee K, Shin S-c, et al., editors. Detecting periodic limb movements in sleep using motion sensor embedded wearable band. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2017: IEEE.
54. Christensen JAE, Munk EGS, Peppard PE, Young T, Mignot E, Sorensen HBD, et al. The diagnostic value of power spectra analysis of the sleep electroencephalography in narcoleptic patients. Sleep medicine. 2015;16(12):1516-27.
55. Guan D, Yuan W, Lee Y-K, Najeebullah K, Rasel MK. A review of ensemble learning based feature selection. IETE Technical Review. 2014;31(3):190-8.
56. Burgos A, Goni A, Illarramendi A, Bermudez J. Real-time detection of apneas on a PDA. IEEE Transactions on Information Technology in Biomedicine. 2009;14(4):995-1002.
57. Xie B, Minn H. Real-time sleep apnea detection by classifier combination. IEEE Transactions on information technology in biomedicine. 2012;16(3):469-77.
58. Avcı C, Akbaş A. Sleep apnea classification based on respiration signals by using ensemble methods. Bio-medical materials and engineering. 2015;26(s1):S1703-S10.
59. Gutiérrez-Tobal GC, Álvarez D, Crespo A, Del Campo F, Hornero R. Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings. IEEE journal of biomedical and health informatics. 2018;23(2):882-92.
60. Sanders TH, McCurry M, Clements MA, editors. Sleep stage classification with cross frequency coupling. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
61. Hassan AR, Bhuiyan MIH. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. Biomedical Signal Processing and Control. 2016;24:1-10.
62. Hassan AR, Bhuiyan MIH. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. Biocybernetics and Biomedical Engineering. 2016;36(1):248-55.
63. Hassan AR, Bhuiyan MIH, editors. Automatic sleep stage classification. 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT); 2015: IEEE.
64. Hassan AR, Bashar SK, Bhuiyan MIH, editors. On the classification of sleep states by means of statistical and spectral features from single channel electroencephalogram. 2015 International conference on advances in computing, communications and informatics (ICACCI); 2015: IEEE.
65. Hassan AR, Bhuiyan MIH. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. Journal of neuroscience methods. 2016;271:107-18.
66. Li X, Cui L, Tao S, Chen J, Zhang X, Zhang G-Q. Hyclasss: a hybrid classifier for automatic sleep stage scoring. IEEE journal of biomedical and health informatics. 2017;22(2):375-85.
67. Klok AB, Edin J, Cesari M, Olesen AN, Jennum P, Sorensen HB, editors. A new fully automated random-forest algorithm for sleep staging. 2018 40th annual international conference of the ieee engineering in medicine and biology society (EMBC); 2018: IEEE.
68. Zhou J, Wang G, Liu J, Wu D, Xu W, Wang Z, et al. Automatic sleep stage classification with single channel EEG signal based on two-layer stacked ensemble model. IEEE Access. 2020;8:57283-97.
69. Cooray N, Andreotti F, Lo C, Symmonds M, Hu MT, De Vos M. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clinical Neurophysiology. 2019;130(4):505-14.

70. Sharma M, Patel V, Tiwari J, Acharya UR. Automated characterization of cyclic alternating pattern using wavelet-based features and ensemble learning techniques with eeg signals. Diagnostics. 2021;11(8):1380.
71. Hassan AR, Bashar SK, Bhuiyan MIH, editors. Automatic classification of sleep stages from single-channel electroencephalogram. 2015 annual IEEE India conference (INDICON); 2015: IEEE.
72. Hassan AR, Bhuiyan MIH. An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting. Neurocomputing. 2017;219:76-87.
73. Hassan AR, Subasi A. A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowledge-Based Systems. 2017;128:115-24.
74. Hassan AR, Bhuiyan MIH. Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. Computer methods and programs in biomedicine. 2017;140:201-10.
75. Liu J, Li Q, Xin Y, Lu X, editors. Obstructive sleep apnea detection using sleep architecture. 2020 IEEE International Conference on Mechatronics and Automation (ICMA); 2020: IEEE.
76. Tharwat A. Linear vs. quadratic discriminant analysis classifier: a tutorial. International Journal of Applied Pattern Recognition. 2016;3(2):145-80.
77. Javaid AQ, Noble CM, Rosenberg R, Weitnauer MA, editors. Towards sleep apnea screening with an under-the-mattress IR-UWB radar using machine learning. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA); 2015: IEEE.
78. De Chazal P, Sadr N, editors. Sleep apnoea classification using heart rate variability, ECG derived respiration and cardiopulmonary coupling parameters. 2016 38th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016: IEEE.
79. Jayawardhana M, de Chazal P, editors. Enhanced detection of sleep apnoea using heart-rate, respiration effort and oxygen saturation derived from a photoplethysmography sensor. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017: IEEE.
80. Albadra MAA, Tiuna S. Extreme learning machine: a review. International Journal of Applied Engineering Research. 2017;12(14):4610-23.
81. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications. 1998;13(4):18-28.
82. Zhang J, Zhang Q, Wang Y, Qiu C, editors. A real-time auto-adjustable smart pillow system for sleep apnea detection and treatment. Proceedings of the 12th international conference on Information processing in sensor networks; 2013.
83. Prabha A, Trivedi A, Kumar AA, Kumar CS, editors. Automated system for obstructive sleep apnea detection using heart rate variability and respiratory rate variability. 2017 international conference on advances in computing, communications and informatics (ICACCI); 2017: IEEE.
84. Deviaene M, Borzée P, Buyse B, Testelmans D, Van Huffel S, Varon C, editors. Pulse oximetry markers for cardiovascular disease in sleep apnea. 2019 Computing in Cardiology (CinC); 2019: IEEE.
85. Bajaj V, Pachori RB. Automatic classification of sleep stages based on the time-frequency image of EEG signals. Computer methods and programs in biomedicine. 2013;112(3):320-8.
86. Zhu G, Li Y, Wen P. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE journal of biomedical and health informatics. 2014;18(6):1813-21.
87. Lajnef T, Chaibi S, Ruby P, Aguera P-E, Eichenlaub J-B, Samet M, et al. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. Journal of neuroscience methods. 2015;250:94-105.

88. Dong H, Supratak A, Pan W, Wu C, Matthews PM, Guo Y. Mixed neural network approach for temporal sleep stage classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;26(2):324-33.
89. Seifpour S, Niknazar H, Mikaeili M, Nasrabadi AM. A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal. Expert Systems with Applications. 2018;104:277-93.
90. Rahman MM, Bhuiyan MIH, Hassan AR. Sleep stage classification using single-channel EOG. Computers in biology and medicine. 2018;102:211-20.
91. Lee MH, Kim N, Yoo J, Kim H-K, Son Y-D, Kim Y-B, et al. Multitask fMRI and machine learning approach improve prediction of differential brain activity pattern in patients with insomnia disorder. Scientific Reports. 2021;11(1):9402.
92. Khandoker AH, Palaniswami M, Karmakar CK. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. IEEE transactions on information technology in biomedicine. 2008;13(1):37-48.
93. Diykh M, Li Y, Abdulla S. EEG sleep stages identification based on weighted undirected complex networks. Computer methods and programs in biomedicine. 2020;184:105116.
94. Shahin M, Mulaffer L, Penzel T, Ahmed B, editors. A two stage approach for the automatic detection of insomnia. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018: IEEE.
95. Hua Y, Guo J, Zhao H, editors. Deep belief networks and deep learning. Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things; 2015: IEEE.
96. Mostafa SS, Mendonça F, Morgado-Dias F, Ravelo-García A, editors. SpO2 based sleep apnea detection using deep learning. 2017 IEEE 21st international conference on intelligent engineering systems (INES); 2017: IEEE.
97. Xia B, Li Q, Jia J, Wang J, Chaudhary U, Ramos-Murguialday A, et al., editors. Electrooculogram based sleep stage classification using deep belief network. 2015 International Joint Conference on Neural Networks (IJCNN); 2015: IEEE.
98. Medsker LR, Jain L. Recurrent neural networks. Design and Applications. 2001;5:64-7.
99. Stuburić K, Gaiduk M, Seepold R. A deep learning approach to detect sleep stages. Procedia Computer Science. 2020;176:2764-72.
100. Bahrami M, Forouzanfar M. Deep learning forecasts the occurrence of sleep apnea from single-lead ECG. Cardiovascular Engineering and Technology. 2022:1-7.
101. Kuo C-E, Chen G-T. Automatic sleep staging based on a hybrid stacked LSTM neural network: verification using large-scale dataset. IEEE access. 2020;8:111837-49.
102. Pathinarupothi RK, Rangan ES, Gopalakrishnan E, Vinaykumar R, Soman K, editors. Single sensor techniques for sleep apnea diagnosis using deep learning. 2017 IEEE international conference on healthcare informatics (ICHI); 2017: IEEE.
103. Pathinarupothi RK, Vinaykumar R, Rangan E, Gopalakrishnan E, Soman K, editors. Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep learning. 2017 IEEE EMBS international conference on biomedical & health informatics (BHI); 2017: IEEE.
104. Ajit A, Acharya K, Samanta A, editors. A review of convolutional neural networks. 2020 international conference on emerging trends in information technology and engineering (ic-ETITE); 2020: IEEE.
105. Sharan RV, Berkovsky S, Xiong H, Coiera E, editors. ECG-derived heart rate variability interpolation and 1-D convolutional neural networks for detecting sleep apnea. 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC); 2020: IEEE.

106. Eldele E, Chen Z, Liu C, Wu M, Kwoh C-K, Li X, et al. An attention-based deep learning approach for sleep stage classification with single-channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2021;29:809-18.
107. Cai Q, Gao Z, An J, Gao S, Grebogi C. A graph-temporal fused dual-input convolutional neural network for detecting sleep stages from EEG signals. IEEE Transactions on Circuits and Systems II: Express Briefs. 2020;68(2):777-81.
108. Tsinalis O, Matthews PM, Guo Y, Zafeiriou S. Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. arXiv preprint arXiv:161001683. 2016.
109. Zhang J, Wu Y. A new method for automatic sleep stage classification. IEEE transactions on biomedical circuits and systems. 2017;11(5):1097-110.
110. Vilamala A, Madsen KH, Hansen LK, editors. Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. 2017 IEEE 27th international workshop on machine learning for signal processing (MLSP); 2017: IEEE.
111. Sors A, Bonnet S, Mirek S, Vercueil L, Payen J-F. A convolutional neural network for sleep stage scoring from raw single-channel EEG. Biomedical Signal Processing and Control. 2018;42:107-14.
112. Islam SM, Mahmood H, Al-Jumaily AA, Claxton S, editors. Deep learning of facial depth maps for obstructive sleep apnea prediction. 2018 international conference on machine learning and data engineering (iCMLDE); 2018: IEEE.
113. Phan H, Andreotti F, Cooray N, Chén OY, De Vos M. Joint classification and prediction CNN framework for automatic sleep stage classification. IEEE Transactions on Biomedical Engineering. 2018;66(5):1285-96.
114. Huang X, Shirahama K, Li F, Grzegorzek M. Sleep stage classification for child patients using DeConvolutional Neural Network. Artificial Intelligence in Medicine. 2020;110:101981.
115. Sridhar N, Shoeb A, Stephens P, Kharbouch A, Shimol DB, Burkart J, et al. Deep learning for automated sleep staging using instantaneous heart rate. NPJ digital medicine. 2020;3(1):106.
116. Petrenko A. Breathmonitor: AI Sleep Apnea Mobile Detector. System Analysis & Intelligent Computing: Theory and Applications: Springer; 2022. p. 201-8.
117. Li F, Yan R, Mahini R, Wei L, Wang Z, Mathiak K, et al. End-to-end sleep staging using convolutional neural network in raw single-channel EEG. Biomedical Signal Processing and Control. 2021;63:102203.
118. Zhang J, Wu Y. Competition convolutional neural network for sleep stage classification. Biomedical Signal Processing and Control. 2021;64:102318.
119. Kwon K, Kwon S, Yeo W-H. Automatic and accurate sleep stage classification via a convolutional deep neural network and nanomembrane electrodes. Biosensors. 2022;12(3):155.
120. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;25(11):1998-2008.
121. Qu W, Kao C-H, Hong H, Chi Z, Grunstein R, Gordon C, et al. Single-channel EEG based insomnia detection with domain adaptation. Computers in biology and medicine. 2021;139:104989.
122. Hafezi M, Montazeri N, Saha S, Zhu K, Gavrilovic B, Yadollahi A, et al. Sleep apnea severity estimation from tracheal movements using a deep learning model. IEEE Access. 2020;8:22641-9.
123. Leino A, Korkalainen H, Kalevo L, Nikkonen S, Kainulainen S, Ryan A, et al. Deep learning enables accurate automatic sleep staging based on ambulatory forehead EEG. IEEE Access. 2022;10:26554-66.
124. Qin H, Liu G. A dual-model deep learning method for sleep apnea detection based on representation learning and temporal dependence. Neurocomputing. 2022;473:24-36.
125. Jothi ESJ, Anitha J, Hemanth DJ. A photoplethysmography-based diagnostic support system for obstructive sleep apnea using deep learning approaches. Computers and Electrical Engineering. 2022;102:108279.

126. Efe E, Ozsen S. CoSleepNet: Automated sleep staging using a hybrid CNN-LSTM network on imbalanced EEG-EOG datasets. Biomedical Signal Processing and Control. 2023;80:104299.